A MESOSCOPIC MODEL FOR MOLECULAR DYNAMICS STUDIES OF RNA NANOSTRUCTURES

Maxim Paliy
M^2NeTLab, Wilfrid Laurier University,
75 University Avenue West, Waterloo, ON, Canada, N2L 3C5
Homepage: http://www.m2netlab.wlu.ca, Email:mpaliy@wlu.ca

Roderick Melnik
M^2NeTLab, Wilfrid Laurier University,
75 University Avenue West, Waterloo, ON, Canada, N2L 3C5
Homepage: http://www.m2netlab.wlu.ca, Email:rmelnik@wlu.ca

Bruce Shapiro
Center for Cancer Research Nanobiology Program
National Cancer Institute, Frederick, MD 21702
Email: bshapiro@ncifcrf.gov

A series of 3-beads-per-nucleotide mesoscopic models is being developed for the study of the RNA nanostructures via a Molecular Dynamics simulation. Such coarse-grained treatment allows us to reach the microsecond time scale that is by three orders of magnitude larger than that achieved in the full atomistic computer simulations, and thus, to study the slowest conformational motions of the RNA, as well as to enable simulations of the larger RNA structures in the context of bionanotechnology. We find that the variant of the model described by a set of just a few universal parameters is able to describe different RNA conformations and is comparable in structural precision to the model variant where the detailed values of the backbone P-C4' dihedrals taken from a reference structure are included. Our findings demonstrate the importance of the RNA conformation classes based on these dihedrals.

1. INTRODUCTION

The development of coarse-grained description of RNA is of paramount importance, in view of its prominent role in Nature, as well as its promising future applications in bionanotechnology [1, 2]. To date, the most detailed and informative method for the *in silico* study of RNA and other biomolecules remains the full-atom Molecular Dynamics (MD) simulation [3]. Unfortunately, time scales that can be achieved in the present-day MD amount to a few (tens) nanoseconds only, which is by many orders of magnitude less than the duration of the processes occuring in biolomecules (where the evolution times span from micro- to milliseconds). In order to reach such scales in the simulations, one needs to consider a coarse-grained (CG) treatment, where, typically, the groups of neighbouring atoms are represented by the CG interaction centres - "beads", and effective interactions between such beads are tweaked in the way to fit the molecule's atomic connectivity, thermal and mechanical properties *etc.* Two sources of data are often used in the fitting process: (i) the experimentally available structural information as well as other known properties of interest (which can be limited and/or incomplete), and (ii) a host of very detailed atomistic data obtained in the full MD simulations. Namely, the parameters for a CG model can be derived from both experimental and full-atom MD data by Boltzmann Inversion (BI) [4] of the Radial Distribution Functions (RDFs), or, in the case of full MD simulation only, with the 'force matching' method [5] (for some recent approaches see e.g. [6, 7]). Finally, the behaviour of such a CG model is further investigated using the coarse-grained Molecular Dynamics (CGMD), which now allows to reach much longer time scales (although the dynamics of such a model does not always represent the dynamics of the original system sufficiently well [6]).

Meanwhile, most recently RNAs has attracted the attention of bionanotechnologists in the context of so-called "RNA architectonics" — a set of

recipes for self-assembly of the RNA nanostructures of the arbitrary size and shape [1, 2]. Smallest RNA building blocks, such as "right angle motif" [1], 'kink-turn motif' [2, 8] or "RNAIi/RNAIIi complex" [9] were manipulated (either experimentally [1, 2] or via computer modelling [9]) into the desired 2D or 3D nanostructures (squares, hexagons, cubes, tetrahedrons etc) that can be further assembled into periodic or quasiperiodic lattices (an intersting example is a RNA nanotube [9]).

Compared to the DNA, the RNA as a nano-engineering material brings some additional challenges. For example, due to the specificity of the interactions in RNA (e.g. noticeable presence of the non-Watson-Crick base pairing) it shows much larger structural modularity and diversity of the tertiary building blocks, ~ 200 versus ~ 20 for DNA [2]. Its structural diversity makes it more difficult to model the RNA on a coarse-grained level as well. Along this direction, a review [10] and some recent attempts [11–13, 15, 14] can be mentioned. The main challenge is to represent the RNA on a coarse-grained level sufficiently accurately with just a few "universal" parameters, thus adopting the strategy of a "CG forcefield". This proved to be a difficult task. Instead, typically, a lot of detailed structural information (such as equilibrium values of bonds, angles, dihedrals, nonbonded interatomic distances from the experimently resolved structures) is supplied to the CG model, thus providing its precision in describing a given structure only. Such structurally biased approach is often termed Self Organized Polymer (SOP) [11]. In the present paper we make an initial comparison of the "SOP" approach to a "RNA CG forcefield" approach and attempt to explore the possible avenues from the former to the latter.

