Taming the Hydroxyl Rotor

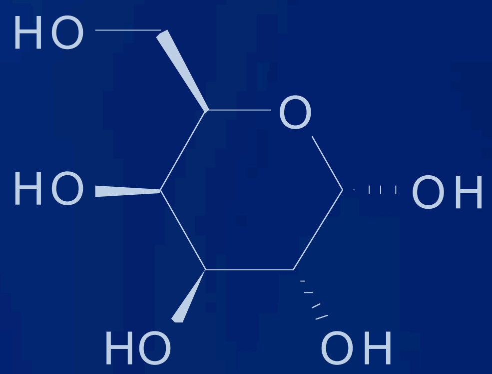
Roger Sayle
OpenEye Scientific Software

Motivation

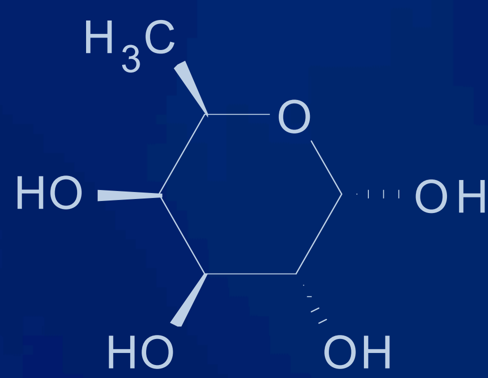
- Electrostatic Evaluation (WABE) Paper
- The Good, the Bad and the Ugly Validation
- First Loss of Signal: Arabinose Binding Protein
- Critical Hydroxyl Rotor Electrostatic Interaction
- What Can Be Done?

L-Arabinose Binding Protein

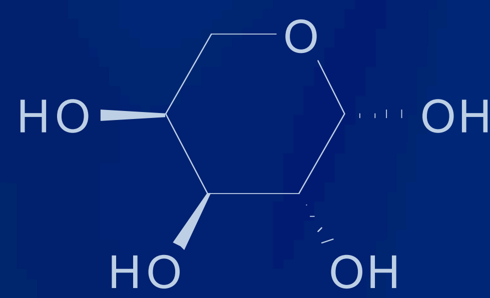
- Jian Shen and Florante A. Quiocho, "**Calculation of Binding Energy Differences for Receptor-Ligand Systems using the Poisson- Boltzmann Method**", *Journal of Computational Chemistry,* Vol. 16, No. 4, pp. 445-448, April 1995.



a-D-Galactose (5abp)

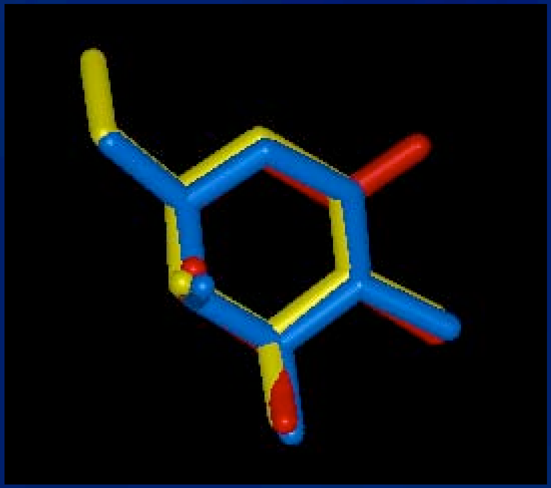


a-D-Fuctose (1abf)



