Winter 12-2019

Exploration of Nucleic Acid Hydrogran Bonding Using Molecular Mechanics and DFT Calculations

Simi Kaur

Exploration of Nucleic Acid Hydrogen Bonding using Molecular Mechanics and DFT Calculations

An honors thesis presented to the
Department of Chemistry,
University at Albany, State University of New York
in partial fulfillment of the requirements
for graduation with Honors in Chemistry
and
graduation from The Honors College

Simi Kaur

Research Mentors: Parisa Ebrahimi, Chris Meyers, M.A.
Research Advisor: Alan A. Chen, Ph.D.

December 2019

Abstract

Many recent theoretical and experimental techniques have been developed to probe the structure-function relationships of complex biomolecules. The roles of RNAs are dependent upon various intricate structural motifs and interactions, including hairpins, pseudoknots, long range territory contacts, bulges and internal loops, that are not easily captured by these methods. We had previously developed an enhanced replica exchange molecular dynamics method that incorporated secondary structure information in the form of distance restraints in order to effectively overcome kinetic barriers and sample conformational space. In several structures, restrained RNA base pairs near large bulges displayed a preference for stacking over hydrogen bonding in simulation. This persisted after usage of numerous enhanced sampling methods. Since RNA tertiary structure is scaffolded by a complex hydrogen bonding network, an accurate depiction of directional hydrogen bonding is essential to obtain predictive nucleic acid models. We compared gas phase DFT and molecular mechanics energies for cytosine-guanine nucleobases and nucleosides in order to determine the accuracy of the classical depiction of hydrogen bonding and evaluate the need for an explicit hydrogen bonding potential for nucleic acids.

Acknowledgements

I would like to sincerely thank Dr. Alan Chen and the entire Chen Lab for their help and guidance in not only this project, but for always going above and beyond. Specifically for this project, Chris Meyers helped me learn how to use the quantum computing software, conduct molecular mechanics calculations, gather ideas, and understand programming. Parisa Ebrahimi assisted me with the initial gromacs simulations that led to this project. This work would not have existed without everyone's help, guidance and patience. I would also like to thank the Department of Chemistry and the Honors College for providing me with the motivation to get involved in research.

List of Figures

Table of Contents

Introduction

All Atom Explicit Solvent Molecular Dynamics Simulations:

Molecular dynamics (MD) simulations use Newtonian motion and atomistic interactions to provide an atomic level understanding of complex molecules. The time evolution of positions generates trajectories using classical physics: at each timestep, forces are computed using Newton's laws of motion, new atomic velocities are calculated, atoms are displaced and coordinates are updated. Simulations can provide explanations of experimental results by analyzing properties that cannot be directly observed, elucidating small timescale dynamics and constructing predictions that can be tested through experiment. A schematic diagram, shown in Figure 1, displays how molecules in solution are modeled in simulations.

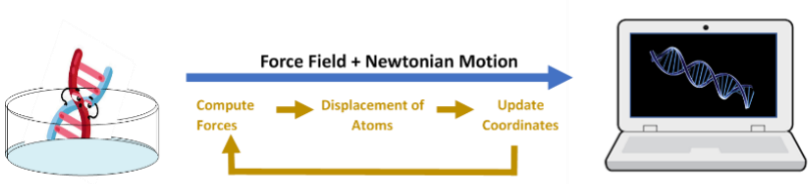

Figure 1. Schematic diagram for MD simulations. Molecular dynamics simulations model molecules in solution using force fields and classical ideas of Newtonian Motion.

The force field, a set of nonbonded and bonded atomic potentials, parameterizes intermolecular and intramolecular forces for a specific type of system. Protein force fields simulate protein folding very well, but are less successful with describing the base stacking and pairing interactions that structurally characterize nucleic acids. After force field selection, the system is assembled through acquisition of the initial molecular coordinates, placement of the molecule in a box and solvation through addition of water and ions. Periodic boundary conditions are used to

mimic an infinite system, the nature of which is determined by the relative volume comparison between the box and molecule: smaller boxes with less solvent imitate an infinite crystal, and larger boxes, containing a higher proportion of solvent, imitate an infinite solution. An example of a simulation with periodic boundary conditions is shown in Figure 2. Simulations can incorporate experimental information in the form of restraints (bias forces) that place an attractive potential between atoms.

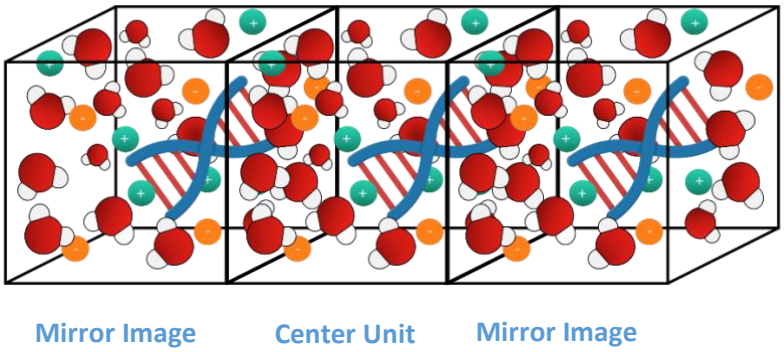

Figure 2. Solvated simulation box. The center unit represents a simulation box containing DNA, solvated with water and ions. The mirror images depict periodic boundary conditions in one direction. There are 3 mirror images above and below the center unit, and nine mirror images behind and in front of the center unit. Thus, one center unit is surrounded by 26 mirror images.

Folding nucleic acids in simulation is a very difficult and tedious process that may require months of supercomputer time with the assumption of ideality. Currently, typical all atom MD simulations operate on the nanosecond timescale, and rarely venture above the microsecond timescale. The limited sampling of conformation space is due to structural confinement in kinetic traps (local minima). To circumvent this, we created a novel 2D replica exchange enhanced sampling protocol which incorporates experimental information in the form of piecewise linear-harmonic bias forces. In this method, 9 copies of the same system, with varying temperatures and base-pairing restraints, exchange states in accordance with the metropolis criterion. This refines

the energy landscape (Figure 3) and guides the structure to preserve only the most stable conformations at lower temperatures[4]. The 2D simulation method is shown as a grid in Figure 4.