2. THE CG MODEL

We use the following full MD data sources to fit the parameters of the presented CG model: first, a 6 ns 300 K trajectory of a simple RNA double A-helix dodecamer (GCGCUUAAGCGC) modelled with the Amber software [21] using Cornell force-field [22]; second, a more complex and bigger RNA nanostructure (13 nm in characteristic size) – hexagon-shaped RNA ring [9] (termed "nanoring" in what follows), that is composed of 6 RNAIi/RNAIIi complexes, joined by the "kissing loop" motifs (Fig. 1). The corners of the hexagon are made with two such interacting "septa-loops" (e.g. AACCAUC loop is paired with UUG-GUAG loop). A sample configuration of one double helix ending with a loop is shown in Fig. 1. In a recent study [17] we generated the full MD trajectories for this structure in the NAMD/VMD package [18, 20] using the CHARMM27 [19] force field. Visualisation and data processing of both full MD and CG simulations is carried with VMD, using in-house developed scripts. The Molecular Dynamics for the CG model is implemented via the DL_POLY 2.19 package [23].

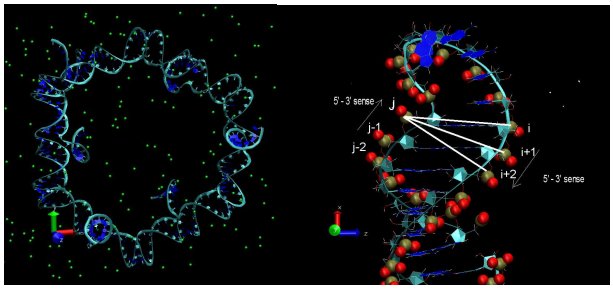

Fig. 1. (color online): Left: The "ribbons" representation of the RNA nanoring in VMD. Right: Detailed view of a part of the double helix and a septa-loop of the RNA nanoring. Phosphate groups are shown with spheres. Layout of the interactions between base pairs in the CG model is shown with white solid lines.

The development of a CG model consists of two major stages: (i) choice of the groups of atoms to be combined in a single CG bead, and (ii) selection of the functional forms and fitting of the parameters for the effective interactions between the beads.

In the case of nucleic acids, the simplest (and the most common up to now) choice for the stage (i) is "one bead per nuclieotide", the beads being normally placed on the phosphate groups. Such a choice allows one to use a host of experimentally available structural data, and at the same time it keeps the model reasonably simple. However, such representation (as our data indicate) may not be sufficiently flexible to represent more complex conformations beyond the simplest double helix (e.g. the "kissing loops"). Furthermore, it has been recently pointed out in [16] that at least two pseudo-torsional angles are required to describe all the available RNA conformations, namely those between the beads placed on the sites of P and C4' atoms (such description is reminis-

cent of the well-known Ramachandran plots for proteins). These findings led us to consider also a two-bead representation. However, the placement of an extra bead at the C4' atomic site turned out to be not particularly suitable in terms of the geometry of base pair bonds drawn between such beads. Instead, a representation with three beads per nucleotide, that correspond to the (P)hosphate, (S)ugar, and nucleic (B)ase, seems to be one of the most natural choices for nucleic acids. That is why we concentrate on the three-beads variant of the model in the present study.

The sample configuration of the RNA nanoring for the three-beads representations of the CG model is depicted in the Fig. 2. First two beads (P) and (S) are placed on P and C4' atoms, while a number of plausible choices is possible for the placement of the third (B) bead (note that for the sake of simplicity we prefer to place beads on the exististing atoms rather then on centres of masses of groups of atoms). We found the following variant to be most convenient: N9 atoms of purines and N1 atoms of pyrimidines. The masses of the beads are taken as $m^{(P)} = 109$ a.m.u. , $m^{(S)} = 120$ a.m.u., $m^{(B)} = 92.5$ a.m.u.