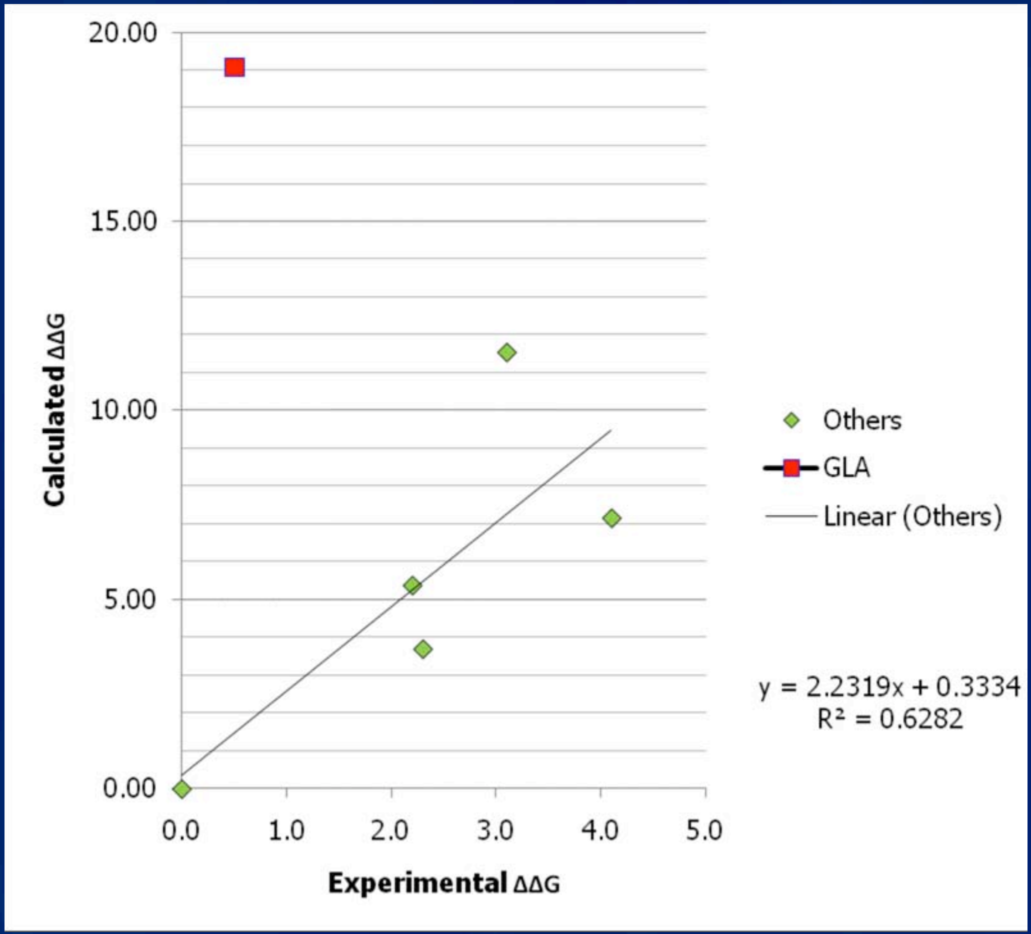
a-L-Arabinose (1abe)