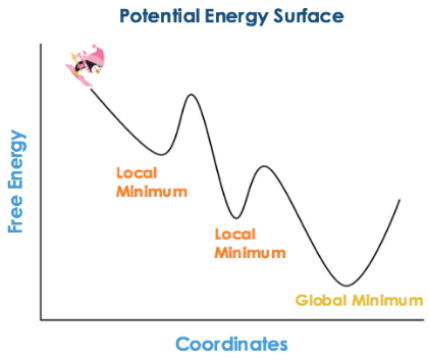

Figure 3. Graphical depiction of a potential energy diagram. The potential energy surface can be compared to a skier traveling down a slope. The skier requires the kinetic energy to overcome the barriers in order to reach the bottom of the slope, otherwise they will get stuck in the valleys. Similarly, molecules require the free energy required to overcome kinetic traps to reach the global minimum structure, otherwise they will get stuck in local minimums.

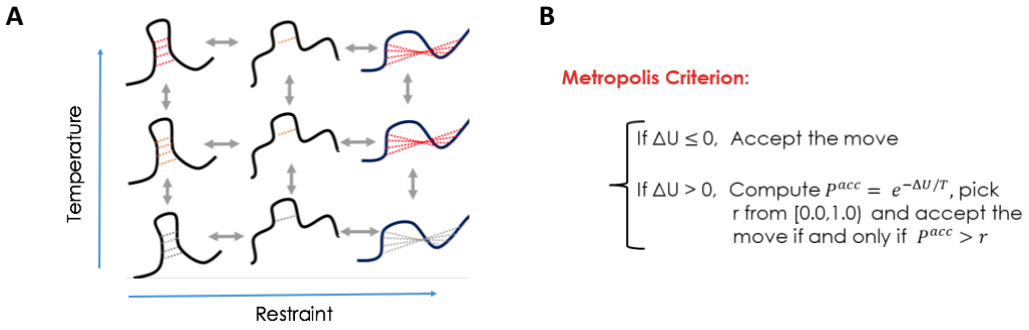

Figure 4. The 2D replica exchange method and the metropolis criterion. A) Replica exchange conducted in both temperature and position space, where replicas (molecules) with variable bias forces and temperatures run in parallel and periodically swap according to the Metropolis Criterion. B) The Metropolis criterion calculates the free energy difference of exchanging two structures. If this energy difference, ΔU, is negative or 0, the exchange is accepted. If it is greater than 0, then the probability of acceptance, P_{acc} is calculated, and compared to a random number, r, chosen from 0 to 1. There is a lower likelihood that an r value that is less than a low P_{acc} will be randomly selected. This method allows the system to access higher energy structures based on a probabilistic approach.

Although this method is fruitful, difficulties arise when portions of a stem melt at lower temperatures than the remainder of the molecule. In these situations, it is difficult to design a temperature protocol where all segments of the molecule melt and reform bonds in a similar fashion. In systems where there was one or several small stems adjacent to a large bulge or flexible region, stacking was preferred in simulation for base pairs adjacent to these regions over Watson Crick base pairing fortified with bias forces. Simulations of a hammerhead ribozyme and the signal recognition particle (SRP) of Escherichia coli exhibited this phenomena. The ribozyme contained a guanine-cytosine (GC) base pair, bordering a large bulge, that was observed to stack for over 20 ns. The stacking was stable at higher temperatures and persisted after energy minimization with increasingly large step sizes. Similar stacking was observed in SRP with a G-U base pair in a long stem near a bulge region[17]. It is unclear whether this type of trap is of a thermodynamic or kinetic nature.

The bias forces that are used to define known base pair hydrogen bonds are satisfied if the hydrogen and acceptor atom are positioned within a specified distance. However, these bias forces do not enforce the directional component of hydrogen bonds: they require a minimal angle of 110°, with the maximum strength interaction occurring at 180° [2,13]. Although several hydrogen bonding potentials exist, most optimize protein interactions or protein-nucleic acid interactions[5,8]. In addition, if force fields implicitly capture the directionality of hydrogen bonding, then the origin of the observed stacking phenomena requires further investigation. Here, we compare gas-phase quantum mechanical and molecular mechanics energy calculations to determine the accuracy of the classical representation of nucleic acid hydrogen bonding.

Density Functional Theory:

A quantum state, represented in real space by the complex wavefunction Ψ, contains all the information for a system. The time-dependent non-relativistic Schrodinger equation is

$$\hat{H}\Psi(\overrightarrow{r_1},\overrightarrow{r_2}\ldots\overrightarrow{r_N},\overrightarrow{R_1},\overrightarrow{R_2}\ldots\overrightarrow{R_M}) = E\Psi(\overrightarrow{r_1},\overrightarrow{r_2}\ldots\overrightarrow{r_N},\overrightarrow{R_1},\overrightarrow{R_2}\ldots\overrightarrow{R_M})$$

Where $\hat{H}$ is the Hamiltonian operator or eigenfunction, r represents the 3N spatial coordinates and N spin coordinates for the N electrons, R represents the 3M spatial coordinates of the M nuclei and E is an energy eigenvalue. The square modulus of the wavefunction represents a probability density function of locating particles at specific coordinates. The Hamiltonian represents an energy constructed by many potential and kinetic terms

$$\hat{H} = \underbrace{-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2}_{\text{Electron KE}} \underbrace{-\frac{1}{2}\sum_{A=1}^{M}\frac{1}{m_A}\nabla_i^2}_{\text{Nuclei KE}} \underbrace{-\sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}}}_{\substack{\text{Electron-Nuclei}\\ \text{Attractive}\\ \text{Potential}}} + \underbrace{\sum_{i=1}^{N}\sum_{J>1}^{N}\frac{1}{r_{ij}}}_{\substack{\text{Electron-Electron}\\ \text{Repulsion}\\ \text{Potential}}} + \underbrace{\sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{r_{ij}}}_{\substack{\text{Nuclear Repulsion}\\ \text{Potential}}}$$

where ∇ is the differential del operator, r is the distance, and Z is the nuclear charge.

Since the Schrodinger equation cannot be solved for many-body problems, many approximations have arisen so the electronic structure of larger molecules can be probed. The Born-Oppenheimer approximation states that since the nuclei are much larger than the electrons, we can consider them to be stationary particles in a field of moving electrons. Considering the kinetic energy of the nuclei to be 0 and the nuclear-nuclear repulsion term to be a constant, C, leads to the electronic wavefunction.