In our model, the beads are organised into several (6 in the case of the nanoring from Fig. 1) backbone chains, which correspond to the basic building blocks of the studied RNA structures. The connectivity inside these units is never broken in the course of simulation (Fig. 2). Besides, all beads interact with each other via the base pair bonding terms (that also maintain the connectivity between the pairs, and thus the secondary structure, during the simulation) as well as via a 'univeral glue' pair potential, which should stabilise the overall structure of the system. The total interaction energy has the following form:

$$V = V_{conn} + V_{bp} + V_{nb}, \qquad (1)$$

where V_{conn} is the energy term that maintains the connectivity within the chains. It is chosen to be of the standard form:

$$V_{conn} = \sum_{c=1}^{C} \Big(\sum_{bonds} V_b(r - r^{(0)}) + \sum_{angles} V_a(\theta - \theta^{(0)}) + \sum_{dihedrals} V_d(\phi - \phi^{(0)}) \Big) \qquad (2)$$

where C is the total number of backbone chains in the nanostructure, and $V_b(r)$, $V_a(\theta)$, $V_d(\phi)$ are intra-chain terms that correspond to the energies of bonds, angles and dihedrals, respectively, while $r^{(0)}$, $\theta^{(0)}$, $\phi^{(0)}$ are the equilibrium values for bonds, angles, and dihedrals. These contributions can include both harmonic and anharmonic (up to quartic) terms.

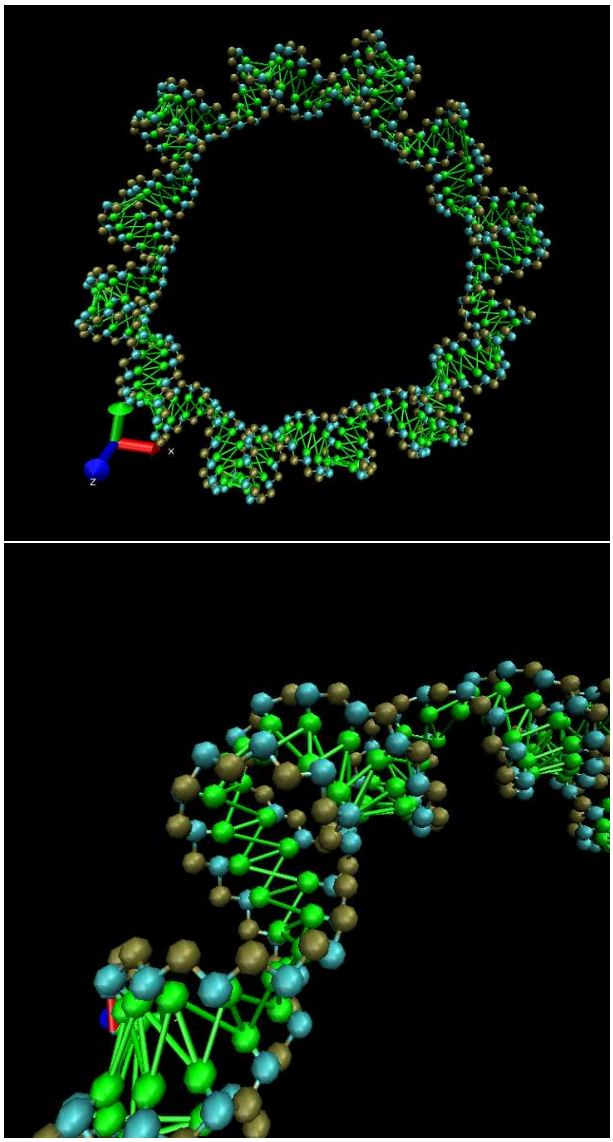

Fig. 2. (color online): Left: 3-beads-per-nucleotide variants of the CG model for the RNA nanoring shown in Fig.1. Right: Zoomed view of one "kissing loop". The phosphate (P) beads are shown in brown, the sugar (S) beads - in cyan, the base beads (B) - in green. The bonding scheme of the backbone and between the bases is shown with the lines.

The energy term V_{bp} accounts for the interactions between the base pairs. In the case of the

nanoring, these include the contributions from the base pairs found inside a single chain, as well as between those septuplets of the base pairs belonging to different chains, that form the "kissing loop" units. It has the following form:

$$V_{bp} = \sum_{i,j\in(basepairs)} u_{i,j}(r_{i,j} - r_{i,j}^{(0)}) + u_{i+1,j}(r_{i+1,j} - r_{i+1,j}^{(0)}) + u_{i+2,j}(r_{i+2,j} - r_{i+2,j}^{(0)}) \quad (3)$$

where all the functions $u(r)$ can include both harmonic and anharmonic terms (corresponding bonds are shown in Fig. 2 with green lines). Following the idea of [7], we express the base pair interaction with three bonds per base pair instead of one (see Fig. 1), which brings enhanced structural accuracy and stability, because such treatment allows us to take into account both the hydrogen bonding between bases i and j, as well as the stacking interaction.

The remaining energy contribution V_{nb} corresponds to the interactions between all bead pairs not involved in the bonded interactions described above. It has the following form:

$$V_{nb} = \sum_{i,j\in(nonbonded)} v(r_{ij}). \quad (4)$$

In the current version of the model we restrict ourselves only with the simplest repulsive (WCA) analytical form of the nonbonded potential $v(r_{ij})$, expressing the steric repulsion between the beads via an energy ε and the bead radius σ.

