34

Molecular Dynamics Simulations of RNA Systems

Pascal Auffinger and Andrea C. Vaiana

34.1 Introduction

As a result of important methodological advances, the number of molecular dynamics (MD) simulations related to RNA systems has become significant, and simulations on more than 30 different RNA, RNA–protein and RNA–ligand systems have been reported (Table 34.1). However, the simulation of the dynamics of such systems presents specific problems associated with the complex three-dimensional folds adopted by their highly charged polyanionic backbone and by the large variety of naturally modified nucleotides they comprise [1]. It is, thus, timely to present the current state of the art of MD simulations of RNA systems in a comprehensive and practically oriented manner. Several reviews on nucleic acid MD simulations [2–10] already address important methodological issues.

Here, we will focus on how to setup a MD simulation of an RNA system by using explicit representation of solvent and Ewald summation methods for the treatment of the long-range electrostatic interactions. Specific features related to MD simulations of RNA systems using implicit solvation models can be found in [5]. The simulation protocols that will be described in the following refer mainly to our own experience with the AMBER program with which we are most familiar and which is used in many laboratories working on nucleic acids. They should be easily transposable to other MD packages.

34.2 MD Methods

Common force field-based simulation methods make use of an empirical potential energy function and parameter set to describe the physical properties of the system to be studied. In MD simulations, the “possible” time evolution of the system starting from an initial set of coordinates and velocities is calculated by integrating Newton’s equations of motion over a certain period of time [11]. Descriptions of the integration algorithms associated with such programs can be found in [12–
Tab. 34.1. List of recent MD simulations of RNA systems (up to December 2003) using an explicit representation of the solvent and Ewald summation methods for the treatment of the long-range electrostatic interactions (simulations using truncation [107–110] or density functional methods [39] are not listed in the table).

