Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations

Alexandre V. Morozov, Tanja Kortemme, Kiril Tsemekhman, David Baker

Andrew Kohlway and Laura Thomas
Mentor: Dr. Jeffry D. Madura
OUTLINE

- Overview of Hydrogen Bonding
- Approaches to Computational Modeling of Hydrogen bonding energy landscapes
- Purpose
- Methods
- Discussion and Results
- Conclusion
- My Research
Hydrogen Bonds

- Partially Covalent Interactions between a hydrogen atom covalently bound to an electronegative atom and an electronegative acceptor atom.

- Important role in
  - Defining structure and function of proteins
  - The formation of secondary structures
  - Energetics of protein folding
Computational Understanding

- Three approaches to modeling hydrogen bonding energy landscapes:
  - Quantum mechanical calculations on small systems
  - Molecular Mechanics Force Fields
  - Knowledge Based Potentials derived from PDB (Protein Data Bank)
Pros and Cons

- Quantum Mechanical “ab initio”–
  - Most general and fundamental → based on first principles
  - Carried out for only small systems → time-consuming
  - May not relate well to larger macromolecules

- Molecular Mechanics –
  - No size constraints
  - Not as accurate as quantum methods because of simplifications

- Knowledge-Based Potentials –
  - Pairwise potentials used so 3 and 4 body systems not taken into account.
  - Physical origin of the PMF not directly inferred
  - In other words specific Coulombic, Van der Waals, etc. interactions
  - Limited by resolution of structure
Quantum Mechanical Calculations

- $\Psi = \text{the wavefunction of a system, a function of all coordinates of all of the particles in the system and time}$
- $\Psi^2 = \text{the probability density for the electrons within a system}$
- The Schrödinger equation
  - $\hat{H} \Psi = E \Psi$
  - Solving this equation for $\Psi$ can yield the energy and ultimately other properties of a system
  - Cannot solve analytically for more than a one electron system
- Molecular Orbital theory decomposes $\Psi$ into combinations of molecular orbitals
- Wavefunction $\rightarrow$ Molecular Orbitals $\rightarrow$ Atomic Orbitals $\rightarrow$ Basis Functions $\rightarrow$ Gaussians

\[
\Psi = f(\psi_1, \psi_2, \psi_3, \psi_M) \quad \phi_i = \sum_{j=1}^{k} d_{uk} \chi_k \\
\psi_i = \sum_{\mu=1}^{N} c_{\mu i} \phi_{\mu} \\
\chi_k = g(\alpha, r) = c x^n y^m z^l e^{-\alpha r^2}
\]
Hartree-Fock Method

- So how does one solve for the molecular orbital expansion coefficients?
- HF uses the variational principle, which states that for any ground state function of the electronic coordinates ($\Phi$), $E(\Phi) > E(\Psi)$
- The energy of the exact wavefunction is a lower bound on the energies calculated by any other normalized antisymmetric function
- The problem is then to solve for the coefficients that minimize $E(\Phi)$
- Roothaan-Hall equation
  - $FC = SC\varepsilon$
    - $F = $ Fock matrix
    - $\varepsilon = $ orbital energy matrix (energy of one specific electron)
    - $S = $ overlap matrix = overlap between orbitals
    - $C = $ coefficient matrix
Møller-Plesset Perturbation Theory

- Includes electron correlation
- The Hamiltonian is divided into two parts
  - $H = H_0 + \lambda V$
    - $H_0$ is the unperturbed component
    - $\lambda V$ is a small perturbation to $H_0$
    - $\lambda V$ is a two electron integral term – incorporates electron correlation
- MP2 computes the second order correction to the energy of the system
  - $E_i^{(0)} = \int \Psi_i^{(0)} H_0 \Psi_i^{(0)} \, dT \rightarrow$ unperturbed energy
  - $E_i^{(1)} = \int \Psi_i^{(0)} V \Psi_i^{(0)} \, dT \rightarrow 0$
  - $E_i^{(2)} = \int \Psi_i^{(0)} V \Psi_i^{(1)} \, dT \rightarrow 1^{st}$ non-zero correction term (overestimation)
  - MP3 and MP4 include third and fourth order corrections respectively
Density Functional Theory

- Models electron correlation based on functions of the electron density

\[ E_{\text{elec}} = E_T + E_V + E_J + E^{\text{XC}} \]

- \( E_T \) = kinetic energy term (electron motion)
- \( E_V \) = potential energy term
  - Nuclear – electron attraction
- \( E_J \) = electron – electron repulsion term
- \( E^{\text{XC}} \) = exchange – correlation term \( \rightarrow \) take guesses at this term
  - Other electron – electron interactions

- If we knew the \( E^{\text{XC}} \) term DFT would be an exact expression for the energy.
Molecular Mechanics

- Mathematical modeling relating the coordinates of a molecule to a certain energy.
- Developed from either experimental data or quantum mechanical calculations.
- Bonds and angles described by a spring equation \( \frac{1}{2}kx^2 \).
- Torsions described by a Fourier Series.
- Van der Waals interactions expressed by the Lennard-Jones 12-6 potential.
- Electrostatics described by Coulomb’s Law.
CHARMM Force Field

