Developing an accurate all-atom fixed-charge force field for RNA with implicitly polarized charges

Chapin E. Cavender, Louis G. Smith, Alan Grossfield, & David H. Mathews
Department of Biochemistry & Biophysics, University of Rochester Medical Center, Rochester, New York, USA

Introduction

Molecular dynamics (MD)

- Use classical physics to obtain a time series of coordinates for every atom
- Model structural ensembles that mediate biological functions of macromolecules
- Generate hypotheses about structure that can then be tested experimentally
- Relies on the accuracy of the energy model, a parameterized function called a force field

RNA force fields

- Accurate for A-form helices composed of canonical Watson-Crick base pairs
- Poor description of non-canonical base pairs[1] and relative orientation of helices interrupted by single-stranded loops[2]
- Development of an accurate fixed-charge force field that is transferable to structurally diverse RNAs remains an open challenge

$$U_{\text{nonbonded}}(R_{jk}) = \sum_{j<k} \left(\varepsilon_{jk} \left[\left(\frac{R^*_{jk}}{R_{jk}} \right)^{12} - 2 \left(\frac{R^*_{jk}}{R_{jk}} \right)^{6} \right] + \frac{Q_j Q_k}{4\pi\epsilon_0 R_{jk}} \right)$$

We are developing an accurate fixed-charge force field for RNA using the Amber functional form by fitting to quantum mechanics (QM) energies that are implicitly polarized to account for the influence of solvent.

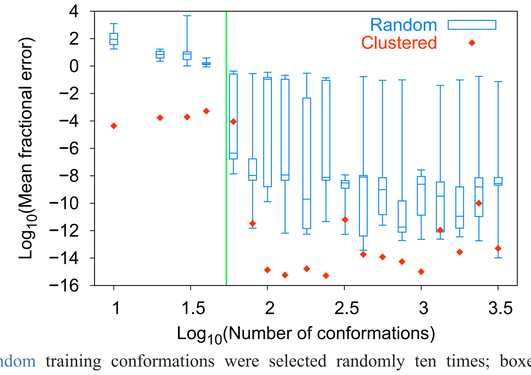
Implicitly polarized charges[3] strategy

Nucleoside dimers and nucleoside-phosphate dimers
↓
Time-averaged solvent density around each RNA
↓
Compute interaction energy ($E_{AB} - E_A - E_B$) by QM
↓
Fit nonbonded parameters to QM interaction energies

Clustering improves efficiency of the fitting dataset

- Initial dataset is non-redundant representative set[4] of experimental RNA structures
- Extract all pairs of RNA residues with interacting nucleobases (~260 000 pairs)
- Compute nonbonded interaction energies with Amber ff99+bsc0+χ_{OL3} force field[5,6]
- Train nonbonded parameters to reproduce Amber energies for subset of conformations

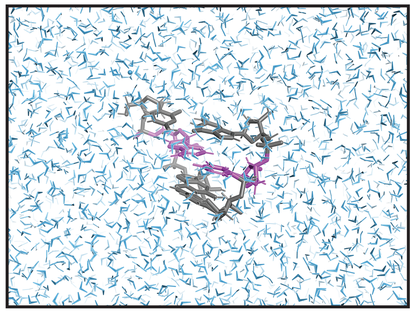
$$\text{Mean fractional error} = \frac{1}{N_{\text{Test}}} \sum \frac{\text{abs}(U_{\text{Fit}} - U_{\text{Amber}})}{\text{abs}(U_{\text{Amber}})}$$



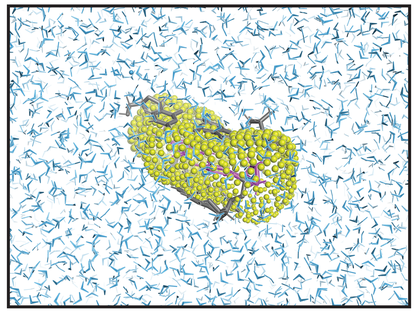
- Random training conformations were selected randomly ten times; boxes show interquartile range with median, and bars show minimum and maximum
- Clustered training conformations were clustered by heavy atom pairwise distance
- Vertical line indicates where there are more conformations than parameters to fit
- Clustering retains diversity of training dataset and outperforms random selection

Strategy for obtaining time-averaged electrostatic interaction with solvent

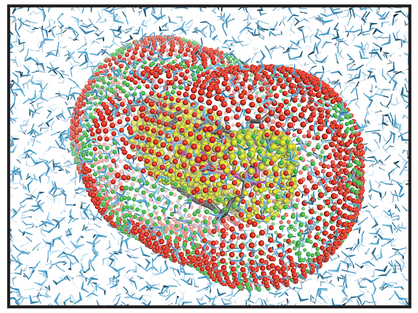
Extract RNA context around QM dimer from experimental structure → MD simulation of water and ions around a restrained RNA → Mesh of equidistant points at radius 3 Å → Compute eletrostatic potential (ESP) at 3 Å due to solvent farther than R → Fit point charges on far mesh to reproduce solvent ESP on close mesh
MD simulation of water and ions around a restrained RNA → Mesh of equidistant points at radius R > 3 Å → Fit point charges on far mesh to reproduce solvent ESP on close mesh



Example of a nucleoside-nucleoside dimer (magenta) and the context from an experimental structure (gray) solvated in water and salt (cyan).