<table>
<thead>
<tr>
<th>Starting structures</th>
<th>nt</th>
<th>Length (ns)</th>
<th>Ions</th>
<th>Modified nt</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNA.RNA duplex</td>
<td>12</td>
<td>2.5</td>
<td>Na⁺</td>
<td>PNA</td>
<td>111</td>
</tr>
<tr>
<td>HNA.RNA duplex</td>
<td>16</td>
<td>1.1</td>
<td>Na⁺</td>
<td>HNA</td>
<td>112</td>
</tr>
<tr>
<td>Model helix r{(ApU)₁₂}₂</td>
<td>48</td>
<td>2.4</td>
<td>~0.2 M KCl</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Model helix r{(CpG)₁₂}₂</td>
<td>48</td>
<td>2.4</td>
<td>~0.2 M KCl</td>
<td>19, 20, 22, 85</td>
<td></td>
</tr>
<tr>
<td>Model helix r{(CmpGm)₁₂}₂</td>
<td>48</td>
<td>4.4</td>
<td>~0.2 M KCl</td>
<td>Cm; Gm</td>
<td>22</td>
</tr>
<tr>
<td>r{(CCAACGUUGG)₁₂}</td>
<td>20</td>
<td>2.0</td>
<td>Na⁺</td>
<td>21, 83</td>
<td></td>
</tr>
<tr>
<td>r{(CCAACGUUGG)₁₂}</td>
<td>20</td>
<td>1.3</td>
<td>Na⁺</td>
<td>MOE</td>
<td>113</td>
</tr>
<tr>
<td>r{(CGGGCG)₁₂}</td>
<td>12</td>
<td>0.7</td>
<td>Na⁺</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>r{(GAGUACUC)₁₂}</td>
<td>16</td>
<td>5.0</td>
<td>~0.3 and ~1.0 M NaCl</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>r{(GCGAGUACUCGCG)₁₂}</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r{(CGCGAUCGCG)₁₂}</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r{(CCUUCGAAGG)₁₂}</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNRA tetraloop hairpins</td>
<td>26</td>
<td>3.0</td>
<td>~0.1 M NaCl</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.4</td>
<td>Na⁺</td>
<td>I</td>
<td>115</td>
</tr>
<tr>
<td>U1A hairpin</td>
<td>21</td>
<td>5.0</td>
<td>NaCl</td>
<td></td>
<td>116</td>
</tr>
<tr>
<td>Human U4 snRNA</td>
<td>62</td>
<td>3.0</td>
<td>K⁺</td>
<td></td>
<td>117</td>
</tr>
<tr>
<td><strong>X-ray</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ApU and GpC steps (in crystal)</td>
<td>8</td>
<td>2.0</td>
<td>Na⁺</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>r{(CmGmCmGmCmGm)₁₂}</td>
<td>12</td>
<td>0.7</td>
<td>Na⁺</td>
<td>Cm; Gm</td>
<td>23</td>
</tr>
<tr>
<td>r{(GGACUUGGUGC)₁₂}</td>
<td>24</td>
<td>4.0</td>
<td>~0.1 M NaCl</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>r{(UAAGGAGGUGAU)₁₂}</td>
<td>24</td>
<td>5.0</td>
<td>~0.3 and ~1.0 M NaCl</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>tRNA^{A¹p}</td>
<td>76</td>
<td>0.5</td>
<td>NH₄⁺</td>
<td>D; Ψ; m¹G; m³C; m³U</td>
<td>38, 89, 91–94, 119</td>
</tr>
<tr>
<td>tRNA^{A¹p} anticondon hairpin</td>
<td>17</td>
<td>0.5</td>
<td>NH₄⁺</td>
<td>Ψ; m¹G</td>
<td>38, 89, 91–94, 119</td>
</tr>
<tr>
<td>SS rRNA loop E</td>
<td>24</td>
<td>10.0</td>
<td>Na⁺ &amp; Mg²⁺</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>11.5</td>
<td>~0.2 and ~1.0 M KCl and Mg²⁺</td>
<td>38, 48, 49</td>
<td></td>
</tr>
<tr>
<td>HIV kissing loop complexes</td>
<td>46</td>
<td>7.5</td>
<td>Na⁺ and Mg²⁺</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>16S rRNA</td>
<td>81</td>
<td>5.5</td>
<td>~0.1 M NaCl</td>
<td></td>
<td>120</td>
</tr>
<tr>
<td>Pseudoknot</td>
<td>26</td>
<td>5.0</td>
<td>Na⁺</td>
<td>C⁺</td>
<td>34</td>
</tr>
<tr>
<td>Hammerhead ribozyme</td>
<td>41</td>
<td>1.1</td>
<td>Na⁺  and Mg²⁺</td>
<td>121, 122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>0.8</td>
<td>~0.1 M NaCl and Mg²⁺</td>
<td>51, 123</td>
<td></td>
</tr>
<tr>
<td><strong>RNA–ligand (X-ray)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16S rRNA site A/neomycin B</td>
<td>21</td>
<td>10.0</td>
<td>Na⁺</td>
<td></td>
<td>124</td>
</tr>
</tbody>
</table>
These are common to all current simulation packages and will not be further discussed here. Various methods having the goal of pushing back currently accessible timescale limits by using multiple time-step integrators are currently under development [12, 13, 16–18].

### 34.3 Simulation Setups

#### 34.3.1 Choosing the Starting Structure

Starting structures are mainly derived from (1) model built systems, (2) low- to high-resolution X-ray data and (3) NMR data.
34.3.1 Model Built Structures
Such structures can be used to study precise issues related to (1) the conformation and solvation of simple RNA building blocks such as Watson–Crick base pairs and base pair steps [19, 20], (2) the differences between RNA and DNA structures of similar sequence [19–21] or (3) the effects associated with the introduction of natural [22, 23] and non-natural modified nucleotides [24, 25]. In order to generate larger systems, model built fragments can also be assembled with motifs derived from X-ray or NMR structures [26].