- Empirical Force Field based on both electron structure calculations and experimental data

\[ E_{\text{bond}} = k_b(b - b_0)^2 \]
PMF (Potentials of Mean Force)

- “Structural frequencies observed from a known structural training database are converted into contact or pairwise potentials.”
- The negative logarithm of the frequency of occurrence of a particular value of p is proportional to the interaction energy for that value of p.
- $E(p) \sim -\ln f_{\text{protein}}(p) \rightarrow$ Inverse Boltzmann Equation
- $E(p) = -\ln[f_{\text{protein}}(p)/f_{\text{ref}}(p)]$
- Used $10^\circ$ angle bins and $.05 \text{ Å}$ bins to collect data
PMF (Potentials of Mean Force)
Purpose

- Comparison of quantum mechanical calculations, molecular mechanical calculations, and protein structural analysis in studying the distance-angle dependence of hydrogen bonding.

Conclusion –

- close agreement between quantum mechanical calculations and protein structural analysis.
- Predictions based on molecular mechanics calculation were different from ab initio calculations and protein structure analysis.
Methods

- Needed to choose a small model that represents hydrogen bonds in protein side chains and main chains.
- NMA (N-methyl-acetamide) often used but the methyl groups affect the dimerization energy.
- Therefore, formamide was used in this study.
- Used an out-of-plane formamide dimer optimized by ab initio calculations with single hydrogen bonds rather than two.
- Also carried out electron structure calculations on the acetamide dimer.
Methods

- Potential energy surface sampled by varying one of four different parameters while keeping three fixed and allowing the molecule to relax for each sample conformation.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\delta_{HA}$ Å</th>
<th>$\psi^*$</th>
<th>$\theta^*$</th>
<th>$X^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFT$_a$</td>
<td>1.94</td>
<td>112.34</td>
<td>159.43</td>
<td>-177.51</td>
</tr>
<tr>
<td>DFT$_b$</td>
<td>1.94</td>
<td>112.91</td>
<td>161.57</td>
<td>179.78</td>
</tr>
<tr>
<td>MP2</td>
<td>1.97</td>
<td>110.49</td>
<td>155.33</td>
<td>-179.49</td>
</tr>
<tr>
<td>HF</td>
<td>2.10</td>
<td>138.16</td>
<td>170.94</td>
<td>-179.54</td>
</tr>
<tr>
<td>CHARMM27</td>
<td>1.82</td>
<td>170.25</td>
<td>170.83</td>
<td>-106.83</td>
</tr>
<tr>
<td>OPLS-AA</td>
<td>1.75</td>
<td>165.04</td>
<td>175.61</td>
<td>145.12</td>
</tr>
<tr>
<td>MM3-2000</td>
<td>1.98</td>
<td>121.16</td>
<td>161.07</td>
<td>149.63</td>
</tr>
<tr>
<td>PDB</td>
<td>1.93</td>
<td>115.00</td>
<td>175.00</td>
<td>175.00</td>
</tr>
</tbody>
</table>

Knowledge-based minima (PDB) are based on the most populated frequency bins (see Methods). The geometry parameters are defined in Fig. 1B.
Methods – Electronic Structure Calculations

- Density Functional Theory (DFT)
- MP2-second order Moller-Plesset perturbation theory using aug-cc-pVDZ
- HF-Hartree-Fock
Methods – Molecular Mechanics

- Carried out using TINKER 4.0
- Used:
  - CHARMM27 – Chemistry at Harvard Macromolecular Mechanics
  - OPLS-AA – Optimized Potentials for Liquid Simulations
  - MM3-2000 – Molecular Mechanics Force Field for small molecules
- Calculations done in the absence of solvent and at zero temperature
Methods – Protein Structures

- Statistics obtained from Dunbrack-culled PDB collection from 698 proteins with resolution of 1.6 Å.
- CHARMM19 used to add hydrogens to donor groups (H,N,Q,R,T).
- Hydrogens with rotatable bonds (Ser, Thr, Tyr, Lys) not considered.
- Unperturbed ionization constants at a pH value of 7.
- Used crystal structure of His, Asn, Gln, without taking into account swapping of N,O,C atoms → incorrect assignment = failure rather than distortion.
Parameters

- Six parameters needed to model the orientation of two molecules.
- Four are measured in this study, the others are the torsional angles around the hydrogen bond and hydrogen bond donor.
- The four-dimensional energy surface is difficult to work with, so a 1D approach was used where only one parameter is changed.
Dimerization Energy Landscapes

- DFT and MP2 calculations very similar, with steep minima in $\delta_{HA}, \Psi, \chi$.
- HF method less accurate because it neglects electron-electron correlations → overestimates hydrogen bonding lengths and favors linear acceptor atom angle.