ABP Isosteric Series

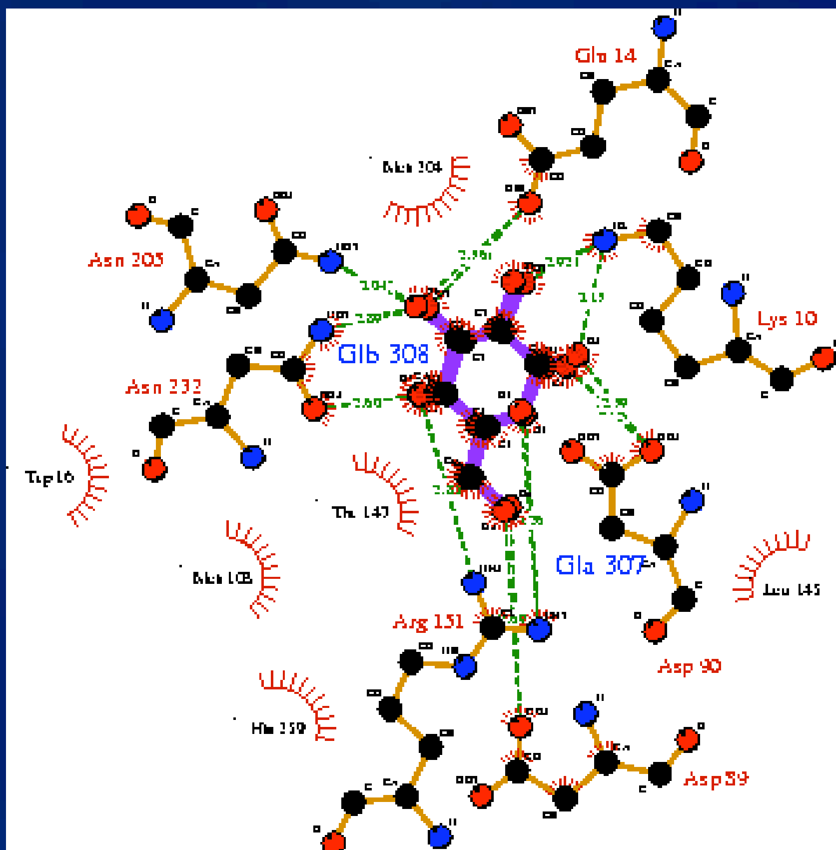


Yellow:	GLA	pdb5abp (1.8Å)
Blue:	FCA	pdb1abf (1.9Å)
		C_ RMS 0.170Å/305 points
Red:	ARA	pdb1abe (1.7Å)
		C_ RMS 0.195Å/305 points

Original ABP Results



ABP post-mortem

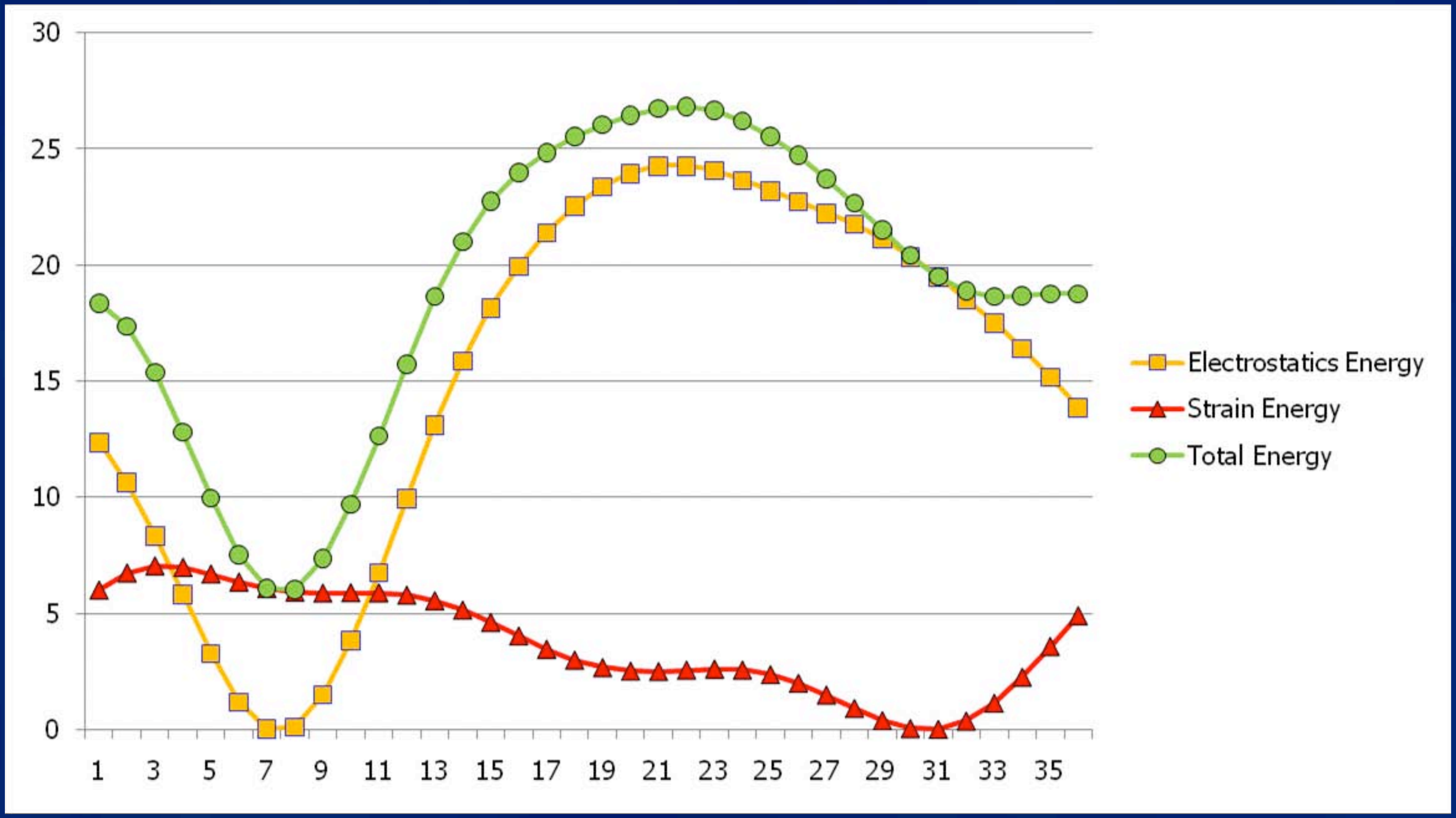


Explicit hydrogen location is important. Depending upon the orientation of the -OH rotor, the interaction unique to _-D-galactose can be favorable or not.