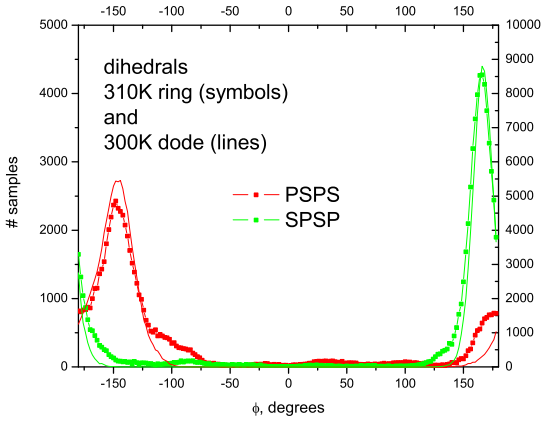

Fig. 3. (color online): Distributions for PSPS and SPSP dihedrals from full atom MD runs. The data for the RNA nanoring are shown with symbols, while those for the RNA dodecamer are shown with thin lines.

3. FITTING THE PARAMETERS OF THE CG MODEL

The total energy of the CG model, Eq. (1) thus contains a number of parameters such as the strengths of the bond, angular, dihedral, base pair terms, as well as those describing the nonbonded interactions. Our general approach to the fitting of these parameteres is based on two sorts of information that is extracted from the full MD runs: (i) distributions of the values of bonds, angles and dihedrals, as well as the RDFs between various sorts of atom pairs are used to fit the CG model parameters via BI method [4]; (ii) atomistic forces are used for the "force matching method" [5].

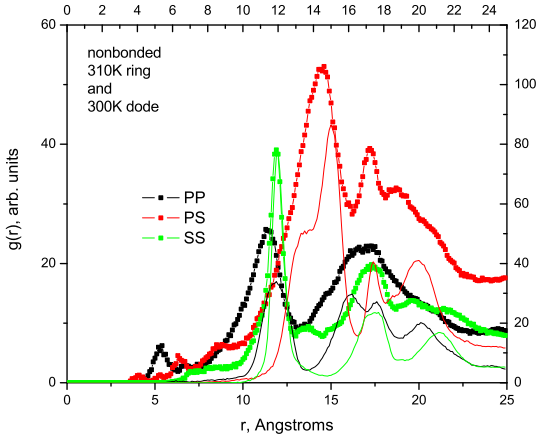

Fig. 4. (color online): Radial Distribution Functions for nonbonded interactions (PP, PS, SS pairs) from full MD runs. The data for the RNA nanoring are shown with symbols, while those for the RNA dodecamer are shown with thin lines.

BI method consists of the following: given the probablity distribution function $P(q)$ for a degree of freedom q it determines the corresponding potential of mean force (PMF) $V_{eff}(q)$ via the following formula:

$$V_{eff}(q) = -k_B T \ln(P(q)). \quad (5)$$

However, thus obtained $V_{eff}(q)$ coincides with the true potential energy only for the case of a single degree of freedom q, and generally it can serve only as a first approximation used in a subsequent iterative

procedure, e.g. of the following kind:

$$V_{eff}^{(i+1)}(q) = V_{eff}^{(i)}(q) - k_B T \ln\left(\frac{P^{(i)}(q)}{P(q)}\right), \qquad (6)$$

where $P^{(i)}(q)$ is the distribution for the variable q obtained in the *coarse-grained* simulation with $V_{eff}^{(i)}(q)$, and the corrections are introduced until the target $P(q)$ is reproduced sufficiently well in CG simulation. This method may not always be successfull, because many variables have to be fitted this way simultaneously, but fortunately usually the fitted energy parameters show certain hierarchy, which allows their refinement in succession, in order of their decreasing strength. Typically [4], such hierachy shows up in the following way $V_{bond} \rightarrow V_{angle} \rightarrow V_{vdw} \rightarrow V_{dihedral}$, though in the case of RNA CG model the order of the terms is not so evident. Therefore, for further refinement of the model, we plan to resort also to a more systematic fitting procedure, involving the force matching method, in particular.

