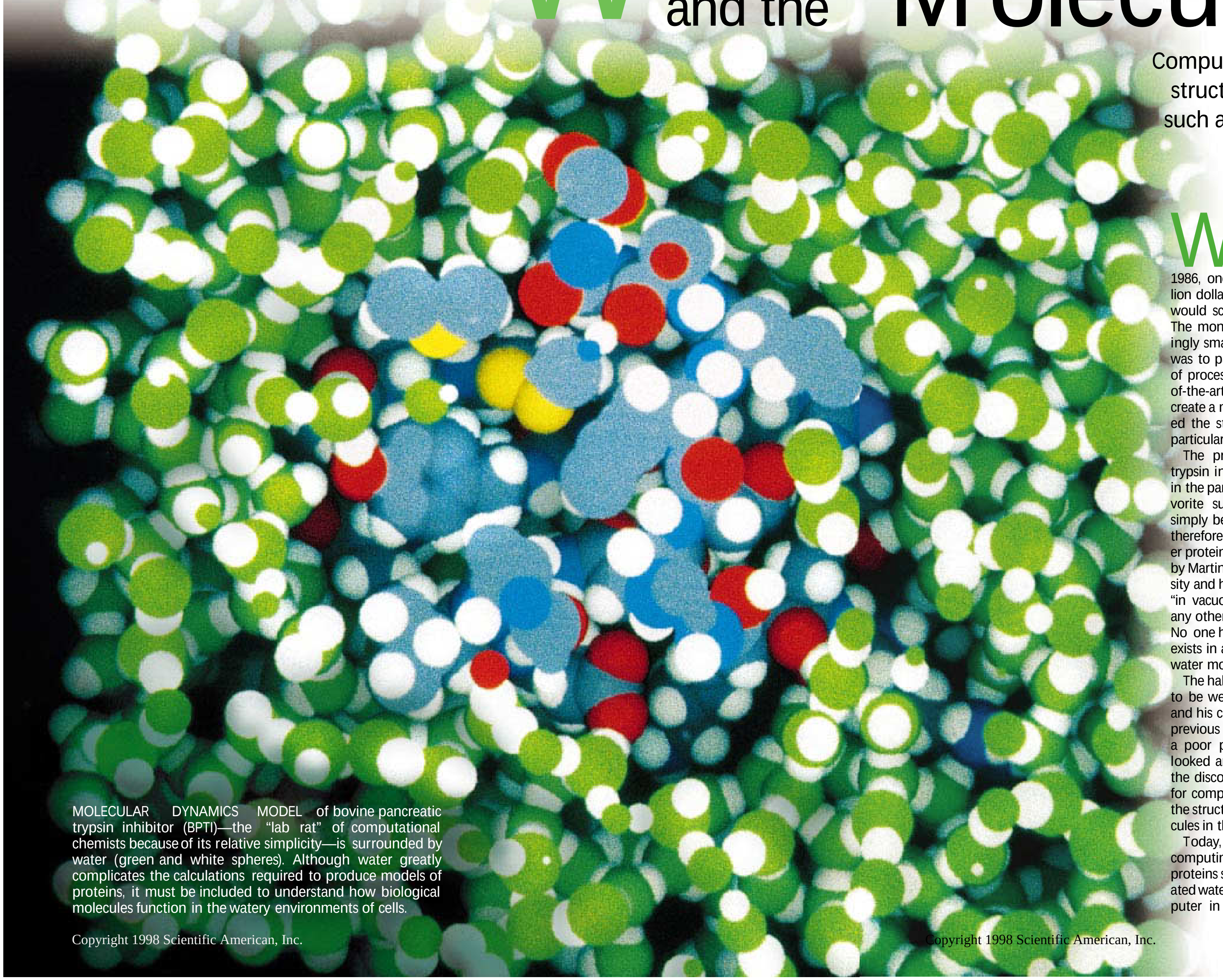
Simulating Water and the Molecules of Life

Computer modeling reveals how water affects the structures and dynamics of biological molecules such as proteins, yielding clues to their functions

by Mark Gerstein and Michael Levitt



MOLECULAR DYNAMICS MODEL of bovine pancreatic trypsin inhibitor (BPTI)—the "lab rat" of computational chemists because of its relative simplicity—is surrounded by water (green and white spheres). Although water greatly complicates the calculations required to produce models of proteins, it must be included to understand how biological molecules function in the watery environments of cells.

Water is cheap, if not free, in most places in the world. But during the summer of 1986, one of us (Levitt) spent half a million dollars on an amount of water that would scarcely wet the point of a pin. The money was not to buy the vanishingly small amount of water. Rather it was to pay for the roughly two weeks of processing time on a gigantic state-of-the-art supercomputer required to create a model of how the water affected the structure and movement of a particular protein.

The protein was bovine pancreatic trypsin inhibitor (BPTI), which is found in the pancreases of cattle. BPTI is a favorite subject of computer modelers simply because it is relatively small and therefore easier to study than most other proteins. It had been modeled before, by Martin Karplus of Harvard University and his colleagues in 1977, but only "in vacuo" (as if in a vacuum)—without any other molecules interacting with it. No one had visualized BPTI as it really exists in a living cell, with thousands of water molecules surrounding it.

The half a million dollars turned out to be well spent. Not only did Levitt and his colleague Ruth Sharon find the previous in vacuo model of BPTI to be a poor predictor of how the protein looked and behaved in the real world, the discovery helped to pave the way for computational chemists to simulate the structures of other biological molecules in their native, watery environs.

Today, given the great advances in computing technology, we can model proteins such as BPTI and their associated water molecules on a desktop computer in a couple of days, spending about 80 cents for electricity. Scientists have now simulated the aqueous ("in water") structures of more than 50 proteins and nucleic acids such as DNA.

Why is it so important to understand the effects of water on the shapes of biological molecules? Principally, because a molecule's structure yields clues to how it functions, helping scientists decipher the biochemical interactions that add up to life. On a more practical level, understanding the structures of biological molecules in water may one day help researchers design new drugs that act by blocking or enhancing various biochemical pathways.

The Water Within

To understand how water affects the structures of biological molecules, we must first appreciate the distinctive properties of water itself. These properties stem from the unique structure of water and the way this structure allows water to "manage" the electric charges of other molecules.

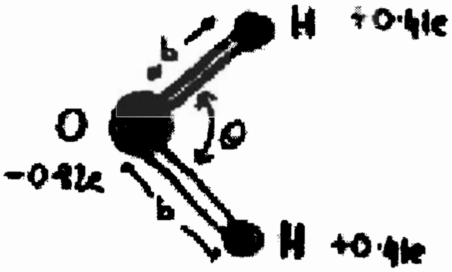
A single water molecule (H_2O) has an essentially tetrahedral geometry, with an oxygen atom at the center of the tetrahedron, hydrogen atoms at two of the four corners and clouds of negative charge at the other two corners. The clouds of negative charge result from the way in which the atomic structures of oxygen and hydrogen combine. In simplified terms, oxygen has eight negatively charged electrons circling its positively charged nucleus: two in an inner shell and six in an outer shell. The inner shell's maximum capacity is two electrons, so it is full, but the outer shell can hold as many as eight. Hydrogen has

ALL ILLUSTRATIONS BY MARK GERSTEIN AND MICHAEL LEVITT

Copyright 1998 Scientific American, Inc.