We were lucky this wasn't an issue with the heterocycles.

This problem is easily fixed.

Electrostatics and Strain



Conclusion

- Time to re-examine how 3D hydrogens are added to PDB structures and their ligands.
- In the OEChem toolkit this is performed by the "OESet3DHydrogenGeom" function.
- No contextual information is used in the placement of hydrogen atoms.

The Classic Paper

Mario Nardelli,

"A Calculator Program for Calculating Hydrogen Atom Coordinates",

Computers in Chemistry,

Vol. 6, No. 3, pp. 139-152, 1982.

M. Nardelli 1922-2004.

Three Degrees of Freedom

- Bond Lengths to Hydrogen, *l*
- Bond Angles to Hydrogen, _
- Torsion Angle to Hydrogen, _

Neutron vs. X-Ray Diffraction

- A complicating factor when positioning protons is their distance from the parent atom.
- Neutron diffraction (which images nuclei) and X-Ray diffraction (which images electron density) give significantly different results.
- One set of bond lengths to hydrogen for working with crystallography and another set of lengths when using quantum mechanics.

A Quick Analysis of CSD

	Neutron	X-Ray
• C	1.0794Å (0.0399)	0.9749Å (0.0506)
	N=3,683	N=2,631,6652
• N	1.0210Å (0.0296)	0.8988Å (0.0791)
	N=411	N=58,674
• O	0.9896Å (0.0669)	0.8918Å (0.1074)
	N=457	N=62,356

Bond Lengths To Hydrogen

H-Csp^3	1.10Å	H-H	0.75Å
H-Csp^2	1.08Å	H-F	0.90Å
H-Csp	1.06Å	H-Cl	1.27Å
H-N	1.00Å	H-Br	1.40Å
H-O	0.95Å	H-I	1.60Å
H-P	1.42Å	H-B	1.20Å
H-S	1.30Å	H-Si	1.48Å
H-As	1.52Å	H-Ge	1.53Å
H-Se	1.46Å	H-Te	1.70Å

Avoid Force Field Parameters

- The angle minima parameters used in force-fields frequently don’t match the actual lowest energy angle, theta.
- Mean *-O-H angles:
  - CSD (xray): 109.36° (N=60,581)
  - CSD (neutron): 110.28° (N=457)
  - MMFF94 validation set: 103.78° (N=274)

Case Study: Phenolic Oxygen

- MMFF Energy Minimum:
  - "CB"(37)-"OC=C"(6)- "HOCC"(29): 105.409°
  - Actual angle for minimized phenol: 108.4°
- Mean Data Set Angle (C.ar-O-H):
  - MMFF94 Validation Set 107.51° (N=27)
  - CSD X-Ray: 109.834° (s.d.=6.07, N=10,740)
  - CSD Neutron: 112.00° (N=39)
- Quantum Chemistry
  - B3LYP/6-31G*: 108.859°

Bond Angles To Hydrogen

- Assume ideal geometries.
  - Sp^3 hybridized: 109.5 degrees.
  - Sp^2 hybridized: 120 degrees.
  - Sp hybridized: 180 degrees.
- Special case terminal heavy atoms.
  - H-O-H: 104.5 degrees.
  - *-S-H: 102 degrees.
  - *=P-H: 103 degrees.
  - *=N-H: 112 degrees.

The Carboxylic Acid Problem

A common example of the potential problems with hydrogen placement is the case of protonated carboxylic acids.