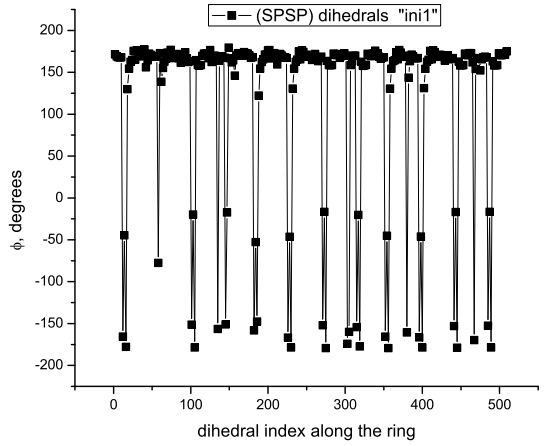

Fig. 5. (color online): The SPSP dihedral angles along the RNA nanoring.

4. RESULTS

The full list of different energy terms for the CG model thus includes the following (the layout of the bonding terms is shown in Fig. 2). (i) Along the nucleic backbone the (PS), (SP), (SB) bonds between the nearest neighbours, as well as (PSP), (SPS), (PSB), (BSP) angles, and (PSPS) and (SPSP) dihedrals are included. (ii) Along the base-paired parts of double helices and kissing loops the $(BB)_{i,j}$, $(BB)_{i+1,j}$, and $(BB)_{i+2,j}$ bonds are included. Besides, due to the topology of the 3-beads nucleic backbone, we introduce the dummy "zero energy" bonds between the next-nearest (SB) neighbours and the nearest (BB) neigbours along the backbone in order to exclude them from the nonbonded interactions, since these bonds are already fixed by the above-mentioned set of backbone terms. The histograms of a few representative degrees of freedom from the full set (namely the PSPS ands SPSP dihedrals) extracted from the full MD runs are plotted in Fig. 3 for both the RNA nanoring and the RNA dodecamer in comparison.

Our model also includes 6 different non-bonded bead pairings (PP), (PS), (PB), (SS), (SB), and (BB). The RDFs (normalized to the ideal gas density) for some of these pairs are shown in Fig. 4, also for both studied systems.

Two important observations can be made based on Figures 3 and 4. The distributions for the RNA nanoring are in general broader, they contain extended tails, that reflect the existence of the important fraction of nonhelical regions. For example, the (PSPS) dihedrals contain two shoulders beside the main peak (at ≈ -150 deg) consistent with two RNA conformational classes found in [16]. Besides, as more careful examination of the dihedral histograms reveals, there exist some dihedral values (mainly in the "kissing loops"), that deviate strongly from the centres of the distributions. Their fraction is not high, and that is why they are not clearly visible in Fig. 3, however they have to be taken into account in a CG model to make the RNA nanoring stable. The detailed dependence of the (SPSP) dihedrals along the ring (versus the dihedral index) is shown in Fig. 5. The strongly deviating dihedrals are clustered in 12 groups that correspond to 12 parts of the nucleic backbones participating in 6 kissing loops. The remarkable fact that one can also observe in Fig. 5 is that in total about 4 dihedrals per "kissing loop" have strongest deviation from the distribution centres, of the order of 180 degrees.

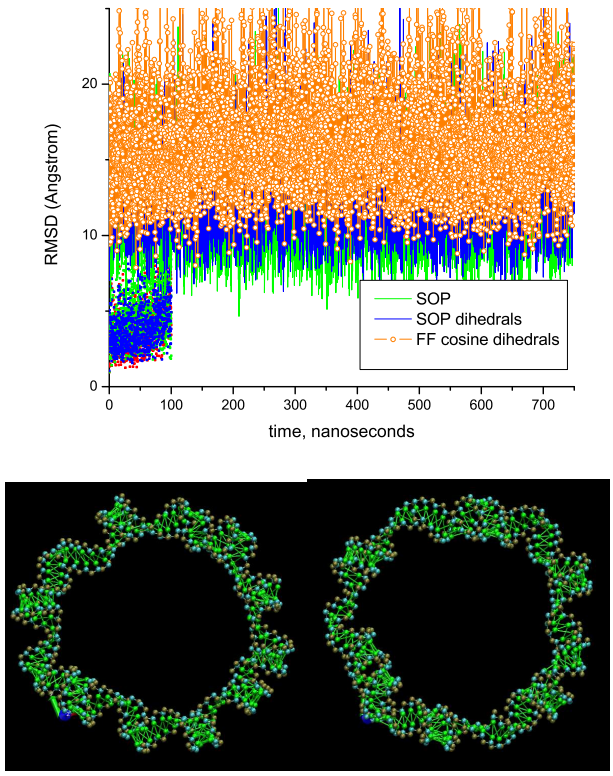

Fig. 6. (color online): Top: RMSD for different variants of the parameter sets of the CG model in the equilibration CG runs. The data for the nanoring are plotted with lines and open circles, while those for the dodecamer are plotted with dots. Bottom: The final snapshots of the RNA nanoring after 7500 ns equilibration in the CG model with the "SOP dihedrals" (left) and "FF cosine dihedrals" (right) parameter sets.