SIMULATING LIQUID WATER

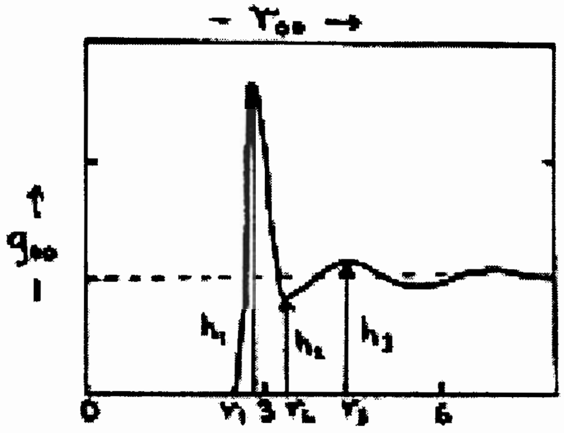
- Very simple model
  - 3 interaction centers
  - Completely flexible
  - Smooth cutoff at 6 Å (list to 8Å)



Electrostatics
Van der Waals } long range forces

- Good fit to experiment

	(25 °C) Experiment	Simulation
Property		
Potential energy (kcal/mol)	−9.2	−9.5
Pressure (atmospheres)	$\boxed{1}$	$\boxed{-61}$
Classical Specific Heat (cal/°K)	27	26
Diffusion Constant (Å²/ps)	0.23	0.22
Rotational Relaxation Time (ps)	2	1.6
Radial Distribution Function r_1	2.8	2.7
h_1	$\boxed{2.5}$ 3.0*	$\boxed{3.2}$
r_2	3.3	3.3
h_2	0.8	0.8
r_3	4.6	4.3
h_3	1.11	1.09

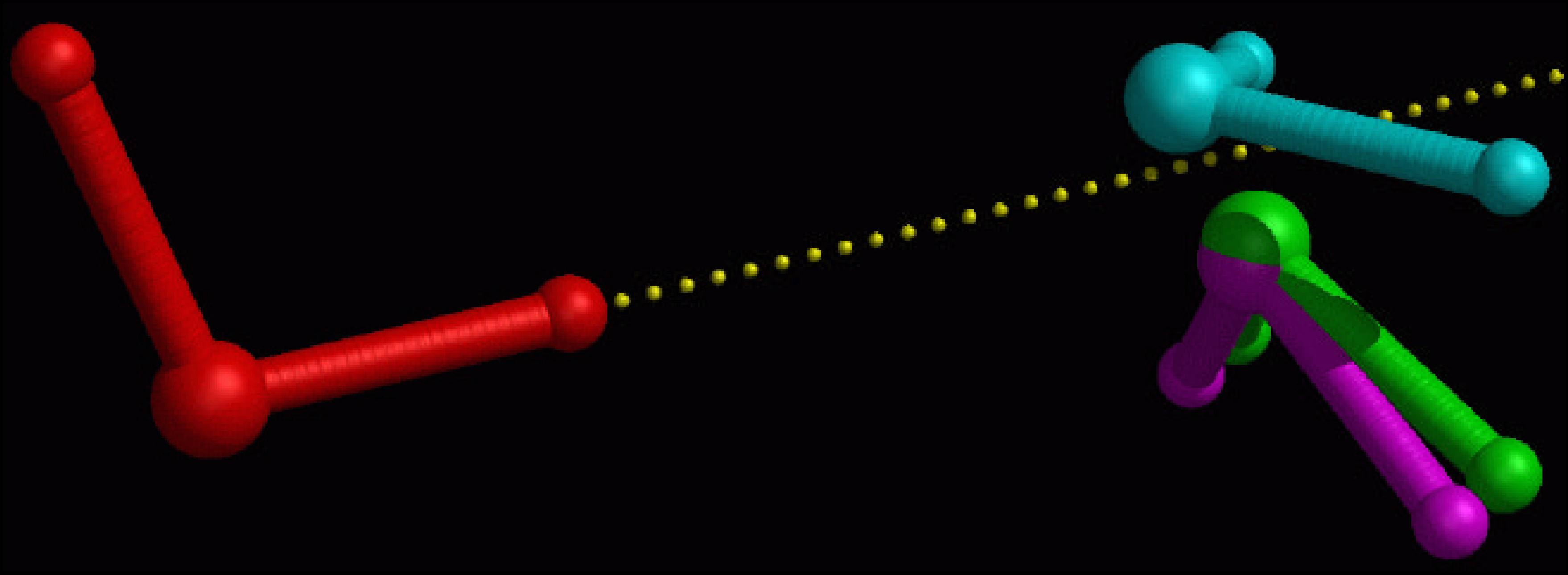


* Calibration error fixed after 15 years of experiment

Water Dimer Structure and Energy

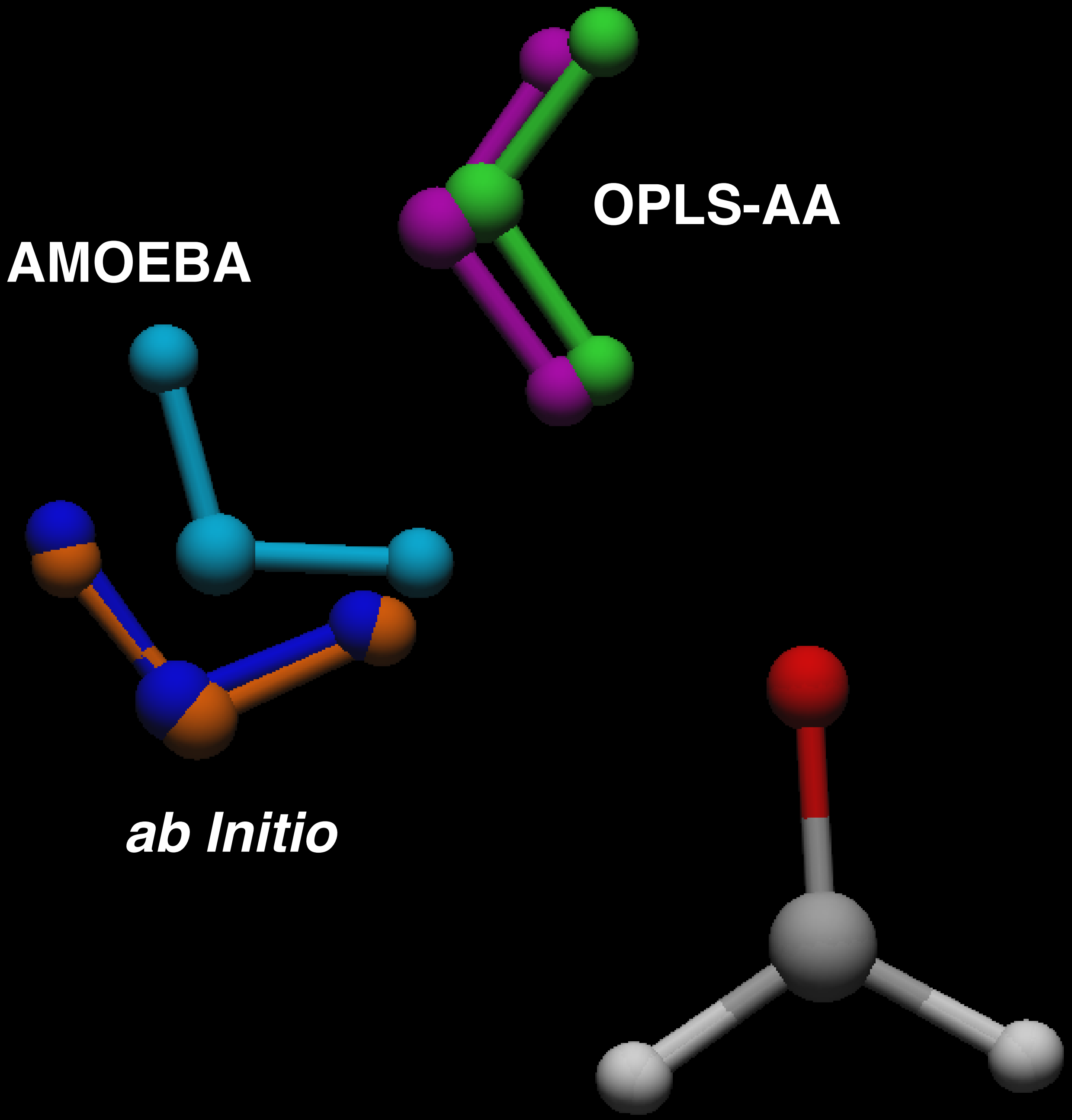
	Expt	QM	AMOEBA
O-O Distance (Ång)	2.98	2.907	2.892
O-O Bisector Angle (°)	57 ± 10	56.9	57.1
Dimerization Energy	5.4 ± 0.7	4.98	4.95