34.3.1.2 X-ray Structures
When derived from moderate- to high-resolution data, X-ray experiments certainly provide the most accurate source of starting structures since, besides precise coordinates for the solute, they also supply important information related to its hydration and ionic environment [27, 28]. The main sources for X-ray structures are the NDB (Nucleic acid Database; http://ndbserver.rutgers.edu/ [29]) and the PDB (Protein Data Bank; http://www.rcsb.org/pdb/ [30]).

34.3.1.3 NMR Structures
NMR structures are similarly deposited in the NDB and PDB. Due to the lower amount of detail contained in these structures compared to those derived from X-ray data, they are generally chosen only when no X-ray data for a given system are available. Furthermore, NMR cannot provide data related to the hydration and ionic environment of the solute.

34.3.2 Checking the Starting Structure
The imprecision associated with model built and NMR structures is usually well characterized. Whereas X-ray structures are generally very precise, these may nonetheless contain local errors which, if not noticed and corrected, may seriously impair the quality of the calculated trajectories which greatly depends on the correctness of the starting structure.

34.3.2.1 Conformational Checks
For proteins, the precise orientation of several residues (essentially His, Asn and Glu) cannot be correctly assessed by a single examination of the electron density maps obtained at resolutions above around 1.0 Å. Some groups have addressed this problem [31] and have proposed various automated solutions based on an examination of “contact maps” (see, e.g. MolProbity; http://kinemage.biochem.duke.edu/ [32]) for detecting incorrect or suspicious orientations of amino acid residues in existing experimental structures. Checking the orientation of such residues is therefore essential if one desires to perform accurate simulations of protein–RNA complexes. In nucleic acids, the orientation of the natural nucleotides can generally be determined without ambiguity even at low resolution. However, this is not
the case for the orientation of the pseudouridine or ψ residue which is a naturally modified nucleotide [33] and possibly other side-chains in modified nucleotides.

34.3.2.2 Protonation Issues
It is now recognized that some nucleic acid residues such as adenines or cytosines can be found in their protonated state [34]. Yet, X-ray structures rarely provide direct evidence for such protonation states. These have to be derived from chemical intuition by inspecting the structural environment of the involved residues. In cases where it is impossible to decide which protonation state is the most probable, it is advisable to perform simulations of both the charged and uncharged states [34]. Evidently, protonation states of charged amino acids and of drugs such as aminoglycosides [35, 36] have also to be clearly defined before starting a simulation.

34.3.2.3 Solvent
The interpretation of solvent density maps derived from X-ray experiments is often difficult. A recent survey of the NDB has revealed that it is not unusual for water molecules to be assigned to sites with very large electron densities (above those expected for a water molecule); this may be indicative of the presence of a metal ion at those locations [37]. It has also recently been emphasized that some density spots resulting from the presence of anions (SO$_4$$^{2-}$, Cl$^{-}$, etc.) in the vicinity of nucleic acids have been wrongly assigned to divalent cations (Mg$^{2+}$ or Mn$^{2+}$) [38]. Incorrectly assigned metal ions may represent an important source of errors in MD simulations.

34.3.3 Adding Hydrogen Atoms
Most of the structures deposited in the NDB or PDB are devoid of hydrogen atoms. Hence, before starting a simulation, it is essential to add them either by using the tools provided by the main MD packages or by using external programs such as MolProbity (http://kinemage.biochem.duke.edu/ [32]).

34.3.4 Choosing the Environment (Crystal, Liquid) and Ions
Two types of environment may be chosen: the crystal and the liquid phase. Simulations of the crystal phase are rare for nucleic acids [39–43] and only the r(CpG) and the r(ApU) dinucleotides have been simulated in this environment [39, 44] even though such simulations probably provide the most accurate confrontation between experimental and theoretical data [3]. Simulations of RNA systems in the aqueous phase are the most common. The ions used to neutralize the nucleic acid can be of various types (Na$^{+}$, K$^{+}$, NH$_4^+$, Mg$^{2+}$) and different ionic conditions ranging from a number of ions sufficient to neutralize the charge carried by the nucleotidic backbone to a number of ions that would appropriately represent
a solution containing 0.1–1.0 M of added salt can be selected (Table 34.1). In this last case, co-ions (Cl\(^-\), SO\(_4^{2-}\), etc.) must be placed in the simulation box as well.