The nonbonded interactions for the RNA nanoring show well pronounced tails in the interval between 5 Å and 10 Å , which are absent in the case of the dodecamer. These features are caused by closely spaced phosphates in the kissing loop regions. This feature dictates that the nonbonded interaction potential should not discriminate such small interbead spacings energetically. That is why we have taken a rather small value of $\sigma = 5.0$ Å (and $\varepsilon = 0.1$ kcal/mol) for the nonbonded WCA parameters.

We use the distributions from Fig. 3 to get the effective CG potentials for bonded degrees of freedom via BI method, Eq. 5, namely an equilibrium value of a degree of freedom q (denoted with the superscript $^{(0)}$ in Eqs. 2 and 3), and the coefficient k for the corresponding harmonic potential $V_{eff}(q)$ are thus obtained:

$$V_{eff}(q) = \frac{k}{2}(q - q_0)^2. \qquad (7)$$

It is important to stress that these degrees of freedom for the RNA nanoring described just above are not distributed according to Boltzmann statistics only, but their distributions also reflect the spatial inhomogeneity of the RNA nanoring. Therefore an attempt to represent such a degree of freedom via a single potential function (as per BI method) would lead to instability of the desired structure in the CG model, because such a degree of freedom would be discriminated energetically in certain regions. Instead, one has to introduce some sort of local modifications to the potential functions for bond, angle, dihedral terms, and the base pairs interactions. The ultimate strategy of this sort is the SOP approach, where each instance of such a degree of freedom has its own equilibrium value depending on its location in the molecule.

Therefore, we decided to start from the CG parameters extracted from the histograms for the dodecamer, and to consider three different parameter sets: (i) "SOP" parameter set, where the coefficients k are uniform throughout the system, while the equilibrium values are unique for each instance of the bond/angle/dihedral *etc.* (ii) "SOP-dihedrals", where the SOP approach (i) is applied only to the PSPS and SPSP dihedrals along the nucleic backbone, but not to all other degrees of freedom, and (iii) "forcefield" (FF) parameter set where each degree of freedom is described by uniform parameters (including its equilibrium value) throughout the ring. The current working sets of parameters are avaialble upon request.

In the FF variant of the model, where all the dihedrals of the same type have the same unique equilibrium value (we term such variant "FF harmonic"), the dihedrals deviating far from the distribution centres (see Fig. 5) would be discriminated energetically very strongly by a harmonic or quartic effective potential of the sort of Eq. 7, and this would strongly distort the equilibrium structure of the kissing loops. However, another simple form of the dihedral potential function, that has an alternative minimum separated from the main minimum by ~ 180 deg might help to stabilize the kissing loops by accommodat-

ing such strongly deviating dihedrals. We tested the following dihedral function:

$$V_{eff}(\phi) = \frac{k}{4}[1 - \cos(2(\phi - \phi_0))], \qquad (8)$$

that has two minima at $\phi = \phi_0$ and $\phi = \phi_0 + 180$, both with the stiffness k. It turns out that this simple function provides an excellent performance of the "FF cosine dihedrals" variant of the model in describing the RNA nanoring.

Using DL_POLY 2.19 program, we relax the resulting CG representations of both the nanoring and the dodecamer and subject them to up to 750 ns equilibration at $T = 300$ K. The Root Mean Square Deviations (RMSDs) from the initial structures during these runs are plotted in Fig. 6.

Typical values of RMSD are $\approx 10 \div 15$ Å in the case of SOP or SOP-dihedrals variants, and RMSD is also about the same ($\approx 12 \div 17$ Å) for the "FF cosine dihedrals" variant. These values are to be compared to the typical values of $\approx 3 \div 4$ Å for the dodecamer (also plotted in Fig. 6) and to $\approx 30 \div 40$ Å for the "FF harmonic" variant (not plotted) where the overall shape of the nanoring is not preserved.

The final snapshot of the nanoring for the "SOP dihedrals" and "FF cosine dihedrals" variants are depicted in the same figure, attesting the preservation of the helical segments, kissing loops structure and the overall shape. We conclude thus, that overall the CG model provides an excellent performance in describing the structurally inhomogeneous RNA aggregate - the nanoring.

Further inquiry into the behavior of all considered CG models variants is provided by the histograms and RDFs for the CG degrees of freedom obtained in the end of the CG runs. Some of them are shown in Figures 7 and 8 in comparison to similar distributions from full MD. Most of the histograms of the bonded terms show reasonable agreement with those for full MD simulations, so no further refinement of these parameters is needed. The dihedral distributions show the required extended tails in all cases except, obviously, "FF harmonic".