● MP2/CBS ● TIP3P ● AMOEBA

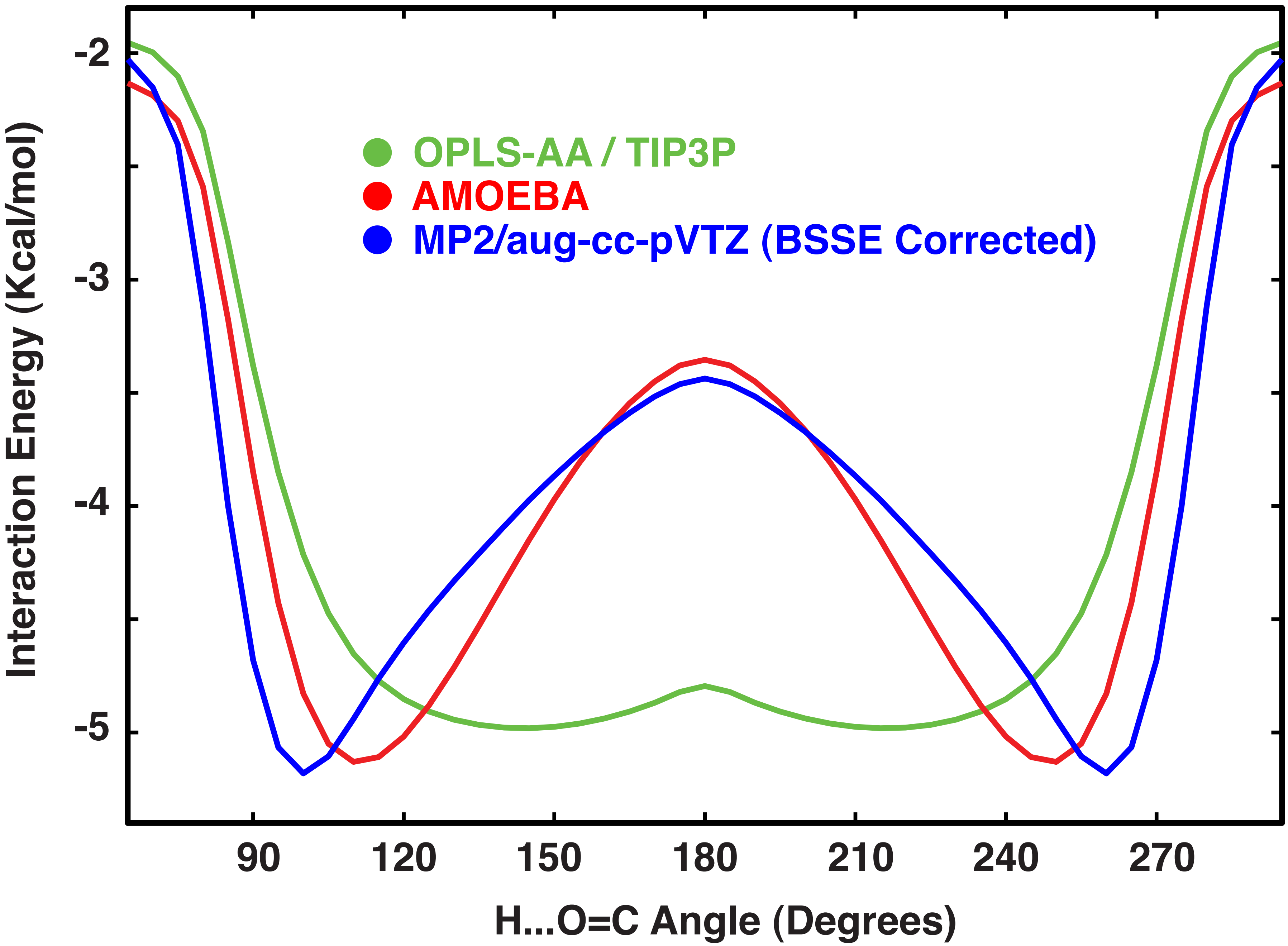


Directionality of Hydrogen Bonds

- determined largely by the potential of the acceptor group
- fixed charge models lacking in directional preference
- a big deal for ligand design *!!*



Formaldehyde-Water Hydrogen Bond Angle



Choice of Permanent Electrostatics

Isotropic Models

Simple Atomic Partial Charges
Diffuse Charges (Guillot & Guissani)

Anisotropic Models

“Extra” Charge Sites (Lone Pairs....)
Atomic Multipole Moments
Gaussian Charge Densities (Darden)

Is anisotropy required for high accuracy?

Electrostatic Potential ✓
Molecular & Functional Group Moments ✓
Hydrogen Bond Directionality ✓

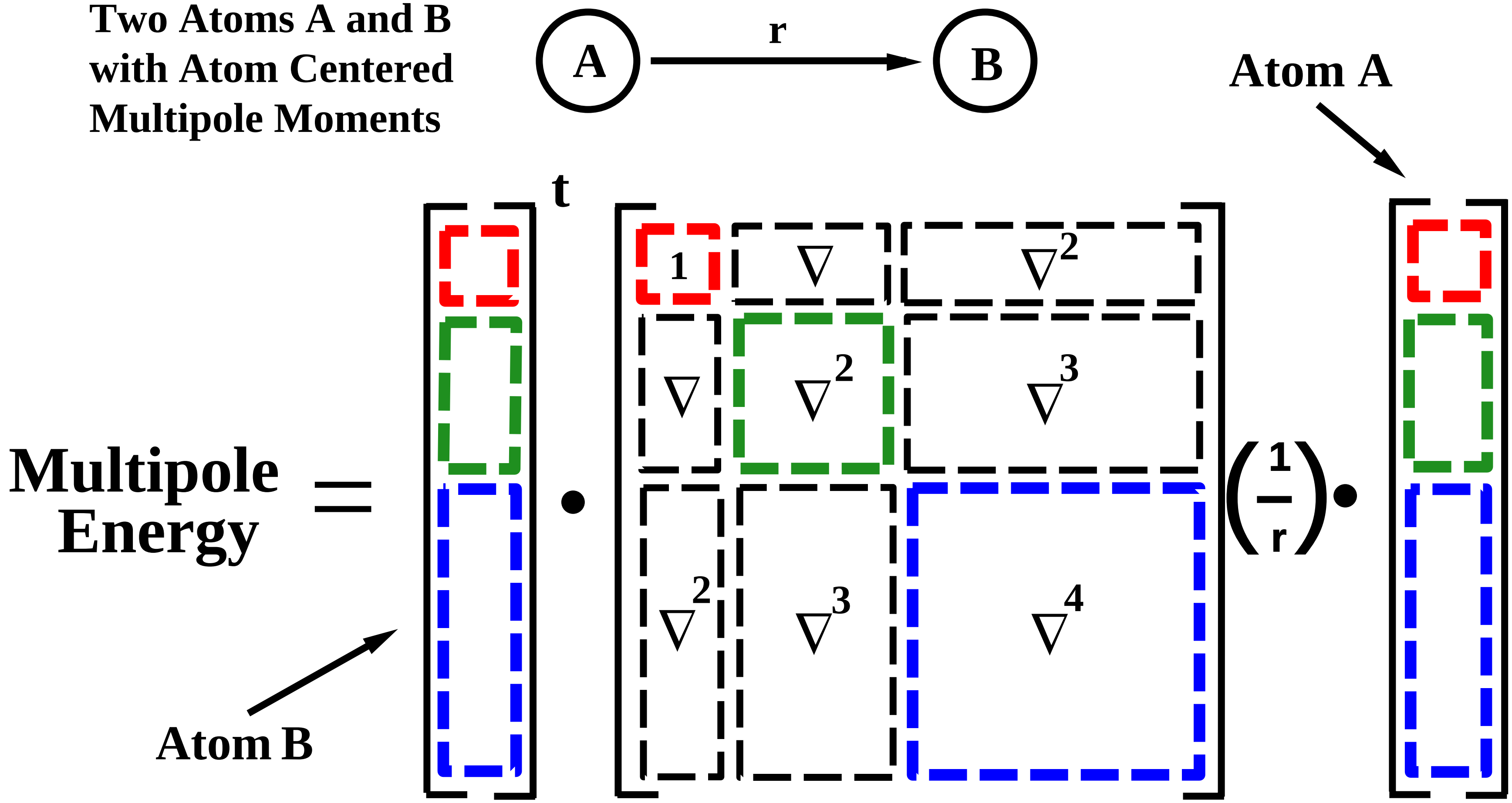
*Multipole Models Fit to 6-31G** Electric Potentials*

Molecule	Points	Relative RMS Error M only	M+D	M+D+Q
Methane	455	13.53	1.04	0.39
Water	363	8.44	0.88	0.01
Ammonia	404	9.90	2.31	0.02
Methanol	483	8.35	1.31	0.02
Acetone	614	2.31	0.98	0.02
Acetylene	438	1.34	0.06	0.02
Formamide	492	3.68	0.65	0.03
Me Acetate	559	6.03	0.67	0.02
DiMe Amine	583	16.27	1.48	0.03
NMA	685	3.26	0.30	0.01

[#] Data from D. E. Williams, *J. Comput. Chem.*, **9**, 745-763 (1988)

Polytensor Formulation of Multipole Interactions

Two Atoms A and B
with Atom Centered
Multipole Moments



- 🟥 **Monopole (1 Component)**
- 🟩 **Dipole (3 Components)**
- 🟦 **Quadrupole (9 Components)**