34.3.5

**Setting the Box Size and Placing the Ions**

34.3.5.1 **Box Size**

For simulations of the crystal phase, the box size is determined by the size of the crystal cell. In solution, the box size is usually allowed to extend by approximately 8–12 Å around the solute, thus ensuring three to four hydration layers in each direction.

34.3.5.2 **Monovalent Ions**

Monovalent ions are then placed in the simulation box by replacing the water molecules with the lowest (for cations) and highest (for anions) electrostatic potential \[19, 45\]. If one wants to avoid biases due to the initial placement of ions, it is also essential to set limits ensuring that no ions are closer than a certain distance, usually 5–6 Å, to any nucleic acid atom or to another ion. In this manner, no ions are initially found in the nucleic acid grooves, in direct proximity of the phosphate groups or forming contact ion pairs.

34.3.5.3 **Divalent Ions**

The limited length of the simulations does not allow reproduction of dehydration processes of Mg\(^{2+}\) ions and formation of direct Mg\(^{2+}\)•••O contacts with the RNA which take microseconds to occur \[46, 47\]. Thus, divalent ions cannot be placed in the simulation box as described above. Such systems can only be realistically simulated if the Mg\(^{2+}\) ions occupy their crystallographic positions already at the beginning of the MD run \[48–52\]. It has to be noted that, although the Mg\(^{2+}\) ions occupy their crystallographic positions, the water molecules located in the first hydration shell of the ions find their equilibrium positions during the equilibration phase or, in other words, the crystallographic position of these water molecules is not taken into account.

34.3.6

**Choosing the Program and Force Field**

34.3.6.1 **Programs**

Two very widely used MD simulation packages are AMBER (Assisted Model Building with Energy Refinement; http://amber.scripps.edu/ \[53\]) and CHARMM (Chemistry at HARward Molecular Mechanics; http://yuri.harvard.edu/). Each of these is associated with its own force field. Other more recent and non-commercial MD packages such as GROMOS (http://www.igc.ethz.ch/gromos/ \[54\]) and NAMD (http://www.ks.uiuc.edu/Research/namd/ \[55\]) are also available. Package choice is a very subjective matter and is generally dictated by the lab history. However, one
should be attentive to the fact that some desired features may not be present in all of these packages.

34.3.6.2 Force Fields
AMBER [56] and CHARMM [57, 58] are the most commonly used force fields for simulating the dynamics of nucleic acids and proteins. They co-exist with the less frequently used BMS [59] and OPLS force fields [60]. However, it is essential to be aware of their limitations. Although they are all able to reproduce quite well most of the important structural and dynamical features of biological systems, they do differ in some specific aspects. Comparisons of the AMBER and CHARMM force fields performed on DNA duplexes [61, 62] have led to the conclusion that in some instances a “force field dependent polymorphism” [3] is observed. This polymorphism contributes to an artificial drift of the duplexes toward an A- or a B-form that is independent of the starting conformation. Some of these problems have been corrected in recent versions of the AMBER and CHARMM force fields [63, 64]. Hence, since these force fields are constantly evolving, it is recommended to check for the latest available updates. Nevertheless, it can be concluded that up to now, none of the widely used force fields is able to reproduce fully all of the structural features experimentally observed for nucleic acids in solution [8, 64]. New, polarizable force fields [65] such as those recently implemented in AMBER are expected to bring significant improvement in the informative and predictive power of MD simulations (although at the expense of the use of more computer resources). These force fields are still at an initial stage of development and will have first to be thoroughly tested on small systems before using them for simulating solvated macromolecules.