What is even more remarkable, all the variants of CG model (except the "FF harmonic") are able to capture the fine features of the nonbonded RDFs. The most important of them is the 5 Å peak for the (PP) pairs, followed by a minimum (Fig. 8). As attested by the corresponding RDFs, this peak is fairly well reproduced, and it is even slightly overemphasized by the "FF cosine dihedrals" variant. Overall, agreement of the CG RDFs with those from the full MD is very good.

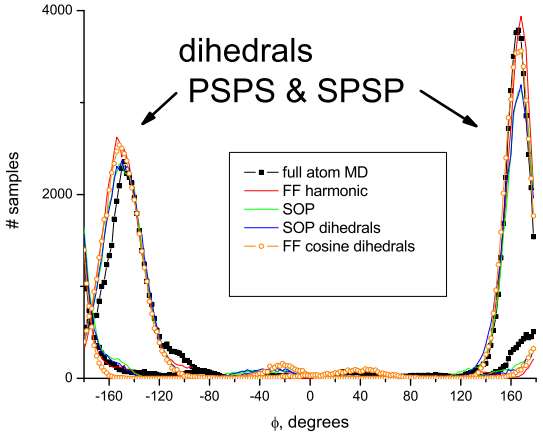

Fig. 7. (color online): Distributions for PSPS and SPSP for RNA nanoring from full atom MD and from CGMD runs in comparison.

5. DISCUSSION AND CONCLUSION

We thus considered a series of four simple variants of the RNA CG model, all of them based on a set of "pure helical" parametres extracted from the RNA dodecamer, and containing a controlled amount of "nonhelicity" included either via SOP approach or via a modification of the diheral functions. Three of these variants allow one to represent the structure of such complicated macromolecule as RNA nanoring reasonably well and are all comparable in performance (namely "SOP", "SOP-dihedrals" and "FF cosine dihedrals" variants). However, the "SOP" variants include a lot (of the order of the number of beads in the model) of specific structural information, namely, all the equilibrium values of the CG degrees of freedom. One can reduce the amount of such information based on an important finding from the previous section - that the inclusion of details about P-C4' dihedral degrees of freedom only is needed for a good representation of the structure of the considered system. This is consistent with a more general observation, that all possible RNA

conformations can be described by just a few regions ("classes") of the P-C4' dihedral space [16], similarly to the famous $\phi-\psi$ Ramachandran plots for proteins. In our opinion this opens up a promising avenue towards the ultimate goal of RNA CG modelling - to make a CG model with just a minimum number of universal parameters (a "forcefield") transferrable to other structures without any modification, and still being able to describe the inhomogeneous nonhelical conformations with a satisfactory precision. While this task is very ambitious in general, and we are not aware of any current advances in this direction (one-bead models seem to be not flexible enough), we believe that our layout of the model (3 beads per nucleiotide and sufficiently flexible interaction scheme) is potentially capable of accomplishing this task.

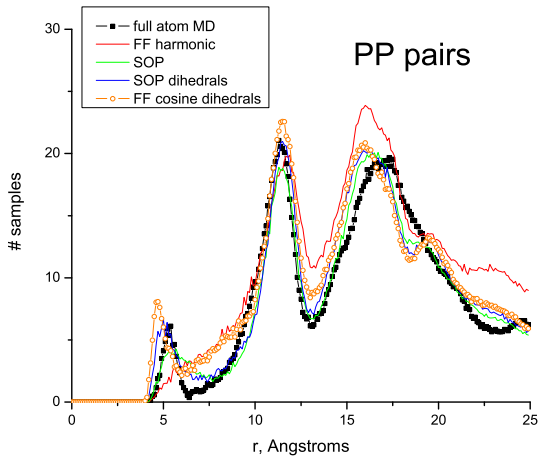

Fig. 8. (color online): Radial Distribution Functions for nonbonded interactions (PP pairs only) for RNA nanoring from full atom MD and from CGMD runs in comparison.

As a first step towards the above-mentioned ultimate goal, another important finding of the current study is that a simple modification of the dihedral potential function Eq. 8 allowed us to reach the SOP precision in describing the RNA nanoring by properly accommodating the distortions of the dihedral angles in the kissing loops. While in the present shape, the function defined by Eq. 8 has only one additional dihedral minimum separated from the original one by 180 degrees, this function can be made more complex (e.g. to contain multiple minima) as dictated by the concrete conformation classes [16] that the structures one aims to describe belong to. What's important to stress again, is that in light of our findings such function [more exactly, a pair of such functions - for both (PSPS) and (SPSP) dihedrals] seem to be the only parts of a CG model requiring an adjustment when the model is transferred to describe other structures.