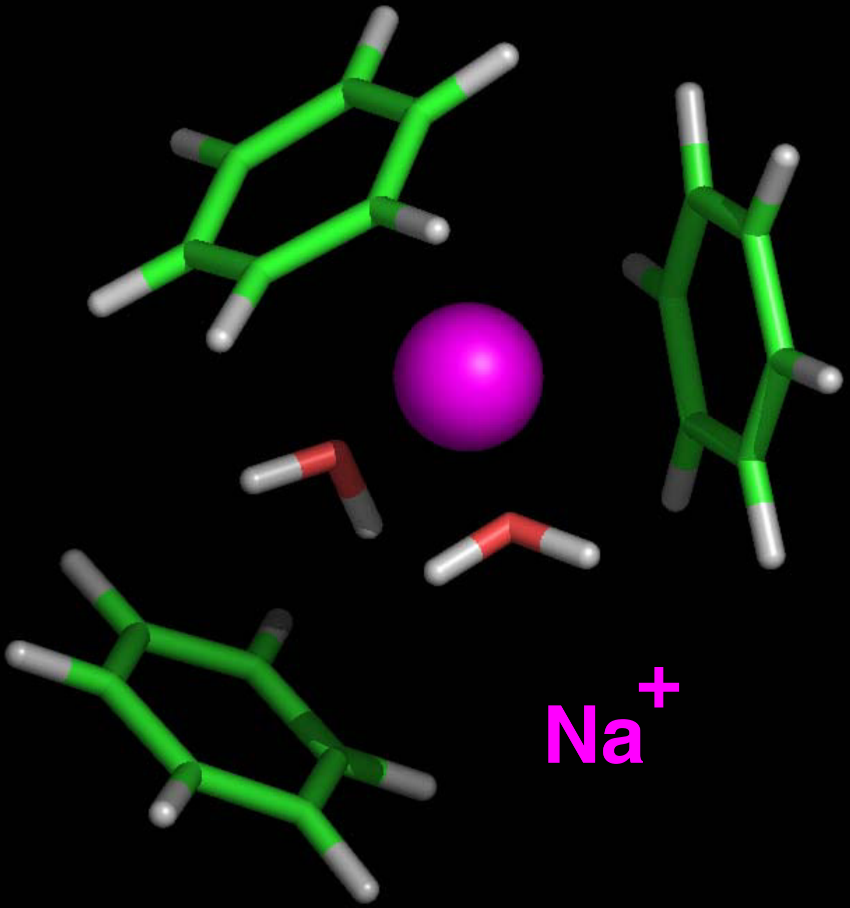
$$\nabla^{(n)} = \frac{\partial^i}{\partial x^i}\,\frac{\partial^j}{\partial y^j}\,\frac{\partial^k}{\partial z^k}\;,\; i+j+k=n$$

π-Cation Interactions

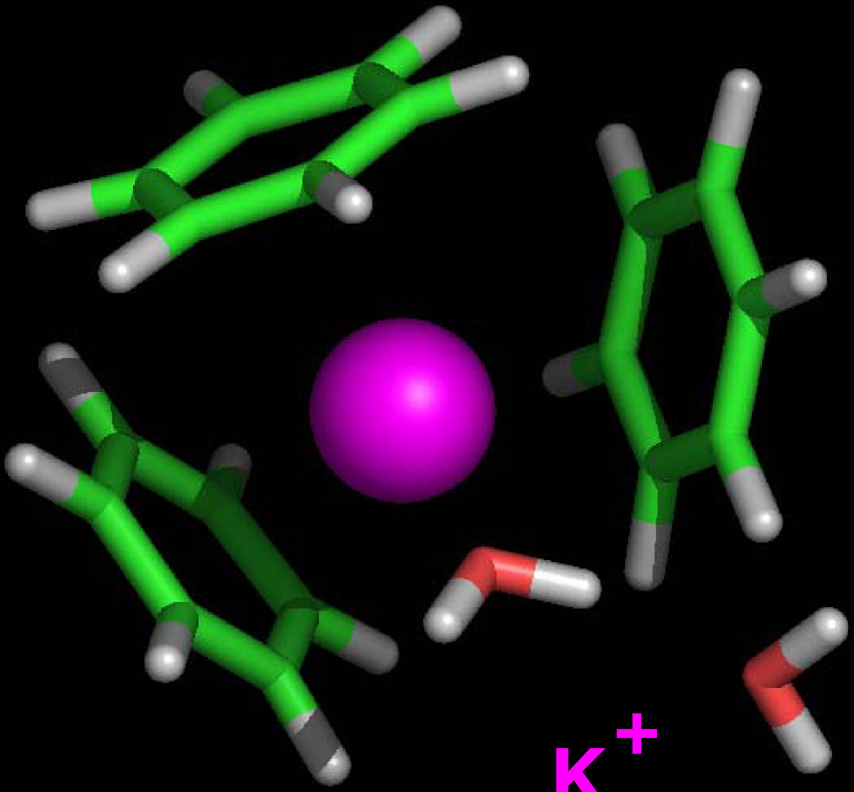
Catastrophic Failure of the Standard Model

	ΔE_0	K^+-Centroid	ΔH_{298}
OPLS-AA	-9.32	2.90	
CHARMM27	-11.06	2.81	
Amber *ff94*	-12.55	2.74	
Amber *ff02*	-15.87	2.63	
AMOEBA	*-19.27*	*2.81*	*-18.15*
MP2/6-311+G(2d,2p)	-18.4	2.81	
MP2/aVQZ	-19.9	2.79	
CCSD(T)/CBS	-20.6	2.79	-20.1
Expt (HPMS)			-18.3
Expt (CID)			-17.7

Ion Selectivity by Benzene-Water



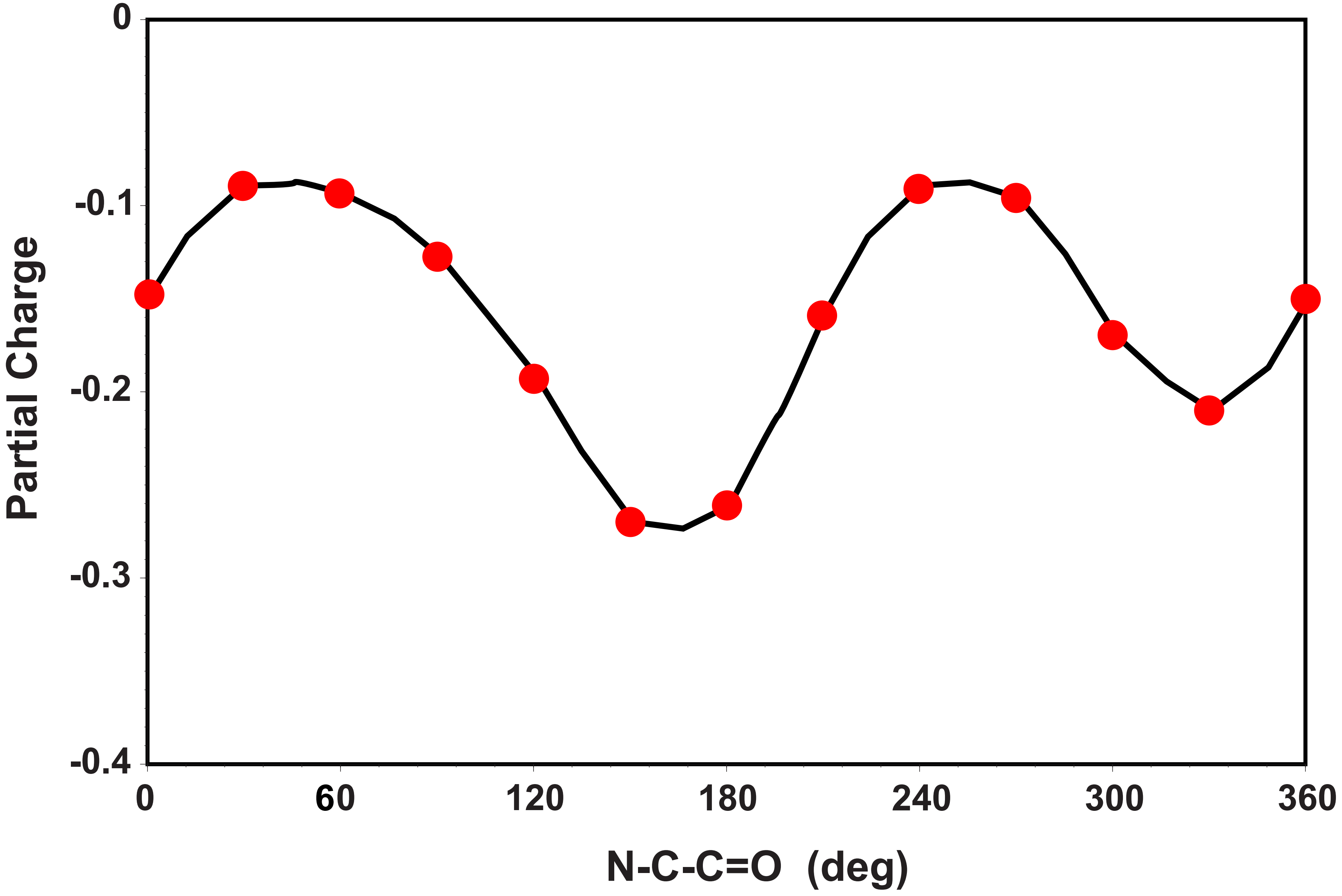
Na^+



K^+

Torsional Dependence of Charges

*(ALA Dipeptide Cα, MP2/6-311G**)*



Dipole Polarizability

All matter is polarized in direct proportion to the strength of an external field, where the proportionality constant is α, the polarizability:

$$\mu_{\text{induced}} = \alpha E \quad \left(\textit{i.e.}, \mu \text{ is linear, provided } E \text{ is not too big!}\right)$$

Imagine a one-electron (*e*) atom with a radius of R placed in an electric field E. The electron's orbit will be shifted away from the nucleus by a distance *d*. Then the induced dipole is given by:

$$\mu_{\text{induced}} = \alpha E = de$$

At the equilibrium value of *d*, the external force on the electron due to the field must exactly counterbalance the internal force of displacement between the nucleus and the electron. These forces are:

$$F_{\text{ext}} = eE\ , \text{ and } F_{\text{int}} = \frac{e^2}{4\pi\varepsilon_0 R^2}\sin\theta \approx \frac{e^2 d}{4\pi\varepsilon_0 R^3} = \frac{e}{4\pi\varepsilon_0 R^3}\mu_{\text{induced}}$$

Since F_{ext} is equal to F_{int}, we obtain for the polarizability: $\alpha = 4\pi\,\varepsilon_0\,R^3$. Thus, neglecting the permittivity term, the polarizability should be roughly equal to the volume of the atom or molecule. For water, the experimental value of α = 1.48 Å^3 suggests a radius of 1.14 Å, about 20% less than the standard water radius of 1.4 Å used in surface area calculations.

The Importance of Polarization

- ***Inter-*****molecular polarization is necessary to describe gas-phase and condensed-phase properties within a single model**

- ***Intra-*****molecular polarization is needed to treat the conformational dependence of electrostatics**

Choice of Polarization Model

- **Fluctuating Charge**
- **Electronegativity Equalization**

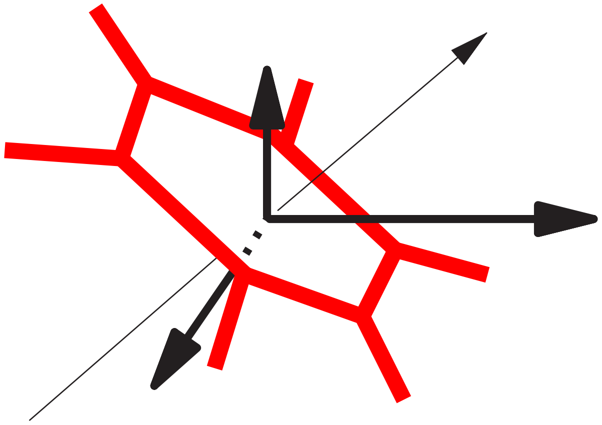
- **Drude Oscillator**
- **“Shell” Method**
- **Charge-on-Spring (COS)**

- **Classical Induced Dipoles (Applequist)**
- **Damped Mutual Induction (Thole)**

- **Various Semi-Empirical QM Methods**

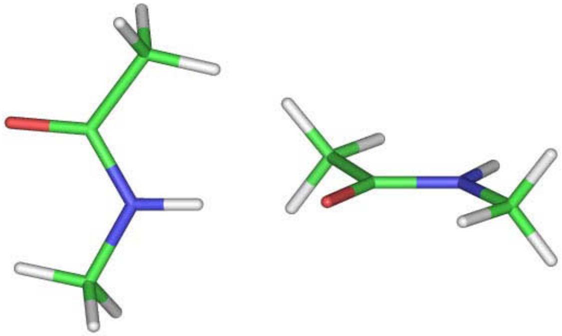
Molecular Dipole Polarizability

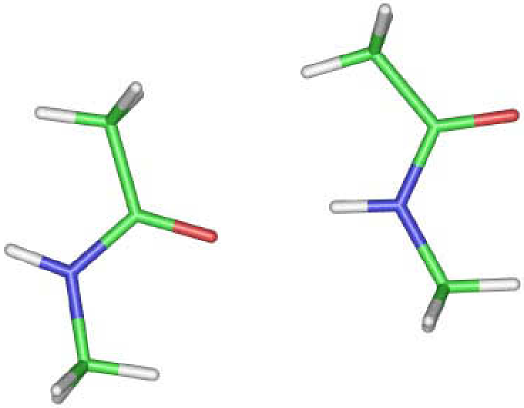
$\boldsymbol{E} = [0.1, 0.1, 0.1]$



- Based on Thole's modified dipole interaction model
- ***Isotropic*** atomic dipole polarizabilities are sufficient to reproduce experimental molecular polarizability tensors
- Induced dipoles further ***interactively*** induce each other within the molecule
- The field and interaction involved in induction are modified ***(damped)*** at short range