34.3.6.3 Parameterization of Modified Nucleotides and Ligands
Most empirical force fields used in common MD packages are equipped with parameters for modeling all canonical nucleotides and amino acids, but not for modified nucleotides and ligands. For simulation of such systems, new parameter sets must be developed. Although some general sets of procedures for automated parameterization based on fitting to both experimental and quantum chemical data have been proposed, force field parameterization remains a matter for experts. A classical review on force field parameterization has been written by Dinur and Hagler [66]. Fortunately some tools and methods for generating reasonable to high-quality force field parameters are available [67, 68]. Among them it is worth mentioning the AMBER module Antechamber which greatly simplifies the generation of new parameters for the AMBER force field and the Automated Frequency Matching Methodology (AFMM) developed for the CHARMM force field but extendible to all other atom based force fields [69]. AFMM allows development of parameter sets for small to medium sized molecules using high quality quantum chemical calculations as reference data. A new set of scripts (RED) has also been made available to facilitate the derivation of partial charge sets from electrostatic potential calculations (http://www.u-picardie.fr/labo/lbpd/RED/). In addition to those for nucleic acids and amino acids, some force fields have been developed
for specific classes of molecules, such as carbohydrates (OPLS-AA [70]; GLYCAM http://glycam.ccr.crc.uga.edu/), sulfates and sulfamates [71] or polyphosphates [72]. Parameters for the ions which are compatible with the chosen force field can be easily retrieved from the literature.

34.3.6.4 Water Models
Among the available water models, the TIP3P (Transferable Interatomic Potential – 3 Point [73]) model is generally associated with the AMBER force field. The more computationally “expensive” four (TIP4P [73]) and five (TIP5P [74]) point models have been developed and tested on various systems. SPC (Single Point Charge [75]) and SPC/E (SPC/Extended [76]) models are also very popular. For DNA, it has been shown that the use of TIP3P and SPC/E water models resulted in comparable hydration patterns although the water densities associated with the water model with the highest diffusion rate (TIP3P) appear to be more blurred than those calculated with the SPC/E model that has calculated diffusion rates in better agreement with experimental data [10]. Besides, a MD simulation of a protein in a crystal environment has led to the conclusion that results obtained with the SPC/E model are in much better agreement with neutron-scattering data than those collected with the TIP3P model [77]. Thus, a particular water model can significantly alter the results from MD simulations and more studies are needed to precisely evaluate their influence.

34.3.7 Treatment of Electrostatic Interactions

The ensemble of parameters in the above referenced force fields is usually considered to be the most important factor in determining the quality of a MD trajectory. Yet, if the empirical equation that models in a condensed way the intermolecular forces at play in biomolecular systems is not accurate enough or if one of its terms is not correctly evaluated then even the most precisely developed set of parameters cannot realistically account for the dynamics of these systems. Indeed, numerous studies have revealed that severe artifacts emerge when the long-range electrostatic contributions are neglected. In other words, when only a part of the Coulomb term \( q_i q_j / 4\pi \varepsilon_0 r_{ij} \) is estimated [78–80]. It has been shown that an accurate treatment of the long-range electrostatic interactions is of paramount importance for generating realistic trajectories of nucleic acid systems. The particle mesh Ewald (PME) summation method [81, 82] has proven to lead to very stable nucleic acid trajectories [80, 83].

The program default parameters are generally adequate for using the PME method. These parameters are such that a cubic interpolation scheme and a \( 10^{-5} \) Å tolerance for the direct space sum cutoff is chosen. To speed up the fast Fourier transform in the calculation of the reciprocal sum, the size of the PME charge grid is chosen to be an integer power of 2, 3 and 5, and to be slightly larger than the size of the periodic box. This leads to a grid spacing of around 1 Å or less.
34.3.8 Other Simulation Parameters

34.3.8.1 Thermodynamic Ensemble
MD simulations can be conducted in a microcanonical ensemble or (N,V,E) for constant number of particles, volume and energy. Yet, the isothermal–isobaric or (N,P,T) ensemble for constant number of particles, pressure and temperature is more commonly chosen. Working at constant pressure is particularly important during the equilibration stage since during the building stage of the system, the water molecule positions are not optimized at the solute/solvent interface. This can lead to the generation of holes in the simulation box when constant volume options are selected.