Fig. 2. Formamide dimer hydrogen bonding energies (kcal/mol) vs. $\delta_{HA}$ (Å), $\psi$, $\theta$, and $\chi$ (°). Green (solid), DFT; blue (short dashes), MP2; cyan (dots), HF SCF; red (dots and dashes), CHARMM27; black (long dashes), OPLS-AA; magenta (long and short dashes), MM3-2000. All abbreviations are defined in the text.
Dimerization Energy Landscapes

- Force Fields model hydrogen bonding as electrostatic interactions $\rightarrow$ minimum energy with collinear atoms.

- Large difference between $\Psi$ dimerization energy for DFT calculations and the MM mechanical calculations.
  - CHARMM27 – $\Psi = 180^\circ$
  - OPLS-AA – shallow minimum allowing for close to linear bond.

- Conclusion $\rightarrow$ Need an orientation-dependent hydrogen bonding potential to correct this.

Fig. 2. Formamide dimer hydrogen bonding energies (kcal/mol) vs. $\delta_{HA}$ (Å), $\psi$, $\theta$, and $X$ (°). Green (solid), DFT; blue (short dashes), MP2; cyan (dots), HF SCF; red (dots and dashes), CHARMM27; black (long dashes), OPLS-AA; magenta (long and short dashes), MM3-2000. All abbreviations are defined in the text.
Dimerization Energy Landscapes

- Close relationship between DFT calculations and PMF.
- DFT and MP2 model hydrogen bonding interactions between amino acid side chains very well.
- Why? Short range covalent nature of hydrogen bonds ignores the surrounding protein.

Fig. 3. Formamide dimer hydrogen bonding energies (kcal/mol) vs. $\delta_{HA}$ (Å), $\psi$, $\theta$, and $X$ (*). Green (solid), DFT (same as in Fig. 2); black (dashes), DFT with constrained optimization; cyan (solid with filled circles), knowledge-based potential (negative logarithm of frequency distributions for side-chain–side-chain interactions in protein structures, binned as described in Methods).
Dimerization Energy Landscapes

As stated before:

- CHARMM27 and OPLS-AA force fields predict lowest energy at 180°
- Protein statistical analysis, DFT, and MP2 predict an angle close to 120°

- $sp^2$ hybridized oxygen atom has lone pairs 120° apart, however there is a low energy difference between 120° and 180° conformations.
- CHARMM dropped a hydrogen bonding potential as a result.

![Graphs showing dimerization energy vs. different angles](image)

Fig. 3. Formamide dimer hydrogen bonding energies (kcal/mol) vs. $\delta_{HA}$ (Å), $\psi$, $\theta$, and $X$ (°). Green (solid), DFT (same as in Fig. 2); black (dashed), DFT with constrained optimization; cyan (solid with filled circles), knowledge-based potential (negative logarithm of frequency distributions for side-chain–side-chain interactions in protein structures, binned as described in Methods).
120° or 180°

- The electronic structure calculations and PDB analysis could have a misleading agreement.
- More work needed to be done to verify that the close data correlation was indeed correct.
Verification of Results

- Optimized MP2 calculations and DFT calculations on formamide dimers at a $\Psi$ angle of 120° and 175°.
- Difference in energy between angles close to 1 kcal $\rightarrow$ larger than other studies using HF or semiempirical methods.
Verification of Results

- Need to check bias for Protein Structural Analysis.
- Always a possibility that there isn’t an energetic preference at 120°.
- Proteins often make two hydrogen bonds which may require the $\Psi$ angle to be at 120°.
Verification of Results

- Graphs show that the number of hydrogen bonds at specific angles for both single hydrogen bonds and all hydrogen bonds are very similar.
- As thought, the acceptor angle must be $120^\circ$ because of an energetic preference.
Conclusions

- Molecular Mechanics force fields are not accurate for side chain bonds.
- May work better for main chain hydrogen bonds which are slightly more linear → combination of steric constraints and bigger dipole.
Conclusions

- Side-chain-side-chain hydrogen bonds modeled well by formamide and acetamide dimers.
- Hydrogen bonds in protein structures follow closely the Boltzmann-like distribution.
Conclusions

- Newer force fields need to come out containing off atom charges, higher multipole interactions and/or electronic polarizability.
Conclusions

Applications of this hydrogen bond analysis includes:
- Protein structure prediction
- Fixed backbone sequence redesign
- Protein-protein docking
- Prediction of hot spots in protein interfaces
Future Applications

- New free energy functions using both knowledge-based potentials along with quantum mechanical calculations representing hydrogen bonding, pi-pi, and cation-pi interactions.

- “Together, the knowledge-based methods can guide the evaluation of the transferability of the ab initio results, and the quantum mechanical methods can then be used to augment and generalize the observed statistics.”
References

Acknowledgements

- Dr. Jeffry D. Madura
- BBSI
- Laura and Me for catering
- Duquesne University, University of Pittsburgh, Carnegie Mellon, and PSC
- NIH/NSF
- Raj
My Research

Docking
And
Homology Modeling