Intermolecular Polarization for NMA

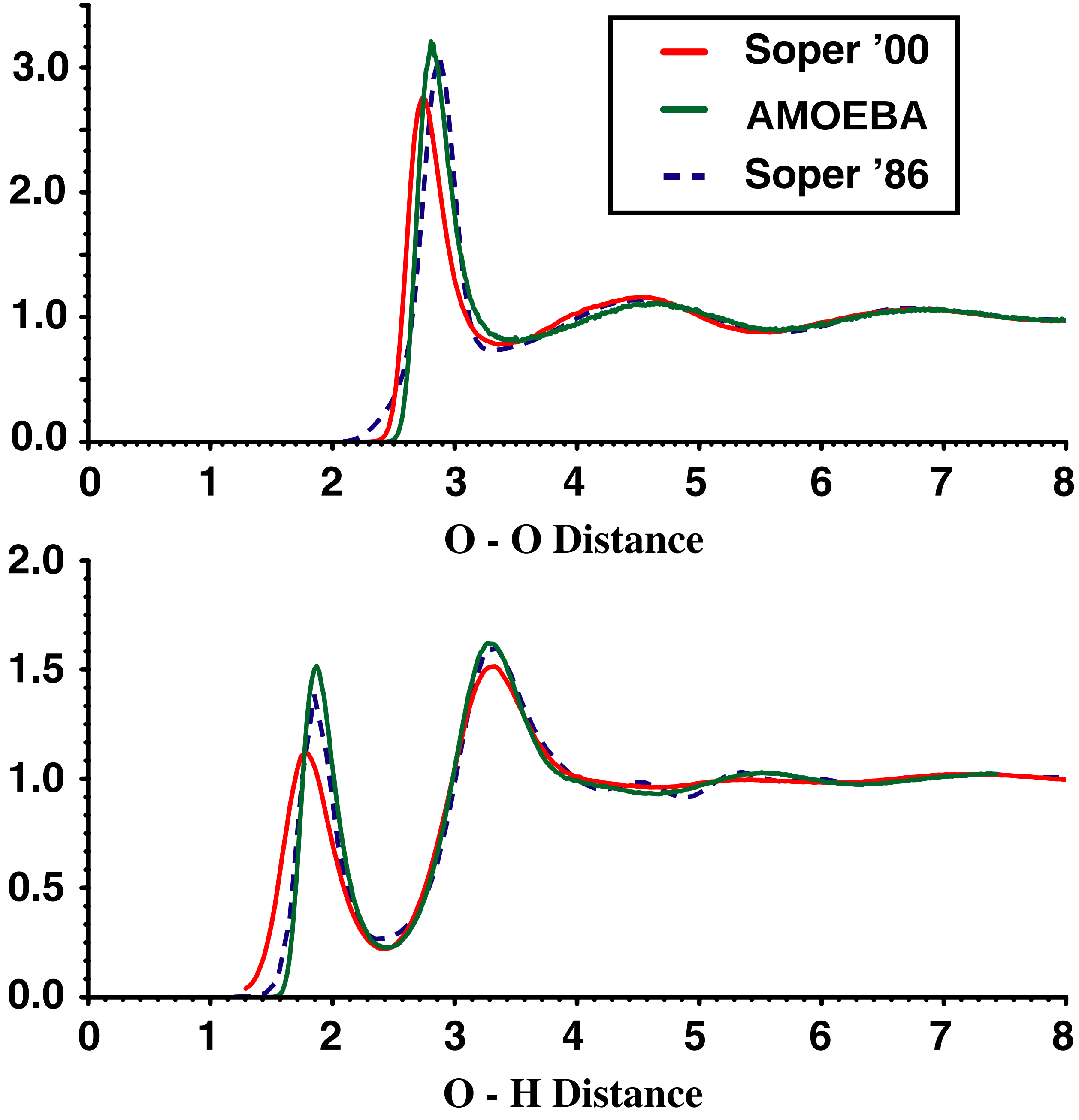




	Dimer DMA/ QM Dipoles	Monomer DMA	Monomer DMA + Polarization
ESP RRMS	6.9%	16.6%	6.5%
Total Dipole	7.88	6.64	7.83
Dx	7.73	6.46	7.69
Dy	0.09	0.01	0.03
Dz	1.51	1.52	1.48
ESP RRMS	6.4%	15.9%	5.6%
Total Dipole	8.85	7.49	8.85
Dx	-8.82	-7.44	-8.81
Dy	0.76	0.75	0.82
Dz	0.02	0.00	0.00

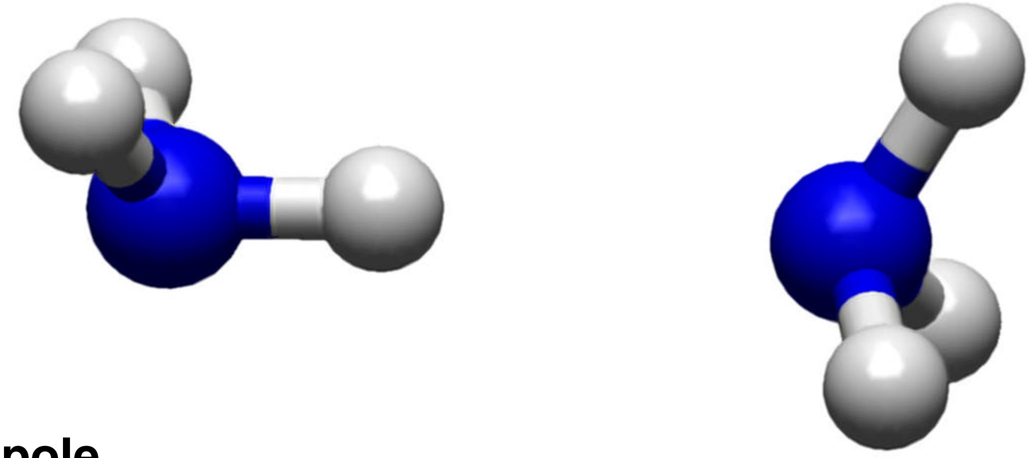
[#] All calculations performed at MP2/6-311G++(2d,2p)

Liquid Water Properties



	Expt	AMOEBA
Heat Vaporization (kcal)		
298K	10.51	10.48
Density (g/cc)		
298K	0.997	1.000
323K	0.988	0.992
363K	0.962	0.964
Dielectric Constant		
273K	87.7	86.8
298K	78.3	80.7
323K	69.9	66.5
Diffusion (10^{-5} cm/s^2)		
298K	2.3	2.0
C_p (cal/mol K)		
298K	18.0	20.9 / 27.6
Avg Mol Dipole (Debye)		
	2.6-3.0	2.78
E_{pol} / (E_{pol}+E_{perm})		30%

Ammonia Monomer, Dimer and Liquid



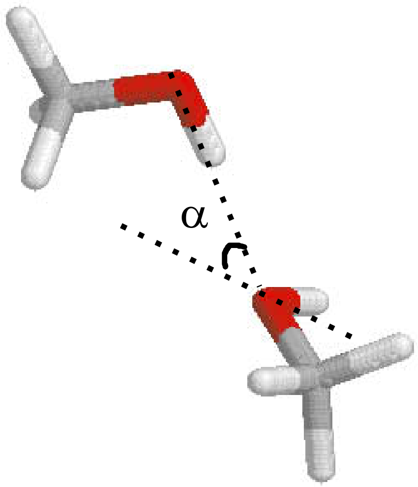
MONOMER	Dipole	Quadrupole
Expt	1.471	-2.42, -2.45
AMOEBA (unscaled)	1.528	-3.093
AMOEBA (60% Q)	1.528	-2.491

DIMER	Energy	N..H	N..N	<HN..H
ab Initio *	3.09	2.226	3.224	135
AMOEBA	3.19	2.248	3.265	120