34.3.8.2 Temperature and Pressure
In order to maintain a constant temperature and pressure, the Berendsen temperature coupling scheme and an isotropic molecule based constant pressure scaling with a time constant of 0.2–0.5 ps for both are often used [84], although other coupling schemes and time constants are found in the literature.

The target temperature and pressure values are usually set to 1 atm and 298 K or 25 °C, also called “room temperature”. At this point, it is interesting to note that the choice of this target temperature is dictated by early studies which used experimental data obtained at room temperature in order to parameterize the force fields. This is close to the temperature at which many \textit{in vitro} experiments are performed in a laboratory. Nonetheless, simulations using different target temperatures (e.g. 5, 25 and 37 °C) have been used to reveal some differences in the hydration of RNA systems [85].

34.3.8.3 Shake, Time Steps and Update of the Non-bonded Pair List
A tolerance of 0.0005 Å is generally used for the SHAKE algorithm [86] that allows to “safely” utilize a 2-ps time step (instead of 1 ps) by artificially freezing the most rapid vibration motions (C–H, O–H, N–H, etc., elongations) which can be observed in biomolecular systems. A strong tolerance value has been found to reduce the so-called “flying ice cube” phenomena described below [87, 88]. Another time saving device is associated with the use of a pair list for calculating non-bonded interactions between all atoms in the system. This pair list is not calculated at each time step, which would be computationally expensive, but only once every \textit{n}th step (usually every 10 steps).

34.3.8.4 The Flying Ice Cube Problem
Since the non-bonded pair list is not updated every step (in order to save CPU resources) and constant pressure algorithms as well as uniform scaling of velocities by the Berendsen coupling scheme are used, some small energy drain during the simulation can occur and the center of mass velocity can slowly grow. Therefore, periodically in the simulation this center of mass velocity has to be removed otherwise the “flying ice cube” syndrome may appear [87, 88]. It is also advised to update the non-bonded pair list more frequently [87, 88].
34.3.9 Equilibration

The main aim of equilibration procedures is to alleviate tensions created in the system during the early building stages. These tensions would, if neglected, lead to unrealistic trajectories. In the following, a typical equilibration protocol used by us will be described.

First, 500 steps of steepest descent minimization are applied to the entire system with periodic boundary conditions in the (N,P,T) ensemble. This is followed by 25 ps of simulation in the same thermodynamic ensemble where only the water molecules and the hydrogen atoms of the solute are allowed to move (the heavy atoms of the solute and the ions are frozen by using the BELLY option in AMBER). During the next 25 ps, the constraints placed on the monovalent ions are released so that they can start to equilibrate around the RNA system while the BELLY option is still used to freeze the heavy atoms of the solute. Then, several rounds of 50-ps MD are performed during which positional constraints of 10, 5, 2, 1, 0.5, 0.1, 0.01 and 0.001 kcal/mol Å² are applied to the heavy atoms of the solute, yielding a partially constrained 450 ps MD trajectory. Then the production phase can take place. However, in order to allow a better sampling of the conformational space by the solute and the ions, the next 0.5–1.0 ns are generally included in the equilibration phase [19].

The length of the equilibration phase can be extended or shortened at will. However, it has to be noted that the phase during which constraints are applied ensures that the solvent is appropriately equilibrated in the vicinity of the solute, but also throughout the simulation box. The part during which no constraints are applied allows then for the solute to relax in the presence of relatively well-equilibrated solvent environment. Other equilibration protocols start at low temperature (50 K); then, the temperature is raised to the target temperature at discrete time intervals [51]. An apparent advantage of the former procedure is that the sampling of the conformational space by the solvent is conducted at the target temperature and, hence, probably more efficient.