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{J>1}^{N}\frac{1}{r_{ij}} + C$$

The electronic repulsion term is a many-body problem for all elements except hydrogen, and cannot be solved for exactly. Different methodologies are defined by how the electronic repulsion term is approximated. Wavefunction based methods, like Hartree-Fock (HF), approximate Ψ using antisymmetric Slater determinants. Density functional theory (DFT) reduces the dimensionality of the problem by relating the energy to the probability density rather than the wavefunction. This is accomplished by rewriting the Schrodinger equation as a functional of the density. Hybrid methods that combine different percentages of HF and DFT, like PBE0 and B3LYP, also exist.

A basis set is a set of atomic centered functions that are used to approximate a molecular orbital or wavefunction of a single electron. We approximate the unknown functional form of the molecular orbital using a linear combination of functions that resemble the hydrogen atom wave functions. Increasing the number of functions used provides a better prediction of the molecular orbital, however, this also increases the expense of the computation.

Molecular Mechanics:

The molecular mechanics energy approximates the quantum energy by using a combination of the Leonard-Jones potential and the Coulomb potential. The 6-12 Leonard Jones potential, consisting of a repulsive term and an attractive term, approximates interactions between neutral atoms and molecules. This potential takes the form of

$$V_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

where ϵ is a measure of the attraction between two atoms or molecules, σ is the van der Waals distance between two nonbonding particles, and r is the distance between two particles. Two non-interacting atoms can be brought together with minimal energy until they are touching. Additional energy is required to bring the atoms closer together because of atomic overlap, which causes repulsions that are much greater than attractions at small distances. The Coulomb potential reflects the long-range electrostatics present in a system. This is easily modeled for two charged, point particles using

$$U = k\frac{q_1 q_2}{r}$$

where U is the potential, k is the Coulomb constant, q represents the charge of a point particle and r is the distance between two point particles.

Materials and Methods

Quantum calculations were used to analyze the angular dependence of the hydrogen bond in guanine-cytosine (GC) Watson-Crick base pairs. For simplicity, initial calculations were only performed on the bases, without inclusion of the phosphates or sugars. Then, the sugars were included to observe if any changes resulted from the inclusion of a dipole. The general process started with a geometry optimization, proceeded with rotation of coordinates and culminated after measurement of quantum and molecular mechanics energies. The Q-Chem quantum chemistry software was used for all geometry optimizations and quantum energy calculations. Python and MATLAB were used to plot all graphs.

Geometry Optimization:

Initial coordinates for the hydrogen bonded bases were retrieved from Šponer et al[15]. The nitrogenous bases were geometry optimized with a PBE0 hybrid DFT-HF method and a D3 energy dispersion correction. The 6-316* basis set was implemented for these calculations because it was a good compromise between the large number of energy calculations and reasonable computational times.

For the nucleoside calculations, two hydrogen bonded nucleotides were generated using the Molecular Operating Environment (MOE) software. Then, the phosphates were manually removed, and sugars capped using Pymol. The structure was geometry optimized using the same parameters as above, except some torsions were constrained to 0° and -180° to ensure planarity.

Rotations:

The same rotation parameters were used for the nitrogenous bases and the nucleosides. Guanine coordinates were rotated about the “y axis,” an axis formed by H1(G) and H21(G), from

180° to -180° in 5° increments, as shown in Figure 5. The initial goal was to rotate the molecule by 180° to achieve a stacked structure, perform a geometry optimization to achieve the optimum stacking distance and then rotate the stacked guanine by 180° to obtain the hydrogen bonded structure. This would result in two slightly different energy curves: one representing motion from the optimized hydrogen bonded position to the stacked position and the other representing motion from the optimized stacked position to the hydrogen bonded position. The average of both curves would be the representative energy curve.

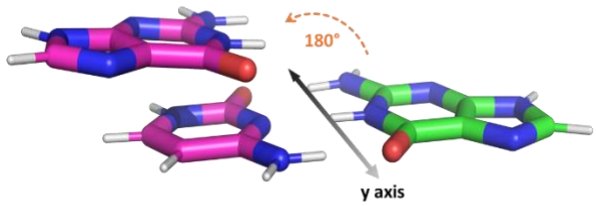

Figure 5. Guanine rotation 180° about the "y axis." The initial, geometry optimized position of the GC base pair is also shown.

Two other rotations were performed. The guanine coordinates were also rotated about the "x axis", the axis intersecting the middle of the N3(C) – H1(G) bond, from an angle of 0° to 360° in 5° increments. A sliding case was also computed to more effectively sample the stacking and hydrogen bonding space. H1 was rotated from -122.5° to 122.5° in 2.5° intervals while the distance between the H1 and N3 was varied between 1.4 Å and 2.4 Å in 0.1 Å increments. In this case, the H1 hydrogen was rotated half the angle about the "y axis" than the entire base to ensure that the middle hydrogen bond was kept at an equilibrium distance of 1.9 Å. Schematic representations of the "x axis" and sliding rotations is shown in Figure 6 and Figure 7, respectively.

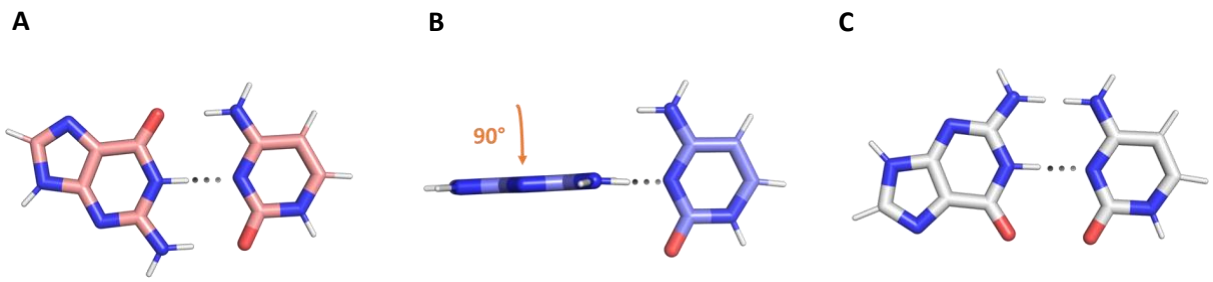

Figure 6. Guanine rotation about the "x axis." A) This is the initial, geometry optimized position of the cytosine and guanine base pair. The distance between the nitrogen donor hydrogen and the accepter is 1.9 Å. B) The guanine is rotated 90° about the x axis. C) The guanine is rotated 180° about the x axis.