* aug-cc-pVQZ energy at 6-31+G* minimum

LIQUID	H_{vap}	Pressure	$Dx10^5$	T(K)
Expt	5.58	1	5.8	240
AMOEBA	5.54	99	5.0	240

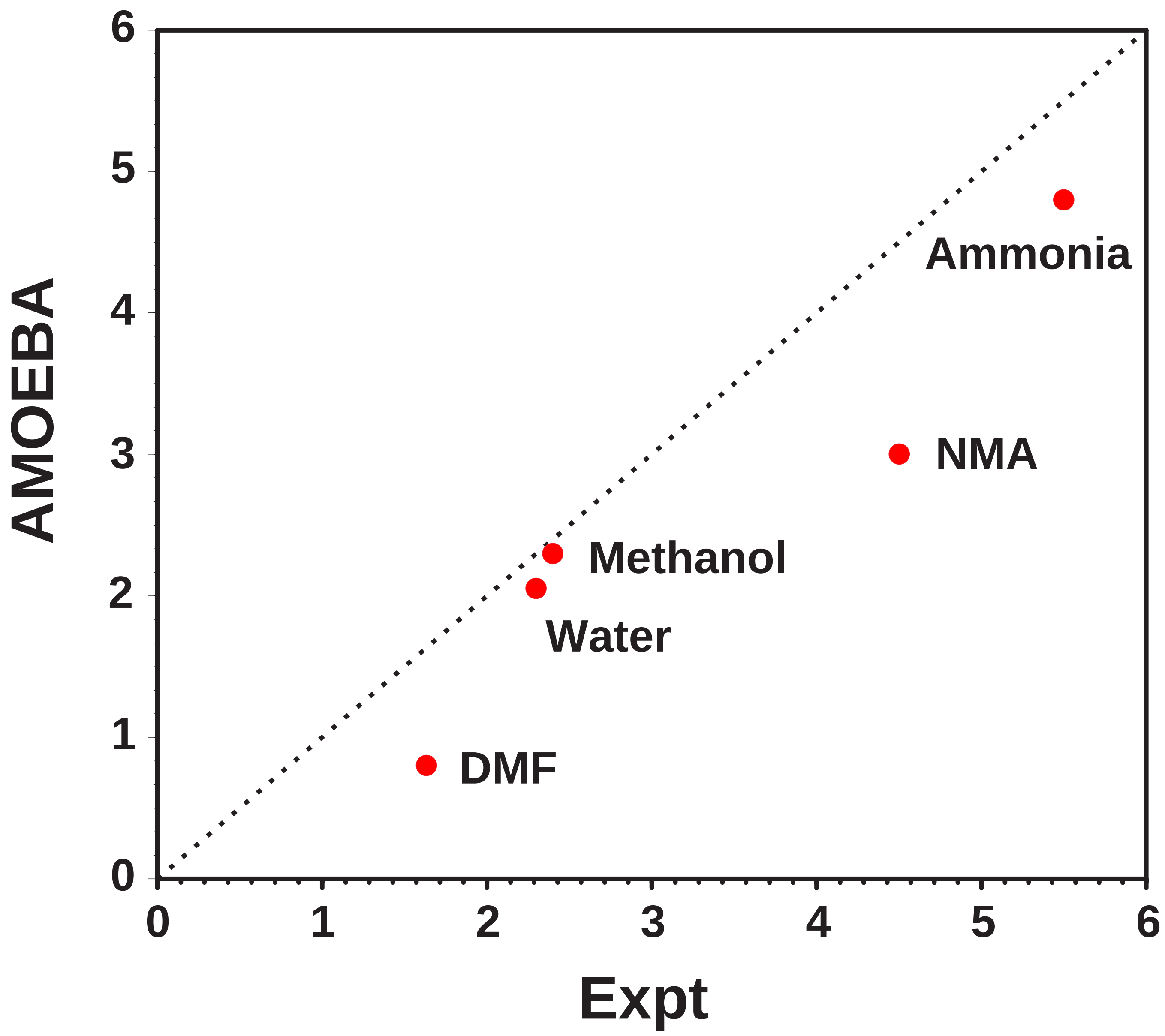
Methanol Dimer



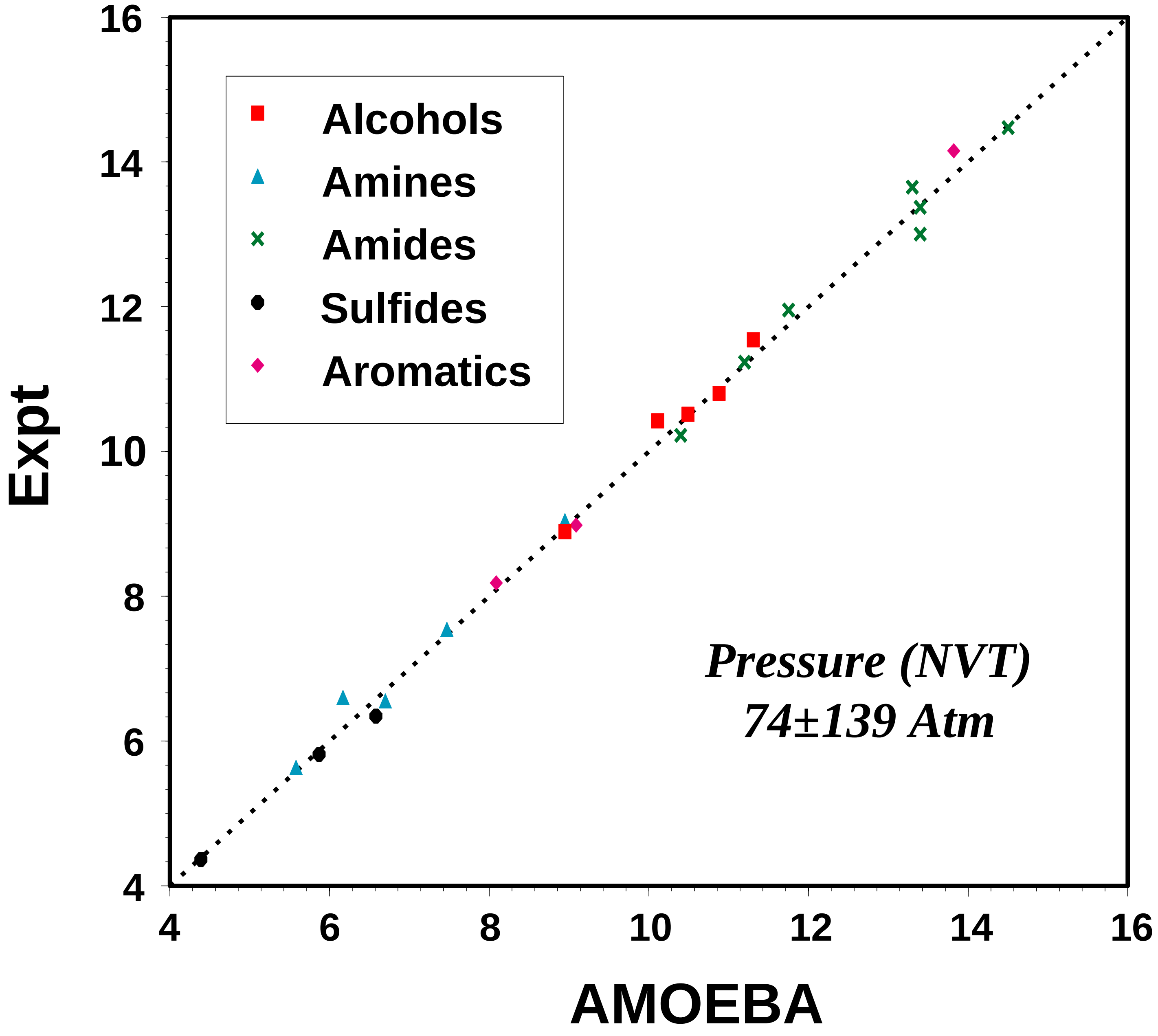
	QM[#]	AMOEBA	CHARMM
E (kcal/mol)	5.44	5.38	6.99
$R_{O..O}$ (Å)	2.87	2.91	2.80
α (deg)	44	44	23
β (deg)	179	174	178

[#] MP2 Calculations from Mooij, *et al.*, 1999

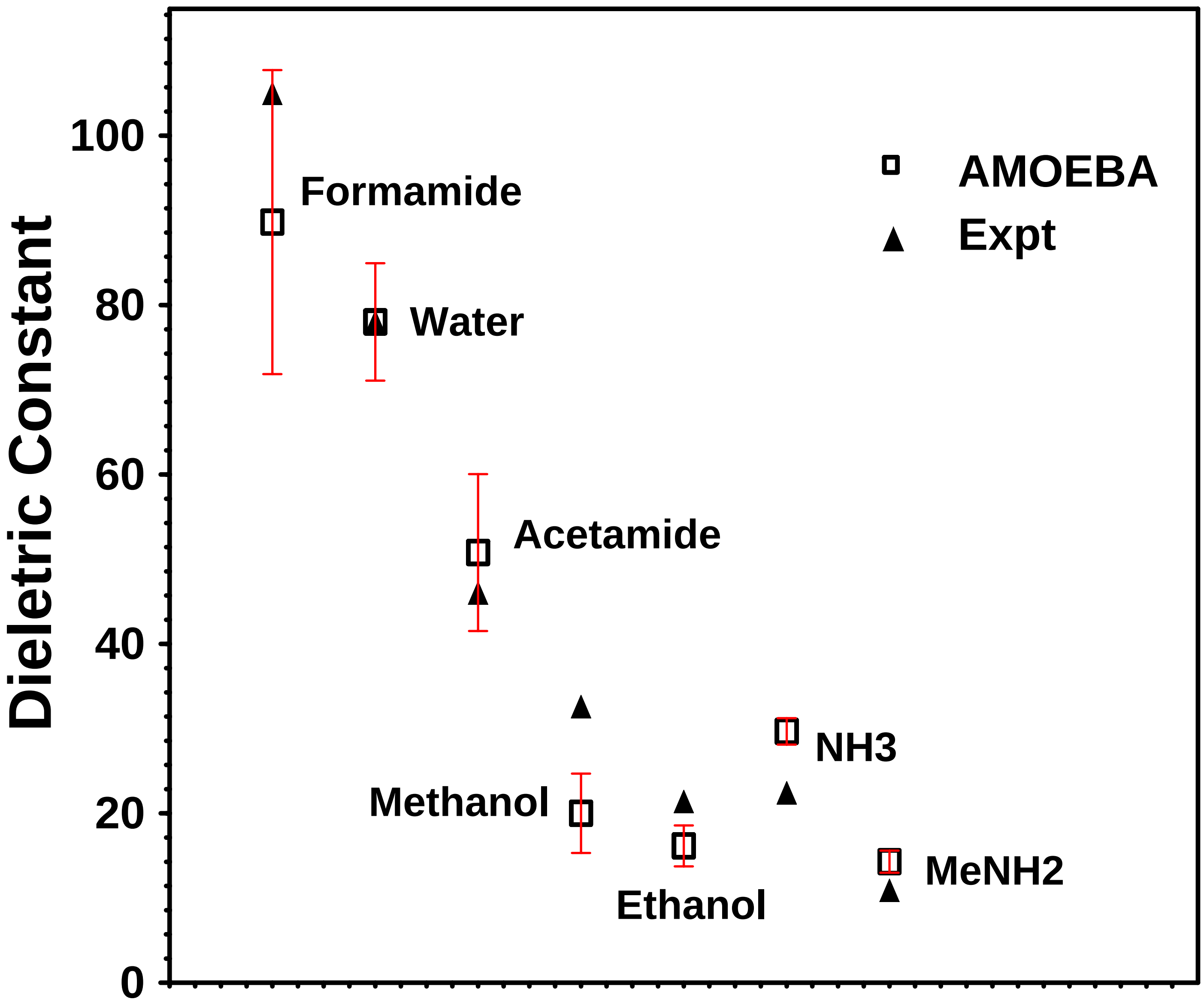
Self-Diffusion Coefficient (10^{-5} cm^2/s)