Mesh of equidistant points (yellow) at radius 3 Å from QM dimer (magenta). Point size represents magnitude of ESP at that point due to solvent farther than R > 3 Å.

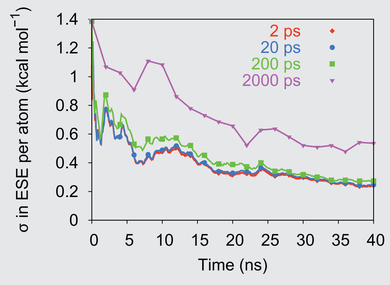


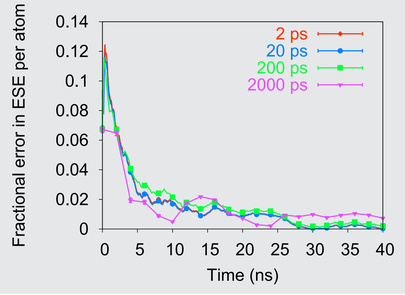
Mesh of point charges at radius R > 3 Å reproduce the solvent ESP at 3 Å (yellow). Point size represents magnitude of the charge, red positive or green negative.

Is solvent density converged from simulation length and amount of solvent?

Replicas give similar estimates of solvent density

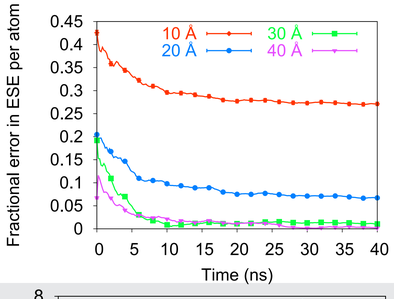
- Assess convergence of solvent density for a single conformation with respect to sampling interval, simulation length, and solvent cutoff distance
- Electrostatic energy (ESE) between RNA and solvent
- Standard deviation, σ, in ESE (left) across 10 replicas
- Relative error in ESE (right) compared to 2 ps, 40 ns, and 40 Å; mean ± SEM across RNA atoms
- Colors represent sampling interval
  - Intrinsic σ across replicas is 0.25 kcal mol^{-1} per atom
  - Sampling at 200 ps does not change variance or relative error in ESE compared to sampling at 2 ps

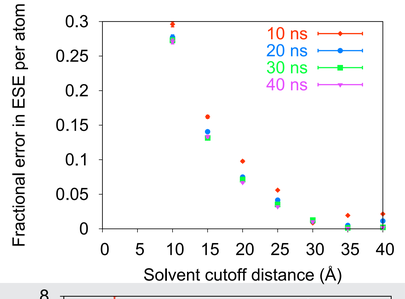




Large system size is required to converge solvent density

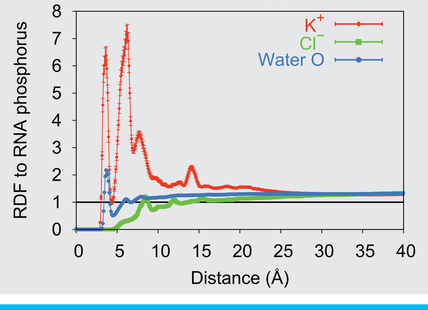
- Relative error in ESE for 200 ps sampling vs. simulation length (left) or solvent cutoff distance (right), mean ± SEM across RNA atoms
  - Relative error in ESE stops changing after 20 ns
  - Relative error in ESE is still changing up to 30 Å

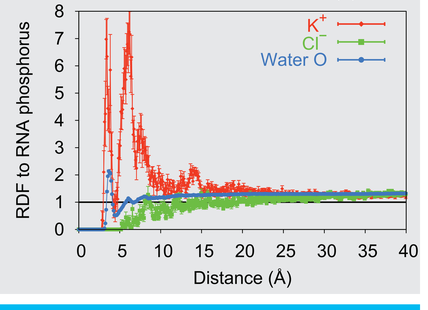




Large size requirement due to salt

- Radial distribution function (RDF) gives the probability density for pairwise distances compared to an ideal gas (RDF = 1)
- RDFs of solvent atoms to RNA phosphorus atoms for sampling rate of 2 ps (left) and 200 ps (right), mean ± SEM across 10 replicas
  - Water oxygens are unstructured after 10 Å
  - Salt ions are structured up to 25 Å
  - Variance across replicas is higher for salt than for water at 200 ps sampling interval





Do meshes capture influence of solvent density on QM energy?