34.3.10 Sampling

34.3.10.1 How Long Should a Simulation Be?

Ideally, MD simulations should be long enough to sample all the conformational transitions occurring in a particular biomolecular system. Given present and predictable computational means, this goal will certainly not be achieved in the next decade. It is, thus, impossible to address in the near future biochemical processes with very long relaxation times by using classical MD simulations with explicit solvent representations. Up to now, the longest MD simulations of RNA systems have barely overcome the 10-ns timescale (Table 34.1). Nevertheless, many processes associated with limited conformational perturbations lie within reach of present methods. Among them, those that involve the binding of the smallest ligands to nucleic acids (monovalent ions and water molecules [19, 20, 48]) or the orientation...
of the 2′-OH hydroxyl groups [89] can trustfully be evaluated on the current nanosecond timescale.

34.3.10.2 When to Stop a Simulation

Another important and often overlooked issue is: when to stop a MD simulation or when does it start to be a waste of time to continue sampling? Indisputably, a very large number of approximations are included in MD simulations. These approximations are, among others, related to: (1) imprecision in the force field parameters, (2) incomplete evaluation of the intermolecular forces, (3) neglect of polarization and charge transfer effects and (4) an incomplete knowledge of the starting conditions. This last point by itself can easily lead to the generation of partially or totally unrealistic simulations. For example, an instability of the tertiary core structure of the yeast tRNAAsp molecule characterized by a reordering of several base triples has been observed during a 500-ps MD simulation [90]. This has first been attributed to the absence of Mg$^{2+}$ ions in the model. Later on, it was proposed that an adenine involved in the above-mentioned base triples is protonated [34]. Probably both effects led to the calculated transitions and further sampling would, in this case, not have contributed to a better understanding of the structure and the dynamics of this molecule. Hence, there is a certain risk for long MD simulations to oversample regions of the “configurational space” attached to the current force field and MD package that are not overlapping with the “true” configurational space explored by the investigated system.

34.3.10.3 Multiple MD (MMD) Simulations

An alternative to long MD simulations is to generate an ensemble of several “shorter” trajectories by using the MMD simulation technique. This method introduces slight perturbations (such as different initial velocities) in the starting conditions and, thus, exploits the chaotic nature of MD simulations in order to generate several uncorrelated trajectories [91–97]. It has been stated that several trajectories generated from similar but slightly different starting conditions may provide, on a statistical basis, more information than a single long trajectory. A parallel view developed by other authors consists of generating and comparing several trajectories obtained by starting from different initial configurations [96]. In short, MMD methods allow a statistical evaluation of MD simulation that cannot be achieved from the analysis of single trajectories [80].

34.4 Analysis

34.4.1 Evaluating the Quality of the Trajectories

Before trying to extract information from MD simulations, it is imperative to evaluate the quality of the generated trajectories by checking their internal consistency and confronting them to all available experimental data [98].
34.4.1 Consistency Checks

Internal consistency checks are related to the detection of unphysical behaviors such as those resulting from the use of truncation methods for the evaluation of the electrostatic interactions [78, 79] or “flying ice cube” syndromes. When such issues come to the foreground, it is clear that the simulation can and should no longer be used to derive biologically relevant information but that the focus of the study must shift toward the correction of the detected artifacts. By doing so, the importance of the neglected contributions is generally brought to light. For example, the stabilizing role of “hydration forces” could be assessed by comparing trajectories issued from simulations including or discarding long-range electrostatic interactions [3, 80, 92].