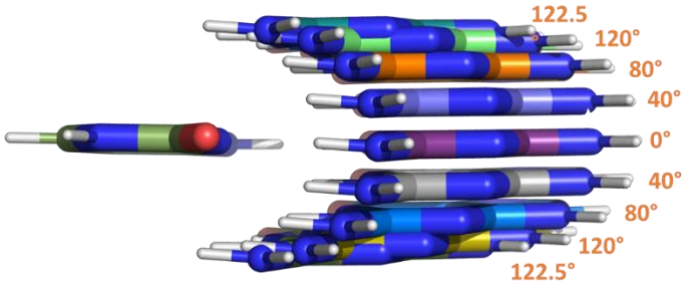

Figure 7. "Base sliding" at a N3-H1 distance of 1.9 Å. Guanine rotations at several different angles are shown.

Energy Calculations:

Quantum and molecular mechanics energy calculations for the nucleobases were conducted at each rotation point. Quantum mechanics (QM) energy calculations were conducted with a D3 correction and the pc-2 basis set. Unless otherwise stated, all geometry optimizations were conducted using radial coordinates. The basis set superposition error (BSSE) job type was implemented find the difference of the calculated the energy of the bases together and the energy of the bases separately. The difference represents the non-bonded interaction energy of the two bases. The molecular mechanics (MM) energies were determined by finding the average sigma

and epsilon parameters to calculate the Leonard-Jones and Coulomb potentials for each pairwise interaction. Charges were obtained using the Amber-99 force field. The same parameters were used for the nucleosides, except the sliding energy calculations were conducted using the cc-pvdz basis set.

Results and Discussion

Cytosine and Guanine Nucleobases:

Geometry Optimization afforded a planar structure with hydrogen bonding distances of 1.9 Å (N3(C)-H1(G)), 1.9 Å (02(C) – H21(G)), and 1.7 Å (H41(C) – O6(G)). QM and MM energies for "y axis" rotation are shown in Figure 8. Rotations of more than 150° caused some overlap between van der Waals radii, resulting in highly repulsive energies. This was counteracted by truncating all graphs to comparable energy scales.

The resultant energy curves for the QM and MM calculations were very similar. The global energy minima at 0° rotation confirms that in both methods, the lowest energy is attributed to the state where the hydrogen bonding interaction is maximized. However, the energy of the hydrogen bonding structure is more favorable in QM than in MM. The QM data appears to follow a linear relationship about the minima whereas the MM data appears to be more parabola-like. The steeper slope of the QM energy well may indicate that the system may be driven faster towards the stable hydrogen bonding equilibrium than in the MM case. In addition, the statistical population about the QM minimum free energy structure would be confined to a tighter geometry space because of smaller angular oscillations at the minimum as compared with the broader, harmonic minimum for the MM case. The MM energy curve also appears to be skewed whereas the QM potential is symmetric. This indicates involvement of asymmetry or directional dependence in the MM calculations since angular rotations in one direction may not be equivalent to rotations in the opposite direction.

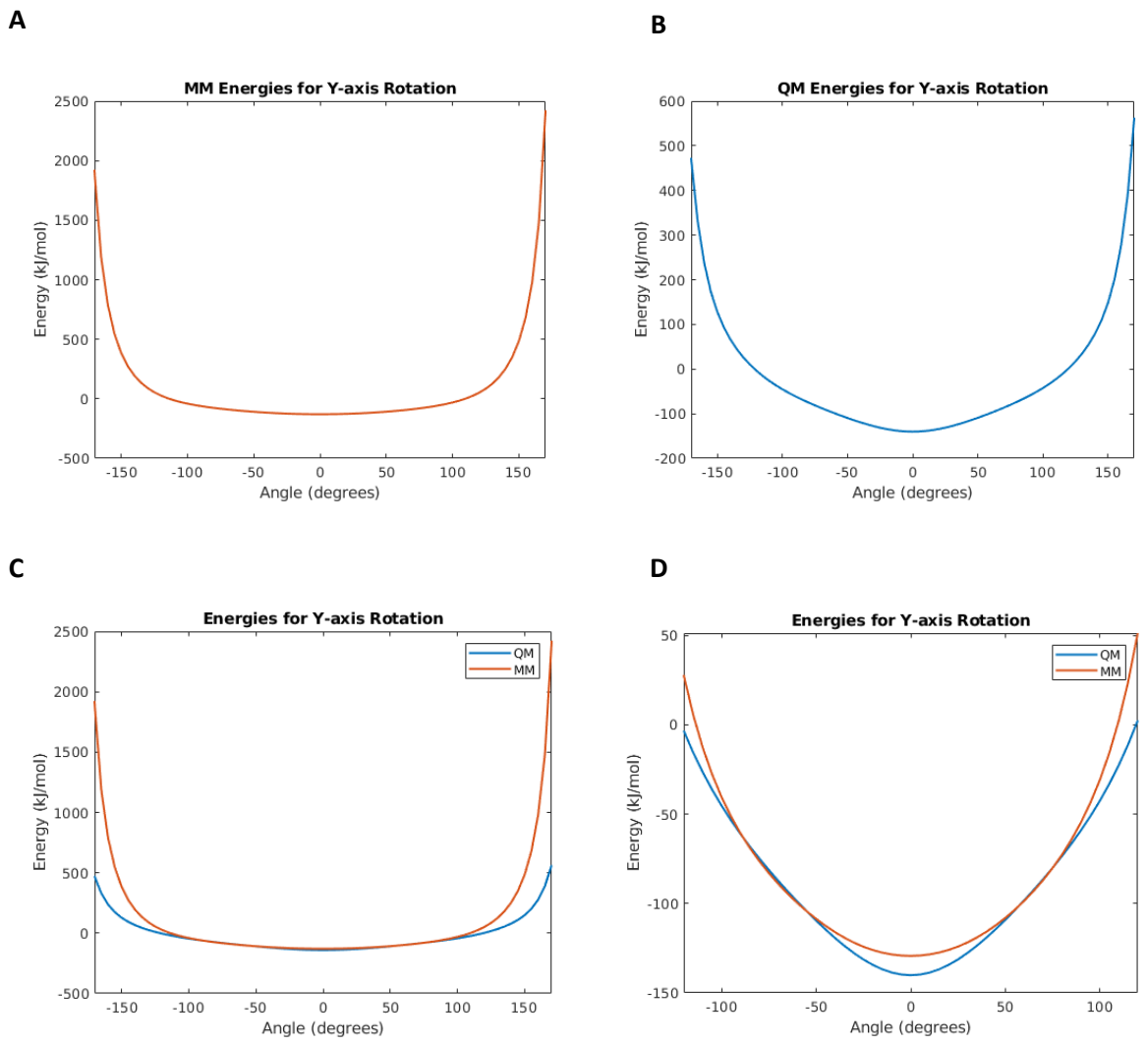

Figure 8. QM and MM energies for nucleobase rotation about the "y axis". A) MM energies for rotations from -160° to 160° about the "y axis." B) QM energies for rotations from -160° to 160° about the "y axis." C) Comparison of QM and MM energies. D) Truncated comparison of QM and MM energies.

