Exploring Hydrogen Bond Geometry in RNA With F-SAPT

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

Introduction

- Modeling Nucleic Acids (NAs) using Molecular Mechanics contributes to biology, biochemistry, and medicine.
- Simulations lose native structure
- Some Hydrogen Bonds (HBs) bent
  - HBs essential for NA structure
  - Rosetta adopted dihedral HB term for peptide modeling
- Orientation preferencess unknown
- Point charges shouldn't work for certain HBs
- Classic force fields (AMBER) only have radial nonbonded terms
- Lack of clean experimental data → computational chemistry
- Functional group (FG) Symmetry Adapted Perturbation Theory (**F-SAPT**) → interaction energy (IE) between two FGs[1]
- FG-IEs sum to SAPT0 IEs with mean absolute error 1.0 kcal/mol
- **Goal**: Patch classic NA force fields

What is the orientation dependence for RNA base-base hydrogen bonds? Can standard force fields mimic this behavior?

Strategy

- Scan of functional group interaction energy over HB orientation
- Guanine-Cytosine (GC) Nucleioside
  - most stable pair
  - largest dipole moment
  - no costly phosphate in F-SAPT
- Explore GC → reaction coordinate
- Distort Watson-Crick (WC) HB
- Comparison of functional form, not parameters → model fitting
- Capability of model → quality of fit
- Favor classic functional form by
  - Overdetermined fitting set
  - No transferability
  - Just fit atoms in HB
  - Just constrain chemically identical atoms (here, H)
  - $-1 \leq$ atomic charges ≤ 1
  - $0 \leq$ hydrogen charges ≤ 1
- Design patch terms
- Fit patch terms in context of global refit of nonbonded terms

Methods

Scan

- Nucleoside dimer built in tLeap
- Manipulated using restrained Simulated Annealing (SA) in openMM 7.x
- φ: dihedral between carboynl C, O (C), amino N and C (G)
- ϑ: angle between carbonyl C, carbonyl O (C), and amino N (G)
- WC (φ, ϑ) = (27.9°, 116.4°)
- Restraints
  - Bumper restraint

$$B(r_{ij}) = \begin{cases} \frac{k}{2}(r_{ij} - r_0)^2 & r_{ij} < r_0, \\ 0 & r_{ij} \geq r_0 \end{cases}$$

    - O2′(C) to C1′(G), $r_0 = 10$ Å
    - O3′(C) to C1′(G), $r_0 = 10$ Å
  - Flat-bottom harmonic (FB)
  - FB angle: carbonyl O (C), amino N (G), and amino C (G) at 120°, width 10°
  - FB dihedral: amino H, N, C, imino N (G) at 0°, width 10°
  - FB distance: imino N (C) and imino N (G) at 3 Å, width 0.4 Å, $k = 10$ kcal ·mol^{-1} Å^{-2}
  - FB distance: amino N (C) and carbonyl O (G), 3 Å with 0.4 Å δ, $k = 10$ kcal ·mol^{-1} Å^{-2}
  - Harmonic distance: amino N (G) and carbonyl O (G).
  - Harmonic restraints enforce (φ,ϑ): $k = 1000$ kcal · mol^{-1} Å^{-2}.

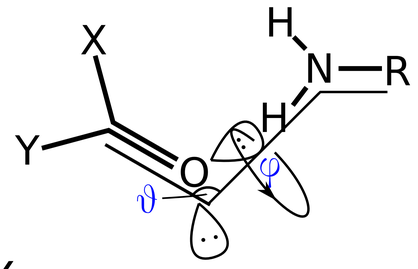
Interaction Energies from F-SAPT

- Used Psi4 v1.2 to perform F-SAPT
- Selected recommended parameters
  - Basis: jun-cc-DVPZ
  - Freeze-Core: True
  - Guess: SAD
- Computed IE for each conformation
- FG partition in postprocessing
- Chose **Functional Groups**
  - C: carbonyl C and O
  - G: amino N, both Hs, and C
- Doesn't cut bonds with DOCC > 0.1

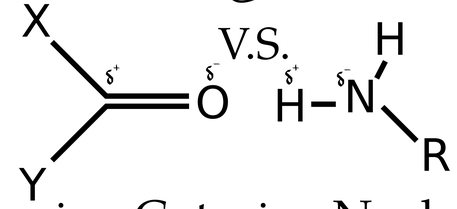
Coordinate Comparison & Fits

- Structures iteratively aligned to average with LOOS 3.0 (See QR)
- mean RMSD from average per (φ,ϑ) plotted as heatmap
- fits performed using basin-hopping routine from scipy.optimize(1.2.1)
  - 1000 Metropolis-Hastings Monte-Carlo steps per fit.
  - Local optimization using L-BFGS-B, then accept.