34.4.1.2 Comparison with Experimental Data

Confrontation with available experimental data (that are most frequently of a structural type) is in all cases mandatory. For instance, if important tertiary interactions (Watson–Crick base pairing, etc.) present in the starting structure break during the simulation over short timescales it is probable that the simulation protocols are not adequate and the reasons underlying such behavior must be uncovered [3, 79]. Early simulations performed in “in vacuo” conditions or with truncation of the electrostatic interactions where disruptions of important tertiary interactions were observed clearly illustrate this point [79, 99]. Moreover, if the structure of the investigated system diverges significantly from the initial structure, in part or entirely, during the time course of the simulation, it must be considered that this observation may result from force field inaccuracies. Only comparisons with experimental data on the same or related systems can help to resolve such issues.

34.4.1.3 Visualization

Visual checks are also mandatory and can be performed by using the programs that are delivered with each main MD package of by using programs such as VMD (Visual Molecular Dynamics; http://www.ks.uiuc.edu/Research/vmd/) that can read most of the available MD outputs. In rare cases, the Open BABEL utility (http://openbabel.sourceforge.net/) is needed in order to convert data written in different formats.

34.4.2 Convergence Issues

Afterwards, the convergence of the simulation with respect to the investigated properties has to be assessed. For this, it is necessary for the simulation length to exceed the average relaxation time of the investigated property which is not necessarily the longest relaxation time of the system that can largely exceed the currently accessible nanosecond timescales [98]. For example, it is possible to estimate on a statistical basis the residence times of water molecules (between 10 and 500 ps) around a regular RNA helix from nanosecond MD simulations, although the statistics will obviously be less good for the water molecules with the longest residence times [19, 20]. For less regular RNA systems, water molecules may be trapped into
specific pockets for much longer times and only lower bounds for the residence times can be obtained [100]. A similar problem is associated with the estimation of the binding properties of monovalent and divalent ions. While it is feasible to estimate the exchange time of monovalent ions located in RNA grooves from nanosecond MD simulations [49], it is impossible on such timescales to estimate the same values for divalent ions [48].

34.4.3 Conformational Parameters

Most of the analysis programs delivered with major MD packages (such as CARNAL for AMBER) can monitor nearly all of the main conformational parameters such as interatomic distances, angles and dihedral angles, and also parameters that are specific to nucleic acids such as sugar puckers or user defined variables. Other programs such as CURVES (http://www.ibpc.fr/UPR9080/Curindex.html) [101] or 3DNA (http://rutchem.rutgers.edu/%7Exiangjun/3DNA/) [102] that allow to estimate parameters specific to nucleic acids such as bending, shift, slide, rise, tilt, roll, twist, etc., can be also very useful in the analysis process. The majority of these parameters are now attached to the experimental structures deposited at the NDB.

34.4.4 Solvent Analysis

Solvent analysis is an important aspect of the evaluation of MD trajectories. It is related to the characterization of water and ion binding sites and to the estimation of their residence times. Several reviews have already addressed these issues for nucleic acid systems [100, 103–106] and we will not discuss them in further details.

34.5 Perspectives

The current simulation protocols are now able to deliver stable trajectories of RNA systems ranging from small duplexes to large RNA–protein complexes over nanosecond timescales. Important conformational transition and solvation processes can be, thus, studied in detail. The current trend is to head toward longer timescales. This would allow us to address, among others, folding and unfolding issues. Nevertheless, one should be aware that by extending the currently accessible timescales, new artifacts that remain up to now hidden will appear and necessitate the development of more refined force fields and simulation methods that will most evidently incorporate polarizability and charge transfer effects. The detection and correction of artifacts that will manifest themselves on the longer timescales must be regarded as an important duty of modelers.
Acknowledgments

A. C. V. acknowledges support from the European contract HPRN-CT2002-00190 (CARBONA).

References

30. H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne,
34 Molecular Dynamics Simulations of RNA Systems

36 D. S. Pilch, M. Kaul, C. M. Barbieri, J. E. Kerrigan, Biopolymers 2003, 70, 58–79.
References

100 P. Auffinger, E. Westhof, Biopolymers 2000, 56, 266–274.