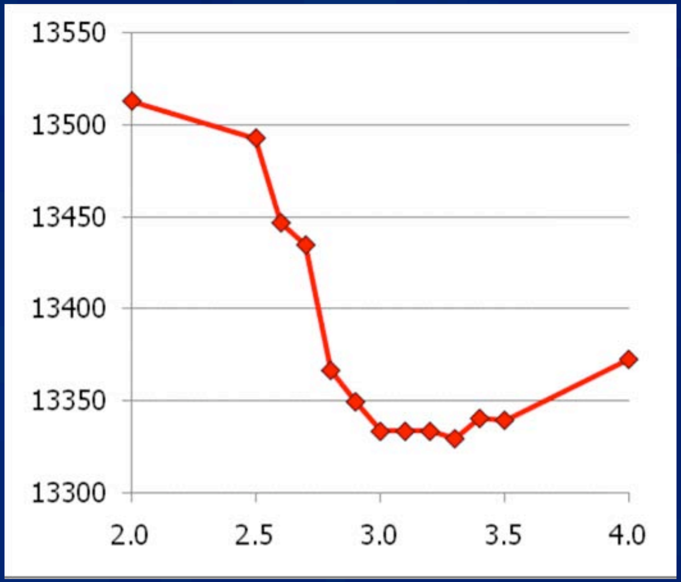
Torsional Preference Heuristics

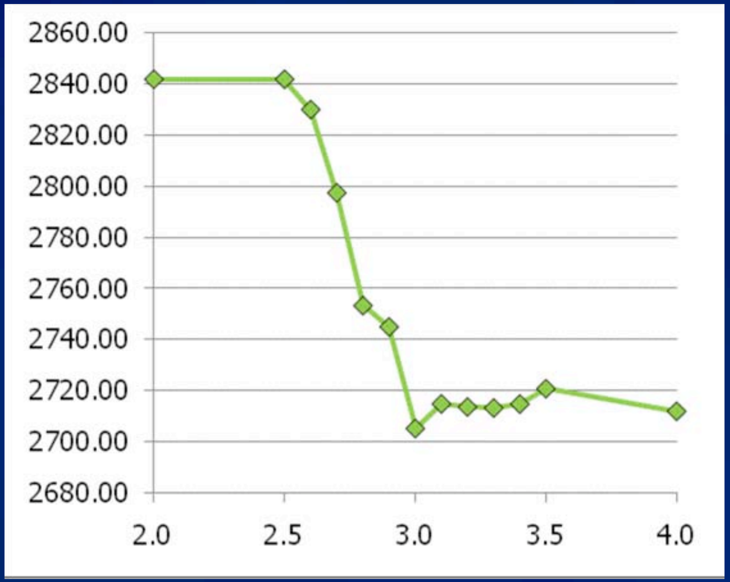
- Trans preference for sp^3-sp^3 chains, such as ethane and methanol.
- In-plane preference atoms attached to sp2 atoms, e.g. ethylene and phenol.
- Perpendicular preference for methyl groups attached to aromatic rings, e.g. toluene.
- Perpendicular 1-5 torsional preference for allenic conjugation, e.g. allene “C=C=C”.
- Cis preference to “acceptor” atoms, e.g. acetic acid or 2-hydroxypyridine.

Charge Proximity (H-Bonding)

- Strong short-range interactions often dominate the proton rotor preference.
- For each suitable rotor, we can identify the closest strong “acceptor” atom within a threshold distance.
- The current implementation uses a 3.0Å distance threshold to the donor “oxygen”.
- This searching is relatively infrequent as both hydroxyl rotors and strong acceptors are rare.