Geometry optimization was conducted on coordinates rotated 180° about the "y axis." Surprisingly, instead of observing a local minima attributed to base stacking, the guanine returned to the hydrogen bonding position. Representative images from the optimization trajectory are shown in Figure 9. A step along the gradient in radial coordinates results in large geometry changes in cartesian coordinates. This relatively large step size may bypass smaller, local minima.

Unfortunately, geometry optimization with cartesian coordinates also converged to the hydrogen bonding state. Then, it was thought that perhaps achieving an optimized stacking distance might uncover the stacking minima. A fixed restraint geometry optimization, where the bases were restricted to motion in the z direction, provided an optimum stacking distance of 3.1 Å. However, the proceeding geometry optimization with cartesian coordinates resulted in a return to the hydrogen bonded state.

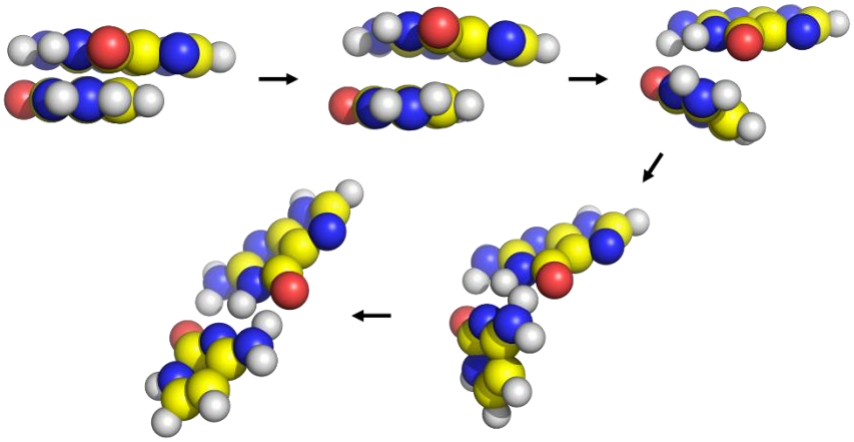

Figure 9. Geometry optimization of 180° rotated structure. Coordinates from a geometry minimization for a structure rotated 180° about the "y axis" are shown. The trajectory evades a plausible stacking minimum to reach the hydrogen bonding minimum.

Rotation about the x-axis provided similar energy curves for the QM and MM methods, as shown in Figure 10. The maxima at 180° rotation for both methods is expected due to steric clashes between the spatially close cytosine and guanine carbonyl groups. Effects of repulsion and stabilization were more amplified in the QM calculations than in the MM calculations.

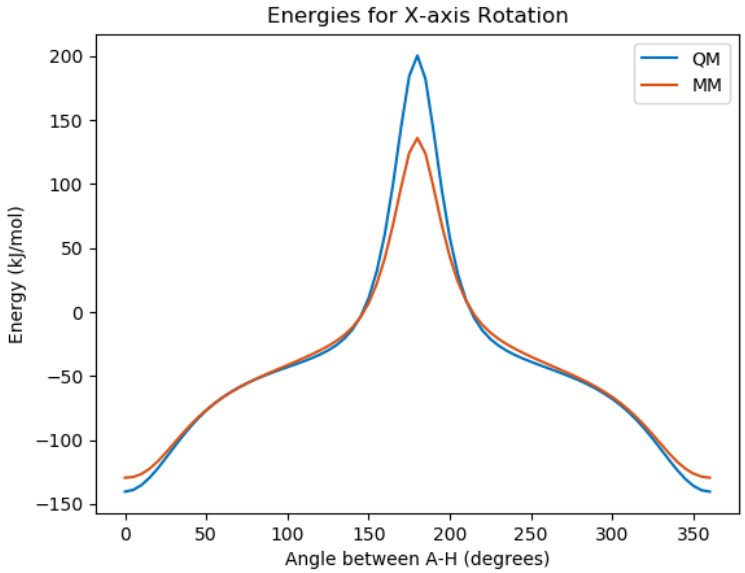

Figure 10. QM and MM energies for rotation about the "x axis."

The greatest differences between the QM data and MM data were observed in the sliding case. QM and MM data comparisons are shown in Figure 11. QM energies remain favorable at larger variations of N1-H1 hydrogen bonding distances whereas MM energies become repulsive at much smaller distances. This indicates that quantum methods allow for a larger range of hydrogen-acceptor distances for effective hydrogen bonding interactions than in MM. As observed in the rotations about the "y axis," the QM energies are more symmetric and ovular, and the MM energies are slightly skewed.