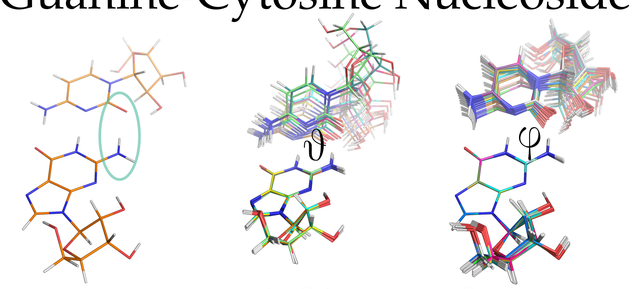
Carbonyl-Amino HB



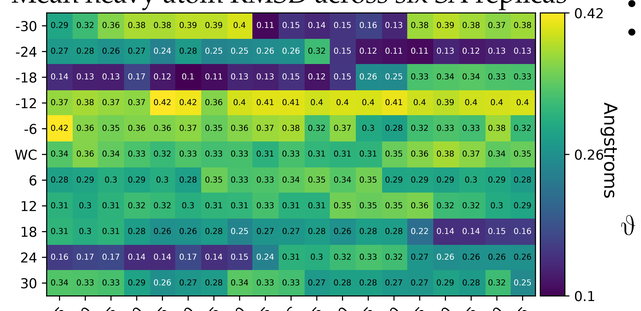
V.S.



Guanine-Cytosine Nucleoside



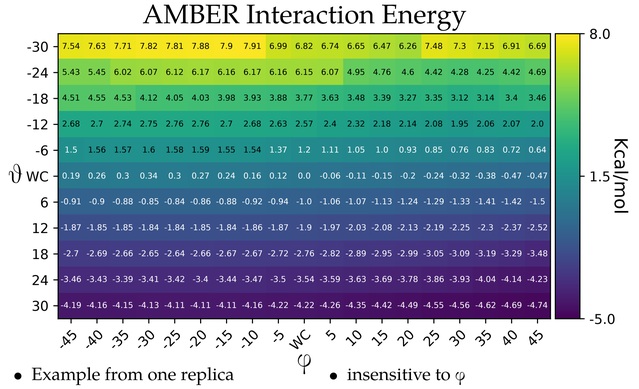
Mean heavy atom RMSD across six SA replicas



- Langevin thermostat is stochastic
  - Replicated process six times.
- Highest RMSD for WC atoms < 0.2 Å
- differences from distant atoms

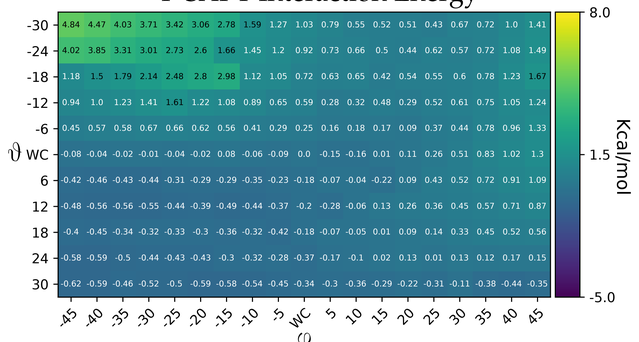
Force Field is Indifferent to Phi

AMBER Interaction Energy



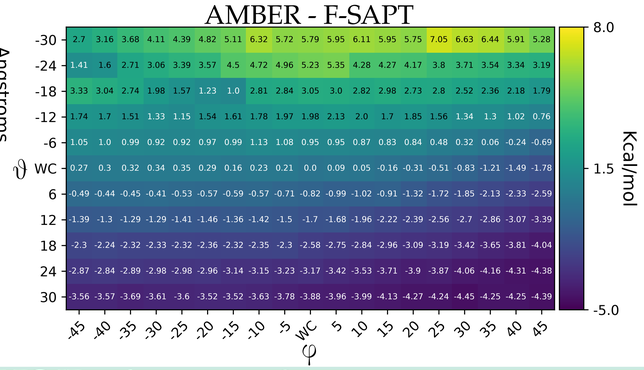
- Example from one replica
- Zeroed at Watson-Crick (WC) (φ,ϑ)
- insensitive to φ

F-SAPT Interaction Energy



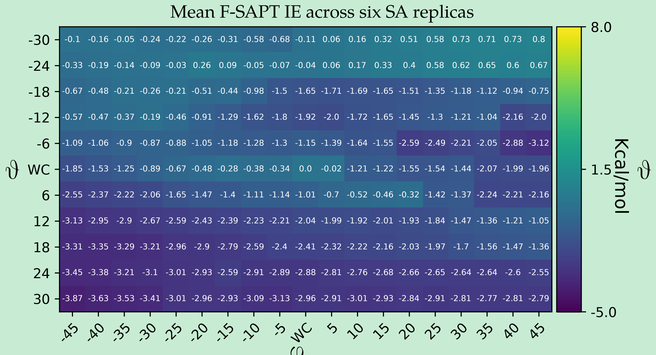
- **Used as fitting target**
- FG-IE for same conformations
- Subtracted from AMBER IEs below
- Δ FG-IE < Δ AMBER IE