Distance Threshold





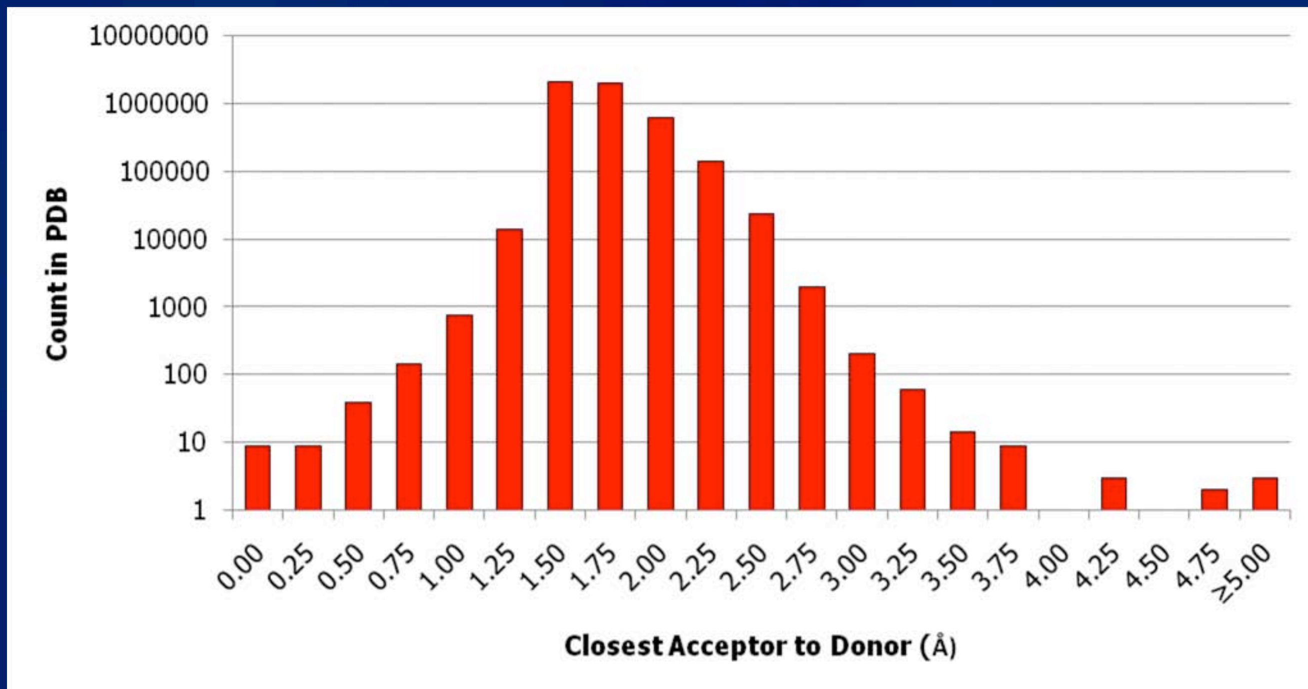
Total single-point MMFF94 energy of the MMFF94 validation suite (left) and 5ABP protein (right) as a function of electrostatic threshold distance (Å).

PDB Statistics (April 2008)

- Number of PDB files: 50,235
- Average number of atoms: 4711
- Average number of donors: 97 (2.06%)
- Average number of acceptors: 697 (14.80%)
- Average number of waters: 244 (5.17%)

- Files with donor & acceptor: 49,882
- Donor-acceptor distances: 4,881,842

Donor-Acceptor Histogram



4,881,545 of 4,881,842 (99.99%) of donors have an acceptor atom within 3.0Å!

MMFF94 Validation Suite

- To assess the quality of OEChem's 3D hydrogen placement algorithm, we can play "pin the tail on the donkey".
- Strip and regrow protons from the 761 small organic compounds in the MMFF validation set. 745 of these molecules have hydrogens.
  - MMFF94 total energy: 5,197 Kcal/mol.
  - Previous OEChem total energy: 19,363 Kcal/mol.
  - Improved OEChem total energy: 13,334 Kcal/mol.

Case Study: 5ABP Protein

- Heavy Atom Count: 2316
- Hydrogen Parents: 1467 (63.3%)
- Hydrogen Count: 2336 (1.59 avg)