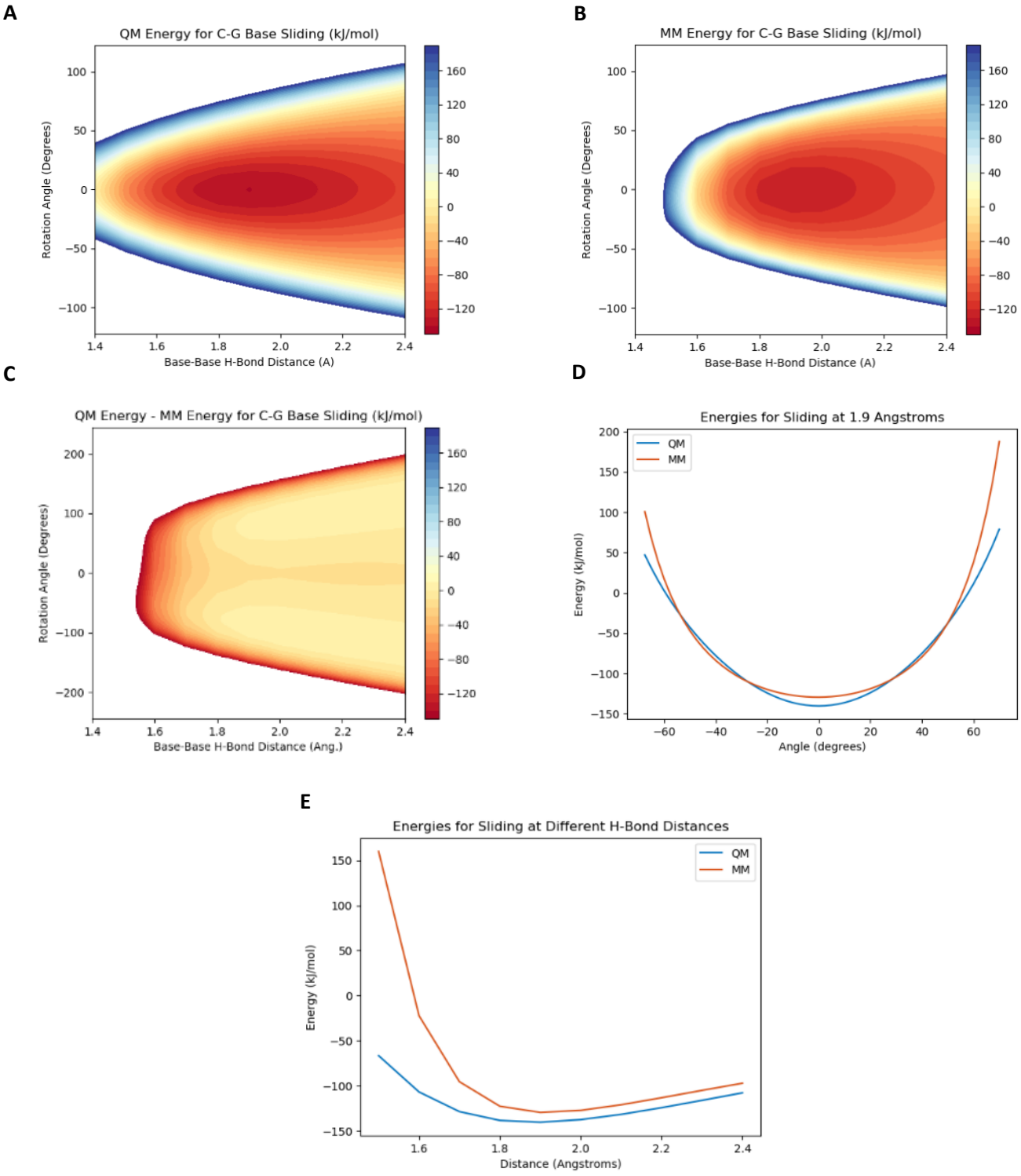

Figure 11. QM and MM energies for base sliding. A) QM energies for base sliding at different acceptor-hydrogen distances and angles. B) MM energies for base sliding at different acceptor-hydrogen distances and angles. C) Differences between the QM energies and the MM energies. D) QM and MM energies for different acceptor-hydrogen distances at the maximum hydrogen bonding angle. E) QM and MM energies for rotation at an acceptor-hydrogen distance of 1.9 Å.

Cytosine and Guanine Nucleosides:

Addition of a dipole did not have a large impact on the rotation results, except for raising the QM and MM minimum energies for the system. Figure 12 shows MM and QM energies produced by rotation about the "y axis." Repulsion and MM energy skewing were further enhanced in the nucleoside. QM and MM energies both displayed a larger increase in energy per degree increase in angle. QM nucleoside and nucleobase, and QM and MM nucleoside energies for rotation about the "x axis" are compared in Figure 13. The QM energy for rotation about the "x-axis" was mostly unaffected by the addition of the dipole. One slight difference is that a more repulsive maxima was observed in the nucleoside. Nucleoside sliding was also very similar to nucleobase sliding. The QM and MM energy countour plots and curves are shown in Figure 14. QM energy gradients are more ovular, and QM and MM energies are more repulsivse in the nucleoside than in the nucleobase.

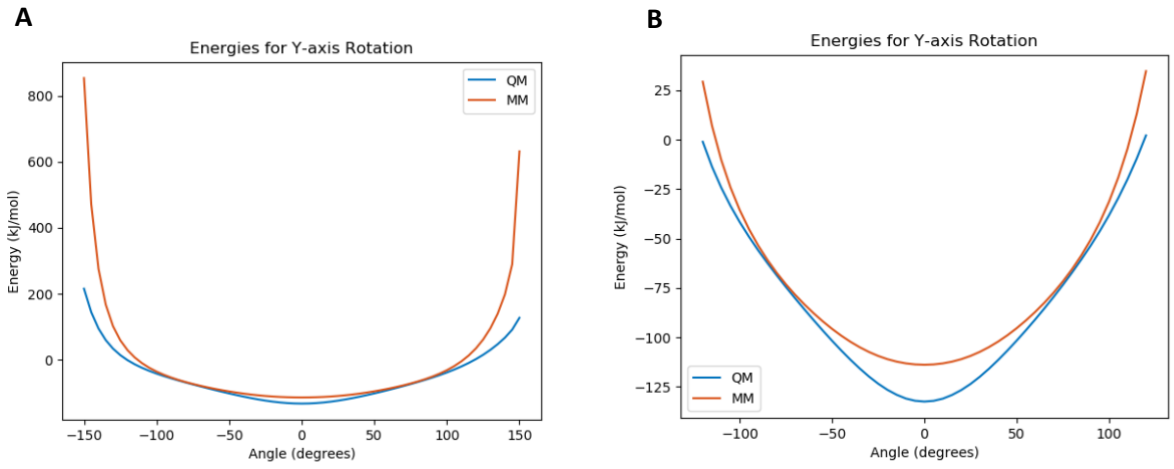

Figure 12. QM and MM energies for nucleoside rotation about the "y axis". A) Comparison of QM and MM energies for rotations of -160° to 160° about the "y axis." B) Truncated Comparison of QM and MM energies for rotation about the "y axis."

