



OPERA: An OPTimized coarse-grained Energy model for RnA

C. Colas, P. H. Nguyen, J.-Ch. Gelly, Ph. Derreumaux

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 185-188, 2008.

© 2008 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

OPERA: An OPTimized coarse-grained Energy model for RnA

Claire Colas¹, Phuong Hoang Nguyen², Jean-Christophe Gelly¹, and Philippe Derreumaux¹

¹ Laboratoire de Biochimie Théorique, UPR 9080 CNRS,
Institut de Biologie Physico-Chimique, Université Paris 7 Denis Diderot,
13 rue Pierre et Marie Curie, 75005 Paris, France
E-mail: Philippe.Derreumaux@ibpc.fr

² Institute of Physical and Theoretical Chemistry, J.W. Goethe University,
Max-von-Laue-Strasse 7, 60438 Frankfurt, Germany

RNAs have many cellular functions ranging from transcription to catalysis. The gap between sequences and 3D structures is increasing and knowledge of RNA dynamics and thermodynamics at an atomic level is missing. In principle, all-atom molecular dynamics (MD) and replica exchange molecular dynamics (REMD) simulations in explicit solvent can investigate these issues. However with current computer facilities, these simulations have been limited to small RNAs. To move to large RNAs, we can resort to coarse-grained models. In this study we present OPERA, a generic coarse-grained model for RNA. We report MD and REMD simulations on two RNAs of 22 nucleotides using a set of non-optimized OPERA parameters. Current results suggest that further optimization of the OPERA force field should open the door to a relevant model for studying large RNA such as riboswitches.

1 Introduction

All-atom molecular dynamics simulations in explicit solvent are often used to investigate the dynamics and thermodynamics of biomolecules, but they are time-consuming and slow to converge to equilibrium. Pande's group for instance managed to fold a 12 nucleotide RNA with Folding@home, but used 150 000 CPUs¹.

Reducing the number of degrees of freedom is one solution to accelerate convergence. Coarse-grained models have long history for proteins, but only 3 models exist for nucleic acids : one for DNA² and two for RNA^{3,4}, with each nucleotide represented by 3 beads : phosphate, sugar, base. Two of them are based on the Gō potential, which requires knowledge of the native structure. The last model developed by Dokholyan *et al.* shows promising results combining DMD and a square-well potential, but is not free of any biases⁴. We present here our *ab initio* OPERA force field.

2 OPERA

In OPERA, each nucleotide is represented by 6 to 7 beads (Fig. 1): 1 bead for the phosphate, 4 beads for the sugar - O5', C5', C4', C1'- and 1 bead for the base in pyrimidines and 2 in purines. The number of particles is reduced of 80%. Three of the six torsional angles are thus conserved : α , β and γ . Note that the OH group specific to RNA is not treated explicitly and the sugar pucker cannot be modeled. The solvent is treated implicitly.

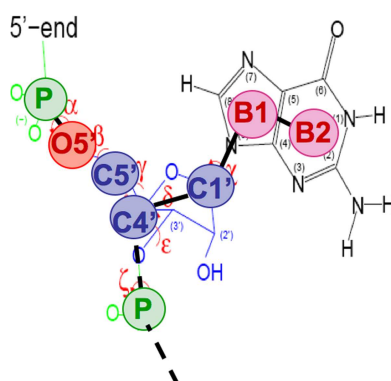


Figure 1. Our coarse-grained model is represented for a guanine superposed on an all-atom representation.

OPERA is based on the OPEP force field (Optimized Potential for Efficient peptide-structure Prediction) developed in our group for proteins^{5,6}. The potential describes short range and long range interactions as follows : $E = E_{bonds} + E_{angles} + E_{torsion} + E_{LJ} + E_{HB}$. E_{LJ} is a 6-12 potential and E_{HB} is represented by two-body and four body interaction terms⁷. The geometrical parameters were derived from a statistical study of 220 structures of the PDB. Here, we use a set of non-optimized force constants.

3 Preliminary Results and Conclusions

MD simulations of 50 ns were performed at 310 K on two RNAs, each of 22 nucleotides : one hairpin (1EOR) and one pseudoknot (2G1W). Figures 2a and 2b show the RMSD of both systems with respect to their NMR structures. The hairpin (Fig. 2a) is rather stable, with the RMSD remaining around 4.5 Å during the first 32 ns and then around 3.5 Å. On the other hand, Figure 2b shows that the pseudoknot deviates much more. The RMSD evolves between 4 and 7 Å until 36 ns when it suddenly decreases to 2.5 Å (Fig. 2c) before reaching 11 Å at 50 ns.