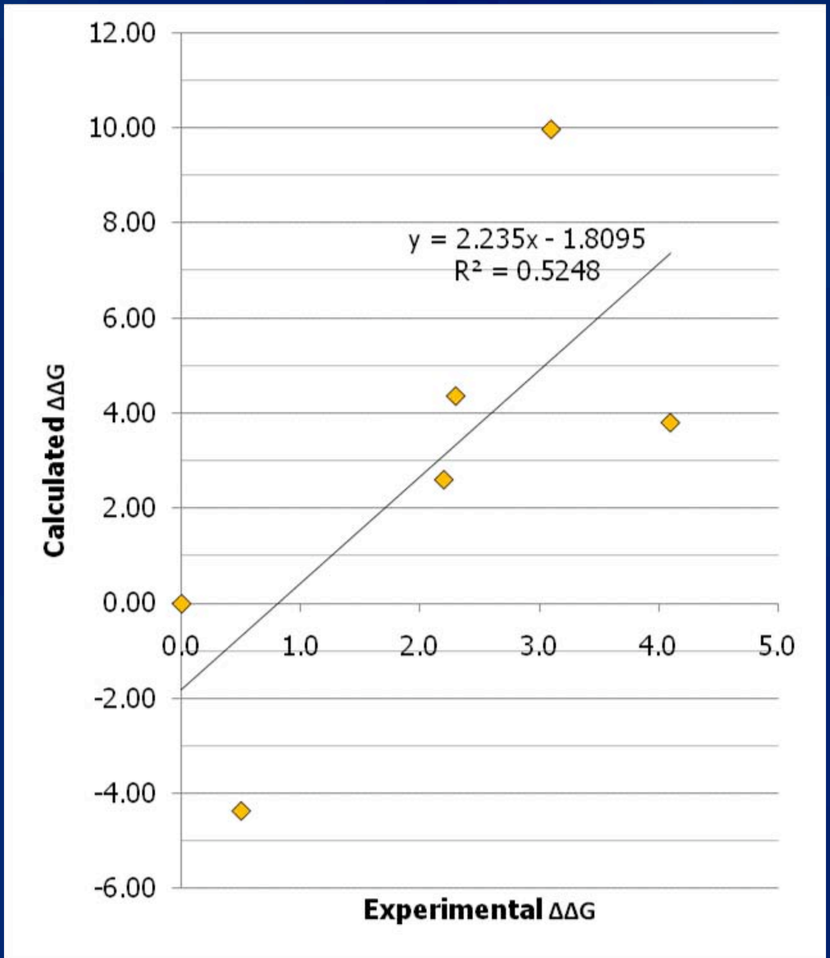
– sp^2	deg=2	hyd=1	420 (28.63%)	
– sp^3	deg=2	hyd=2	407 (27.74%)	
– sp^3	deg=3	hyd=1	353 (24.06%)	
– sp^3	deg=1	hyd=3	213 (14.52%)	Y
– sp^3	deg=1	hyd=1	38 (2.59%)	Y
– sp^2	deg=1	hyd=2	36 (2.45%)	

5ABP MMFF94 Energy

- Calculate the single point energy for 5ABP after sprouting hydrogens with OEChem.
  - Original implementation: 3866 Kcal/mol.
  - Current implementation: 2705 Kcal/mol.

- The new algorithm is also faster than the original implementation.

Final ABP Results



Solvation Energy Calculation

The benefits of improved hydrogen placement are also seen on in the transfer energies predicted by zap's "transfer" example program on the 63 blind compounds in the SAMPL-1 challenge.

	Original	Improved	Provided
_err	-510.06	-492.25	-470.10
_|err|	513.98	498.79	479.96
avg|err|	8.16	7.92	7.62
rms(err)	10.11	9.67	9.38
max|err|	21.32	21.35	23.63

Future Work

- We need a better (quantum) benchmark set.
- Perhaps consider acceptor strength, $1/q^2$.
- The myth of hybridization; primary amines.
- No attempt is made to handle or identify hydrogen bonding networks.
- Requires discrete GA-like optimization, such as Hooft, Sander and Vriend, Proteins, '96.
- The classic worst case is a large water box.

Acknowledgments

- Cambridge Crystallographic Data Center (CCDC) for access to the CSD database.
- Pat Walters and Matt Stahl for Babel.
- Matt Stahl and OpenEye for Omega.
- Anthony Nicholls for Zap and Delphi.
- Stan Wlodek for Szybki.
- Tom Halgren for MMFF94.
- Simon Kearsley for MMFF validation suite.

Medical Benefits

http://www.oxygenmedicine.com/oxygenated_ozonated_water.html

Research has shown that water whose bond angle is 101 degrees is 'dead' water, bereft of life-giving energy. When water is distilled the bond angle expands to 120 degrees upon evaporation, but collapses to 101 degrees upon condensation, and is therefore dead. A bond angle of 103 degrees corresponds to average water.

A bond angle of 106 degrees produces activated, energized water, and is attainable by placing a magnet, north pole inward, against the water container. The highest energy obtainable in liquid water is a bond angle of l09.5 degrees, and this is attainable only by ozonating water at 4 degrees C.