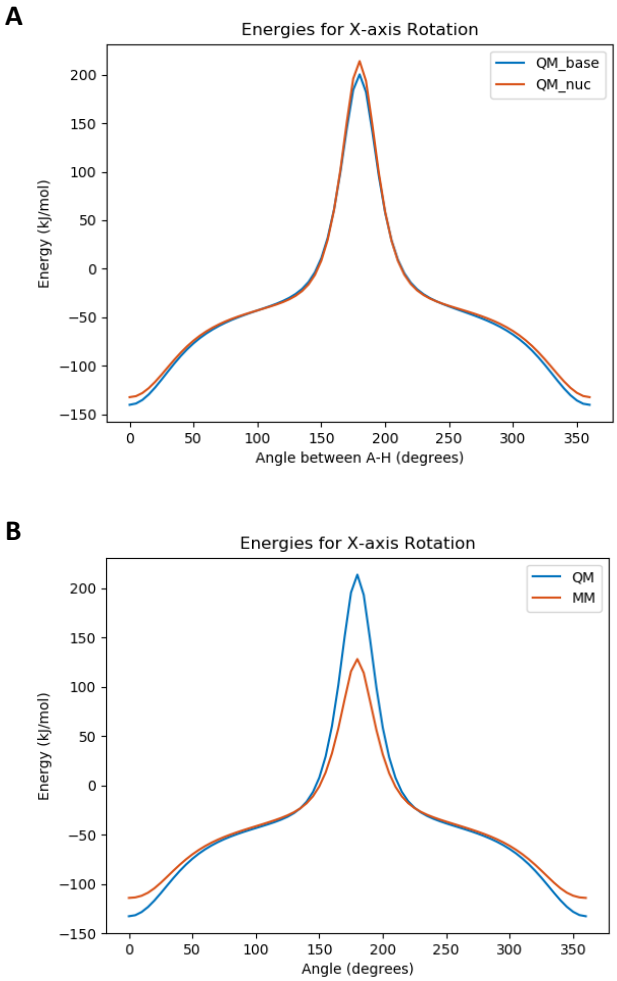

Figure 13. QM and MM energies for nucleoside rotation about the "x axis." A) Comparison of QM energies for the nucleobase and nucleoside. B) Comparison of QM and MM energies.

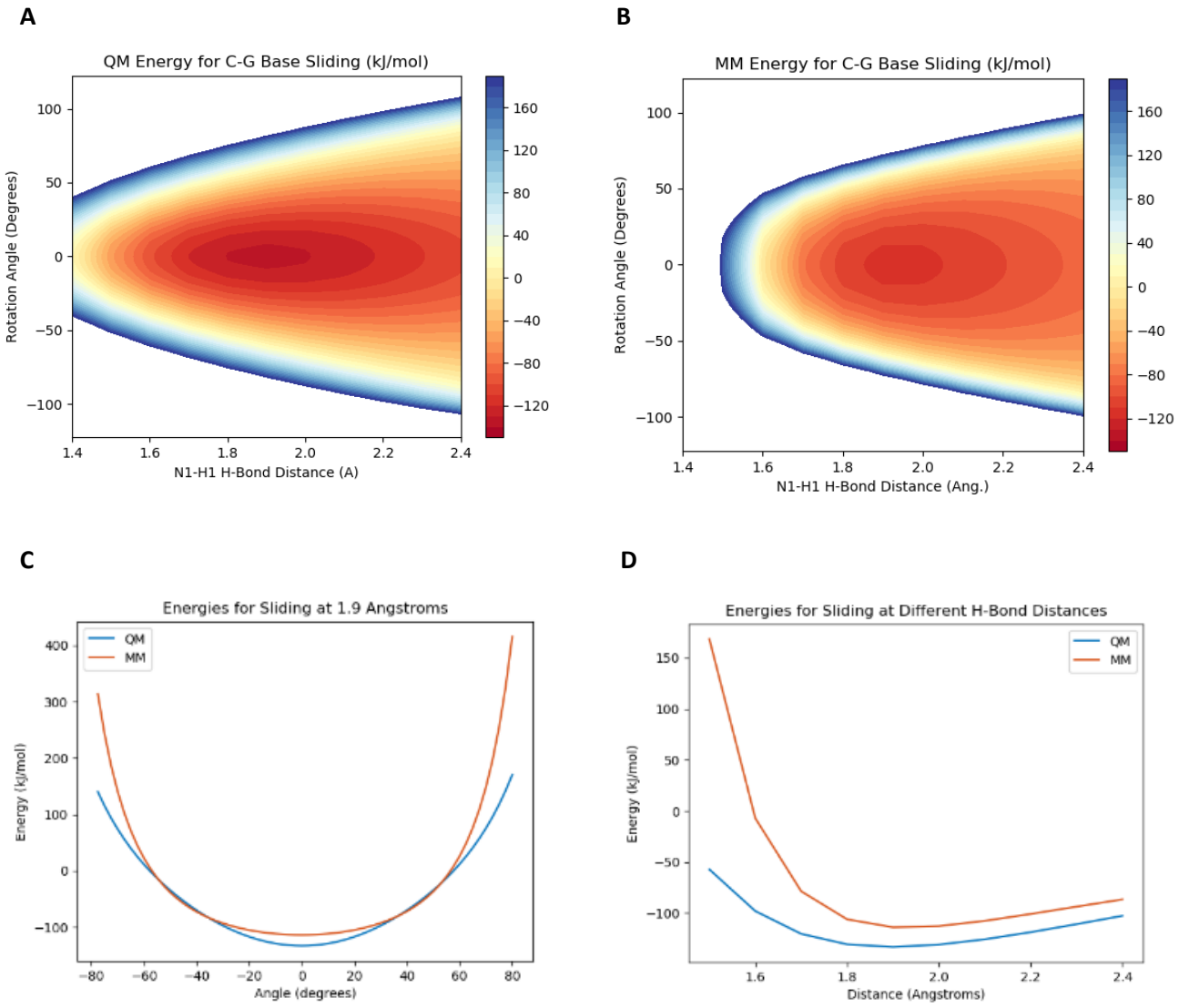

Figure 14. QM and MM energies for Nucleoside Sliding. A) QM energies for base sliding at different acceptor-hydrogen distances and angles. B) MM energies for base sliding at different acceptor-hydrogen distances and angles. C) QM and MM energies for different acceptor-hydrogen distances at the maximum hydrogen bonding angle. D) QM and MM energies for rotation at an acceptor-hydrogen distance of 1.9 Å.

Conclusions

An accurate depiction of hydrogen bonding is essential to generating more predictive and descriptive biomolecule models. We sought to compare the classical interpretation for directional hydrogen bonding to the quantum description using CG base pairs. For all rotations, the appropriate points of maximum repulsion and attraction overlapped in location, but differed in magnitude. The MM energy curve was observed to be skewed for the nucleobase and the effect was enhanced by the presence of the dipole in the nucleoside. Quantum mechanically, the hydrogen bonding interaction is effective at a larger range of hydrogen-acceptor distances, but requires smaller angular deviations. MM describes the opposite situation: hydrogen bonding distances are more rigid and the angular component is observed to more flexible. The addition of a dipole had negligible effects except for increasing the energy of the system.