Meshes reproduce QM energy of explicit solvent density

- QM interaction energy of a nucleoside-nucleoside dimer computed using symmetry-adapted perturbation theory[10], sSAPT0/jun-cc-pVDZ, in Psi4[11]
- Approximate solvent density with point charge mesh to reduce computational burden
- "RNA" contains only point charges from RNA context of experimental structure
- "Solvent" contains explicit solvent density for solvent within 10 Å of the QM atoms
- "Mesh" contains explicit solvent density up to a radius R and fits point charges on a mesh to reproduce solvent ESP for solvent between R and 10 Å

	Radius (Å)	Number of point charges	QM energy (kcal mol^{-1})	Time (h)	Solvent ESP relative error
RNA		134	−14.804380	0.22	
Solvent	10	134415	−15.050562	22.51	Reference
Mesh	9	105813	−15.050563	17.66	0.003068
Mesh	8	80388	−15.050567	13.46	0.003064
Mesh	7	59322	−15.050566	10.03	0.002997
Mesh	6	39819	−15.050557	6.08	0.003914

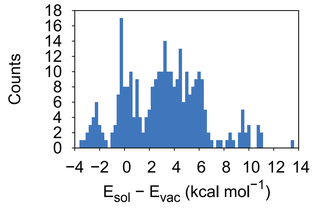
Meshes reduce QM job time by 10-fold

- Include solvent within 40 Å; vary number of meshes, mesh radii, and point density
- QM energy relative error with respect to 10 Å mesh with 1 Å^2 per point
  - Best mesh reproduces QM energy within 0.1 % and reduces job time by 10-fold

Mesh radii (Å)	Density (Å^{-2})	Number of point charges	QM energy relative error	Time (h)	Solvent ESP relative error
10	1.000	122009	0.0000	21.02	0.0007
4	1.000	1030	0.0005	2.67	0.0032
10 4	1.000	3968	0.0005	3.12	0.0033
10 8 6 4	1.000	7754	0.0017	3.70	0.0032
4	0.333	341	0.0012	2.56	0.0032
10 4	0.333	1324	0.0033	2.73	0.0032
10 8 6 4	0.333	2590	0.0030	2.90	0.0033
4	0.083	91	0.0481	2.51	0.0036
10 4	0.083	324	0.0440	2.55	0.0044
10 8 6 4	0.083	650	0.0126	2.62	0.0037

Extent of solvent influence is diverse across clustered dataset

- Compute QM interaction energy for 300 clustered nucleoside dimers
- Histogram of difference in QM energy with (E_{sol}) and without (E_{vac}) mesh
  - Solvent mesh typically makes interaction energy more unfavorable
  - Magnitude of energy difference varies across the dataset



Conclusions

- Clustering RNA dataset on heavy atom pairwise distance improves efficiency
- Estimation of solvent density from MD is converged in time and space
- Point charge meshes reproduce influence of solvent density on QM energy

Future directions

- Compute QM interaction energies for nucleoside-phosphate dimers
- Fit nonbonded parameters to QM energies
- Refit torsions in the context of new nonbonded parameters
- RNA simulations to benchmark force field

Simulation details

- Amber ff99+bsc0+χ_{OL3}[5,6], 2 fs time step, SHAKE
- PME electrostatics with 10 Å cutoff
- OPC water[7] and 150 mM KCl with 40 Å padding
- 10 kcal mol^{-1} Å^{-2} position restraints on RNA
- Langevin thermostat at 310.15 K
- Monte Carlo pressure equilibration at 1 atm
- 10 replicas of 40 ns NVT dynamics in Amber 16[8]
- Data analysis performed in Loos[9]

References

[1] Kuhrova et al. 2016. *Journal of Chemical Theory and Computation* **12**, 4534–4548
[2] Andralojc W et al. 2016. *Physical Chemistry Chemical Physics* **18**, 5743–5752
[3] Cerutti DS et al. 2013. *Journal of Physical Chemistry B* **117**, 2328–2338
[4] Leontis NB & Zirbel CL. 2012. In *RNA 3D Structure Analysis and Prediction.* Springer, 281–298
[5] Perez A et al. 2007. *Biophysical Journal* **92**, 3817–3829
[6] Zgarbova M et al. 2011. *Journal of Chemical Theory and Computation* **7**, 2886–2902
[7] Izadi S et al. 2014. *Journal of Physical Chemistry Letters* **5**, 3863–3871
[8] Case DA et al. 2017. AMBER 2017. University of California, San Francisco.
[9] Romo TD et al. 2014. *Journal of Computational Chemistry* **35**, 2305–2318
[10] Parker TM et al. 2014. *Journal of Chemical Physics* **140**, 094106
[11] Parrish RM et al. 2017. *Journal of Chemical Theory and Computation* **13**, 3185–3197