REMD simulations were also performed on the hairpin, using 14 replicas with T ranging from 310 K to 360 K. Each replica is simulated for 150 ns and exchange events are attempted every 10 ps. Cluster analysis was done at 310 K using the 20-150 ns time interval with a RMSD cutoff of 2.5 Å. A total of 33 clusters is found and two clusters represent 75% of all conformations. The centers of these clusters are at 14.9 Å and 4.6 Å from the NMR structure. Figure 2d superposes the NMR structure on the center of the second cluster. Note that this state does display three non native H-bonds.

We have presented a new coarse-grained model for RNA. Our preliminary results, using a set of non-optimized force constants, are very encouraging since the hairpin is apparently stable at 310 K and the pseudoknot is in equilibrium between native-like and unfolded states. Analysis, based on 150 ns REMD simulations, confirms that the hairpin can adopt a native-like structure, albeit with a lower probability than an unfolded state. We believe that optimization of OPERA should stabilize the structure and capture the thermodynamics of large RNAs.

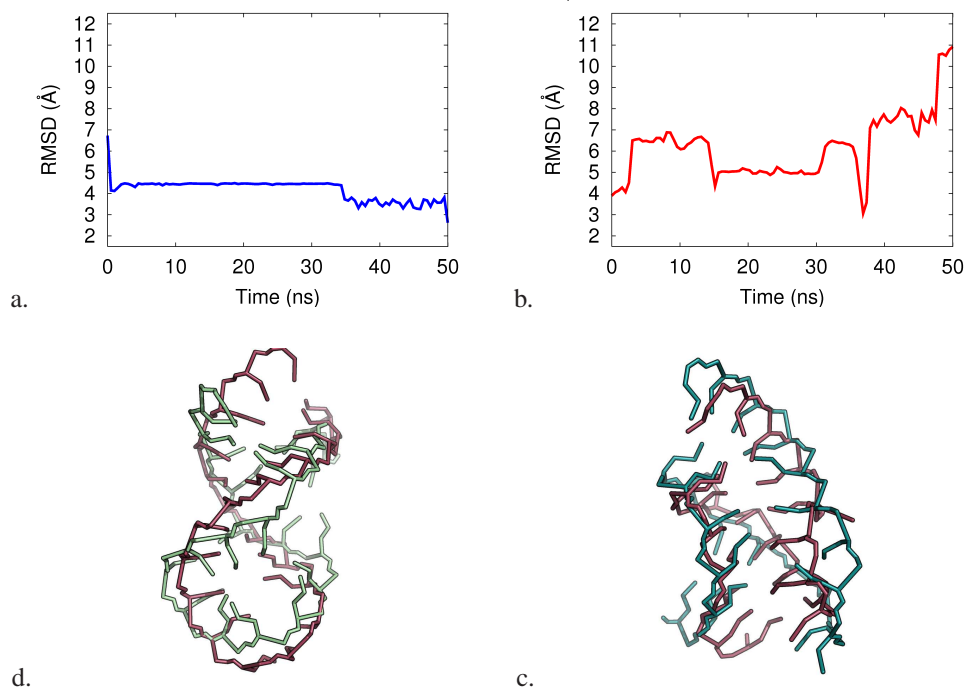


Figure 2. MD-evolution of the RMSD with respect to the NMR structure of 1EOR (a) and 2G1W (b). The RMSD is calculated on the P, O5', C5', C4' and C1' particles using the nucleotides 2-21 (the first and last are flexible by NMR). (c) 2G1W : superposition of the NMR structure (pink) on the MD-structure found at 36 ns (blue). This structure has 3 among 7 native H-bonds. (d) 1EOR : superposition of the NMR structure (pink) on the center of the second cluster predicted by REMD (green).

References

1. Sorin, E.J., Bradley, J., Nakatani, J., Rhee, Y. M., Jayachandran, G., Vishal, S., Pande, V.S. *Does native state topology determine the RNA folding mechanism?* J. Mol. Biol. **337**, 789-797, 2004.
2. Knotts, T. A., Rathoe, N., Schwartz, D. C., de Pablo, J. J. *A coarse grain model for DNA.* J. Chem. Phys. **126**, 2007.
3. Hyeon, C., Thirumalai, D. *Chemical theory and computation special feature: mechanical unfolding of RNA hairpins.* Proc Nucl. Acad. Sc. USA. **102**, 6789-6794, 2007.
4. Ding, F., Sharama, S., Chalasani, P., Demidov, V. V., Broude N. E., Dokholyan, N. V. *Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms.* RNA **14**, 1-10, 2008.
5. Derreumaux, P., Mousseau, N. *Coarse grained protein molecular dynamics simulations.* J. Chem. Phys. **126**, 025101, 2007.
6. Maupetit, J., Tuffery, P., Derreumaux, P. *A coarse grained protein force field for folding and structure prediction.* Proteins **69**, 394-408, 2007.
7. Derreumaux, P. *Simulating the early steps of amyloid fibril formation and disassembly.* See accompanying paper in this volume.