The simple WCA nonbonded interactions $v(r)$ mimicking just the steric repulsion between the beads in the current version of the model proved to be sufficient to give a fairly good match between the full MD and the CGMD RDFs (Fig. 8). For further improvement, a more complex analytical functional form of $v(r_{ij})$ may be justified by some theoretical arguments, for example, they can be chosen as the functions of the "well-barrier-tail" shape [7], where the attractive part of the interaction describes the hydrophobic interactions between the beads, while the repulsive part serves the electrostatics. Alternatively, following [6] one may represent these functions with cubic splines (this allows for much greater flexibility in fitting the data), and use an automated procedure for optimisation. Our work along these lines is in progress.

Besides, a number of further useful extensions of the model is anticipated. (i) In the current version of the model no distinction is made between different base pairs. However, it is easy to introduce *sequence-specificity*, that can be important to represent even better the structure of RNA nano-assemblies. (ii) *Electrostatic interactions* between the negatively charged phosphate groups are to be included explicitely in the future. It has been shown, that counterions concentrate strongly at certain places inside the nanoring (in the "hairpin loops"), and therefore some sort of nonhomogeneous description of the electrostatics may be required, including a coarse-grained representation of the the counter-ions together with their solvation spheres. (iii) *Reactivity.* If the interaction between base pairs were treated in a non-bonding manner, this would allow one to study the association/dissociation reactions between the RNA nanoring building blocks. This feature of the model is important in order to answer the most interesting experimental question: how the character of the self-assebly of the nanostructures depends on

the specifc sequences of the constituents.

ACKNOWLEDGEMENTS

M.P. and R.M. are grateful to the NSERC CRC Program for support. B.S. was supported in part by the intramural research program of the NIH, National Cancer Institute, Center for Cancer Research. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca). The authors are grateful to Valentina Tozzini for helpful comments and suggestions.

References

1. Jaeger, L., and A. Chworos, Current Opinion in Structural Biology 16(4), 531 (2006).
2. Jaeger, L., E. Westhof and N.B. Leontis, Nucleic Acids Res. 29, 455 (2001).
3. Computational studies of RNA and DNA. Series: Challenges and Advances in Computational Chemistry and Physics , Vol. 2, Sponer, J.; Lankas, F. (Eds.) 2006, XI, 638 p.
4. Reith, D, M. Putz, and F. Muller-Plathe, J. Comp. Chem. 24, 1624 (2003).
5. Ercolessi, F. and J. B. Adams, Europhys. Lett., **26**, 583 (1994).
6. Izvekov, S., G. Voth, J Phys. Chem. B 109, 2469 (2005).
7. Trovato, F. and V.Tozzini, J. Phys. Chem. B. 112(42), 13197 (2008).
8. Holbrook, S. R., Current Opinion in Structural Biology 15(3), 302 (2005).
9. Yingling, Y. G., and B. A. Shapiro, Nano Letters 7, 2328 (2007).
10. Tozzini, V., Current Opinion in Structural Biology 15, 144 (2005).
11. Hyeon, C., and D. Thirumalai, Biophys. Journal 92(3), 731 (2007)
12. Hyeon, C., R. Dima and D. Thirumalai, Structure 14, 1644 (2006).
13. Hyeon, C., and D. Thirumalai, PNAS 102(19), 6789 (2005).
14. Jonikas, M.A. *et al.*, RNA 15, 189 (2009).
15. Trylska, J., V. Tozzini,and J. A. McCammon, Biophysical Journal 89, 1455 (2005).
16. Wadley, L.M., K.S. Keating, C.M. Duarte, and A.M. Pyle, J. Mol. Biol. 372, 942 (2007).
17. Paliy, M., R.Melnik, B.Shapiro, A Molecular Dynamics study of the RNA ring nanostructure, Submitted to Physical Biology (2009).
18. Phillips, J. C. *et al.* Journal of Computational Chemistry 26, 1781 (2005).
19. MacKerell, A.D. *et al.* J. Phys. Chem. B 102, 3586 (1998).
20. Humphrey, W., A. Dalke and K. Schulten, J. Mol. Graphics 14, 33 (1996).
21. Case, D.A. *et al.* (2008), AMBER 10, University of California, San Francisco.
22. Cornell, W. D. *et al.*, J. Am. Chem. Soc. 117, 5179 (1995).
23. Smith, W. and T.R. Forester, Journal of Molecular Graphics 14, 136 (1996); www.ccp5.ac.uk/DL_POLY/ .
24. Lee, A.J. and D.M. Crothers, Structure 6(8), 993 (1998).