AMBER - F-SAPT



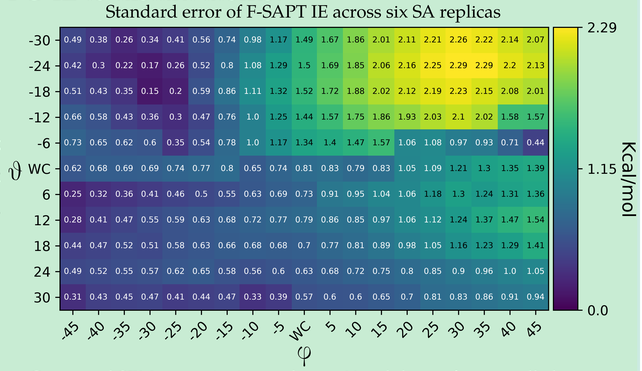
How sensitive is F-SAPT FG-IE to distant atoms?

Mean F-SAPT IE across six SA replicas



- Calculated 209×6 = 1254 QM IEs
- Colorbar from AMBER IE scan
- IEs lower, smoother than AMBER
- Retains trends in φ

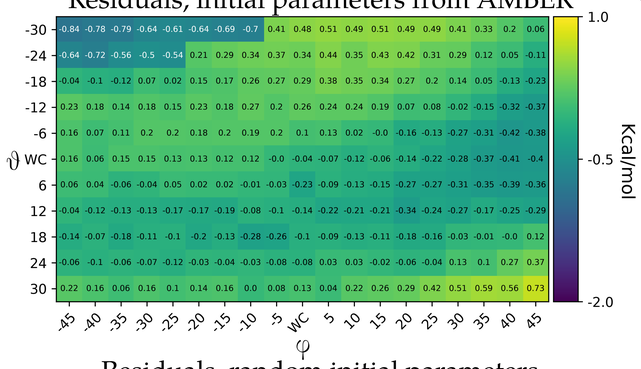
Standard error of F-SAPT IE across six SA replicas



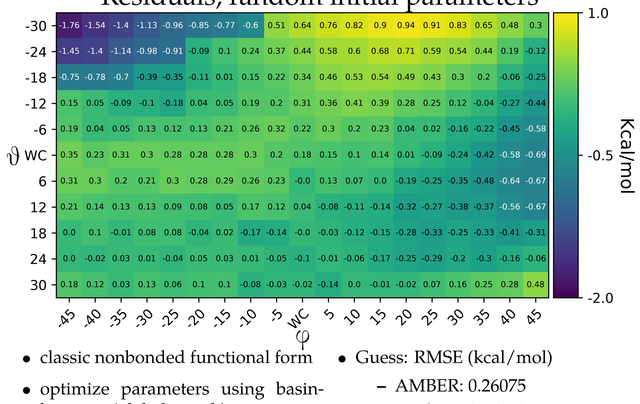
- Computed SEM using 6 – 1 samples
- No SEM for AMBER > 0.1 kcal/mol
- Variability is from small changes in distant atoms

Fit to F-SAPT

Residuals, initial parameters from AMBER



Residuals, random initial parameters



- classic nonbonded functional form
- optimize parameters using basin-hopping (global search)
- Cell values = modeled - target
- IE surfaces (targets) plotted to left
- AMBER target: positive control
- AMBER guess: AMBER parameters
- random guess: Random parameters
- Guess: RMSE (kcal/mol)
  - AMBER: 0.26075
  - random: 0.26069
  - AMBER target: 0.00488
- Constraints
  - H parameters must be equal
  - H charge > 0
  - −1 < partial charges < 1
  - all VDW parameters > 0

Conclusions

- Fitting without the restrictions of Atom type or broader context
- best-possible case for a parameter set or model.
- regions very close to WC are more accurately predicted.
- Unzeroed interaction energies in those regions are similar.
- **Classic force field functional form cannot match QM.**

Future Directions

- Need greater diversity of systems
- Cross-validate above fits
- Compare parameters to AMBERs
  - scan other HBs in GC pair
  - scan AU pair
  - structures from full nonbond fit
- fit all replicas simultaneously
- Scan distance with this system
- Competitive fit of trial patch terms
- Fit patch term in the context of refitting other parameters

⟹ See our other poster, **B109**, on Tuesday in RNA Structure and Dynamics!

References

1. Parrish, R. M.; Parker, T. M.; Sherrill, C. D. *J. Chem. Theory Comput.* 2014, 10 (10), 4417–4431.
2. Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. *RNA* 2015, 21 (9), 1578–1590.

LooS Grossfield Lab

LOOS is an open source library designed for rapid development of software to analyze molecular dynamics simulations. It is available for download from `https://github.com/GrossfieldLab/loos`