DFT calculations did not uncover a stacking minima since all geometry optimizations resulted in a return to the planar hydrogen bonded state. Although the AMBER force field has previously been observed to overestimate gas phase base stacking energies, MM calculations only indicated sizeable repulsions at large angles of rotation[9]. Since stacking interactions are electrostatic and hydrophobic in nature, solvent effects may be required to reveal this minima using our protocol.

In the future, we would like to expand this analysis with nucleotides and conduct similar analyses with longer double strands, in order to obtain the inter and intra strand interactions that also contribute to effective stacking. Due to the similarity between the QM and MM data in gas phase, the stacking preference near large bulges may have other origins than lack of directional hydrogen bonding restraints. Force distribution analysis will be used to study pairwise forces in the molecule to identify other stabilizing and destabilizing interactions. Although directional hydrogen bonding potentials for restraints would provide a better description of the system than

current potentials that only use distance restraints, QM energy calculations do not provide much additional information about hydrogen bond directionality that is not currently encompassed in MM simulations.

References

Abraham, M., Hess, B., van der Spoel, D., & the GROMACS Development Team. (2016). *GROMACS user manual version 2016.4.* Retrieved from www.gromacs.org

Arunan, E., Desiraju, G. R., Klein, R. A., Sadle, J., Scheiner, S., Alkorta I., Clary, D. C., Crabtree, R. H., Dannenberg, J. J., Hobza, P., Kjaergaard, H. G., Legon, A. C., Mennucci, B., & Nesbitt, D. J. (2011). Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure and Applied Chemistry, 83*(8), 1637-1641.

Chemical Computing Group. (2019). Molecular operating environment. Montreal, Quebec, Canada: Chemical Computing Group. Retrieved from www.chemcomp.com.

Chen, A. A., & García, A. E. (2013) High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proceedings of the National Academy of Sciences U.S.A, 110*(42), 16820-16825. doi:10.1073/pnas.1309392110

Ebrahimi, P., Kaur, S., Baronti, L., Petzold, K., & Chen, A. A. (2019) A multi-dimensional Hamiltonian replica-exchange method for simulating RNA folding using experimental constraints. *Methods*. Advance online publication. doi: 1016/j.ymeth.2019.05.001

Fabiola, F., Bertram, R., Korostelev, A., & Chapman, M. S. (2002) An improved hydrogen bond potential: Impact on medium resolution protein structures. *Protein Science*, 11, 1415-1423.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD-Visual molecular dynamics. *Journal of Molecular Graphics and Modeling*, 14, 33-38.

Koch., W., & Holthausen, M.C., (2001) A chemist's guide to density functional theory. New York, NY: Wiley-VCH.

Kortemme, T., Morozov, A. V., & Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *Journal of Molecular Biology*, 326, 1239-1259.

Kruse, H., Banas, P., & Sponer, J. (2019). Investigations of stacked DNA base-pair steps: Highly accurate stacking interaction energies, energy decomposition, and many body stacking effects. *Journal of Chemical Theory and Computation, 15*(1), 95-115.

MATLAB. (2018). *MATLAB version 9.5 (2018b)*. Natick, MA: The MathWorks Inc. Retrieved from www.mathworks.com

Python Core Team. (2018). *Python: A dynamic, open source programming language, version 3.7.* Python Software Foundation. Retrieved from http://www.python.org

Shahi, A., & Arunan, E. (2016). Why are hydrogen bonds directional? *Journal of Chemical Sciences, 128*(10), 1571-1577.

Shao, Y., Gan, Z., Epifanovsky, E., Gilbert, A. T. B., Wormit, M., Kussmann, J. & the Q-Chem Development team. (2015). Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics,* 113, 184-215. doi:10.1080/00268976.2014.952696.

Šponer, J., Jurečka, P., & Hobza, P. (2004) Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *Journal of the American Chemical Society,* 126(32), 10142-10151.

Schrodinger. (2019). *The PYMOL molecular graphics system, version 2.3.* New York, NY: Schrödinger, LLC.

Yu, A., Gasper, P., Strobel, E., Watters, K., Kaur, S., Chen, A. A., & Lucks, J. B. (Under review). Computationally reconstructing cotranscriptional RNA folding pathways from

experimental data reveals rearrangement of non-native folding intermediates. Preprint. doi:10.1101/379222

Appendix: Table of Basis Sets

Basis Set	Number of Basis Functions	Energy	Wall Time (s)
6-31Gs	667	-2415.5210340231	79.78
6-31Gss	739	-2415.5802935176	86.93
6-311Gs	859	-2416.0203674544	218.82
6-311Gss	931	-2416.0825545710	239.71
6-31+Gs	831	-2415.5965339238	330.99
6-31++Gss	927	-2415.6546117697	366.16
6-311+Gs	1023	-2416.0620861807	603.24
6-311++Gss	1119	-2416.1242364526	644.51
cc-pvdz	739	-2415.6911968057	167.34
cc-pcvdz	909	-2415.7414684305	234.32
cc-pvtz	1799	-2416.3131722037	1275.26
cc-pcvtz	2389	-2416.3731946396	2165.15
cc-pvqz	3714	-2416.4885392642	13900.17
cc-pcvqz	5139	-2416.5245336929	21134.94
aug-cc-pvdz	1245	-2415.8363316366	2170.52
aug-cc-pcvdz	1415	-2415.8857616299	2284.39
aug-cc-pvtz	2859	-2416.3254947457	22438.30
aug-cc-pcvtz	3449	-2416.3521703614	32979.51
pcseg-0	421	-2409.3098466403	45.16
pcseg-1	739	-2415.1308677461	113.23
pcseg-2	1799	-2416.3811524497	1397.90
pcseg-3	4032	-2416.5454689073	23425.26
aug-pcseg-0	609	-2409.5269351366	298.34
aug-pcseg-1	1245	-2415.2586509688	1958.29
aug-pcseg-2	2859	-2416.3872801807	32288.74
pc-0	421	-2409.4150476172	39.68
pc-1	739	-2415.1752592252	142.98
pc-2	1799	-2416.4341716601	1478.56
pc-3	4216	-2416.5488849880	27131.68