Heat of Vaporization (kcal/mol)

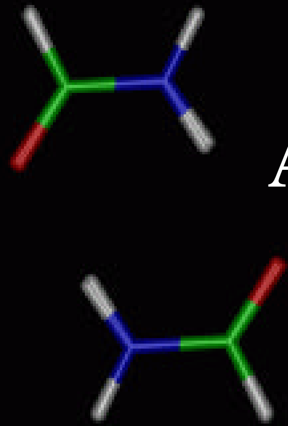


Dielectric Constants: AMOEBA vs. Expt



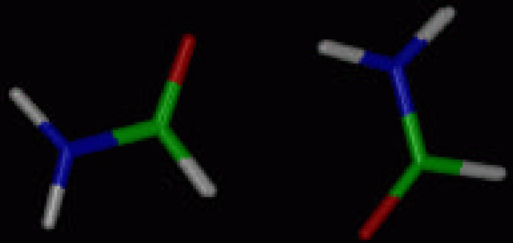
Formamide Dimer Energy Minima

A



E = -16.0 kcal/mol
rms = 0.02 Å

B

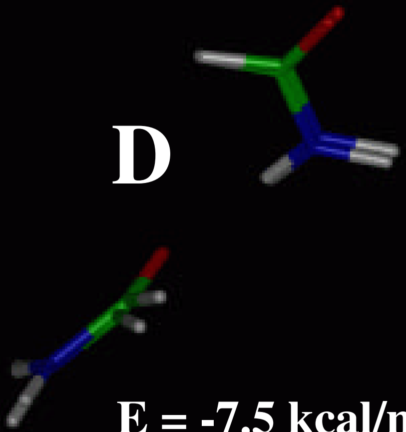


E = -10.3 kcal/mol
rms = 0.04 Å

C

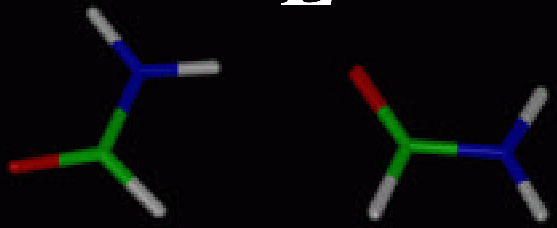
E = -9.0 kcal/mol
rms = 0.09 Å

D



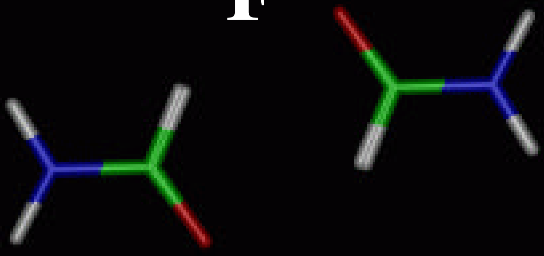
E = -7.5 kcal/mol
rms = 0.28 Å

E



E = -7.3 kcal/mol
rms = 0.03 Å

F



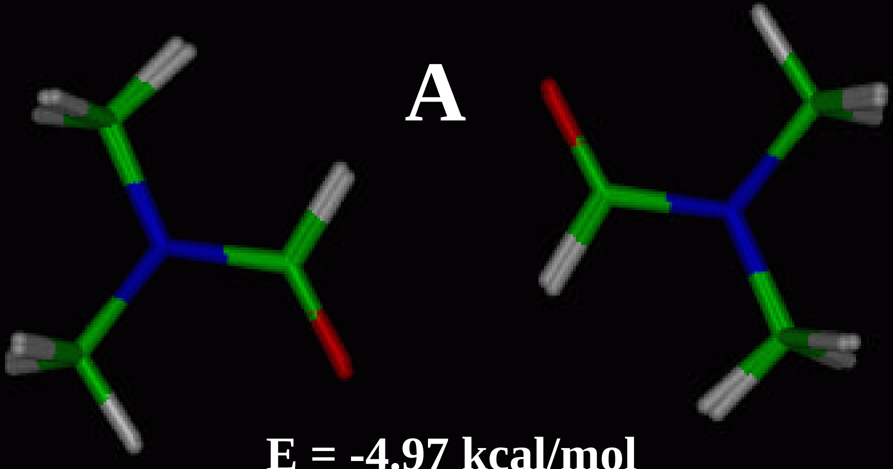
E = -5.5 kcal/mol
rms = 0.05 Å

Formamide Dimer Association Energies

	A cyclic	B side	C nonplan	D nonplan	E side	F head-tail
MP2/6-31G**	**-13.4**	**-8.5**	**-6.9**	**-5.8**	**-6.1**	**-3.6**
B3LYP/6-31G(d)	**-13.4**	**-8.3**	**-6.7**	**-6.0**	**-6.7**	**-3.2**
MP2/aug-cc-pVTZ	***-16.1***	***-10.6***	***-8.2***	***-7.2***	***-6.9***	***-5.4***
AMOEBA	**-16.0**	**-10.3**	**-9.0**	**-7.5**	**-7.3**	**-5.5**
RMS	***0.02***	***0.04***	***0.09***	***0.28***	***0.03***	***0.05***
OPLS-AA	**-14.2**	**-7.8**	**-8.2**	**-8.2**	**-8.0**	**-2.7**
RMS	***0.06***	***0.24***	***0.82***	***1.03***	***0.63***	***0.16***
AMBER	**-16.8**	**-9.5**	**-9.6**	**-9.0**	**-8.9**	**-3.8**
RMS	***0.06***	***0.09***	***0.22***	***1.03***	***0.67***	***0.12***
CHARMM	**-13.0**	**-8.2**	**-8.0**	**-7.7**	**-7.6**	**-4.3**
RMS	***0.05***	***0.13***	***0.21***	***1.06***	***0.74***	***0.10***
MM3	**-12.0**	**-6.5**	**-6.8**	**-6.8**	**-6.4**	**-1.5**
RMS	***0.06***	***0.24***	***0.38***	***1.37***	***0.24***	***0.25***

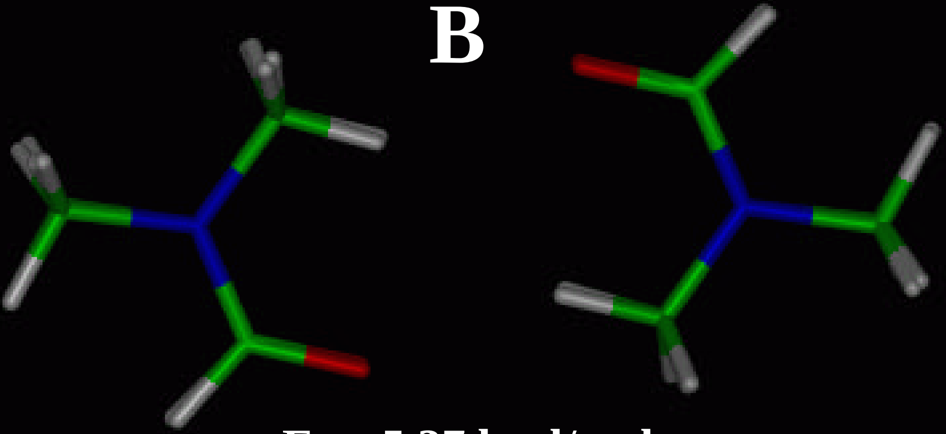
Comparison of DMF Dimers: AMOEBA vs Dixon/Hay

A



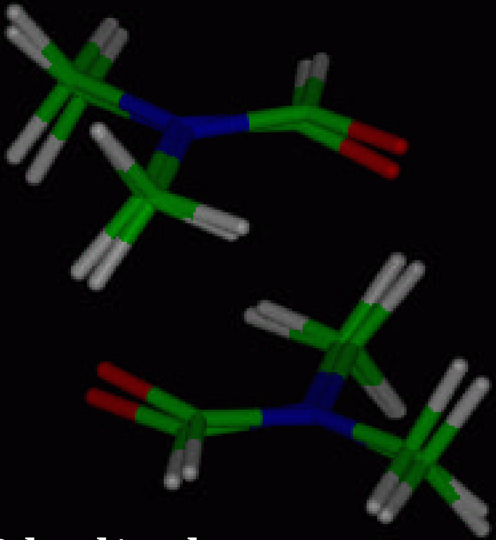
E = -4.97 kcal/mol
***rms* = 0.08Å**

B



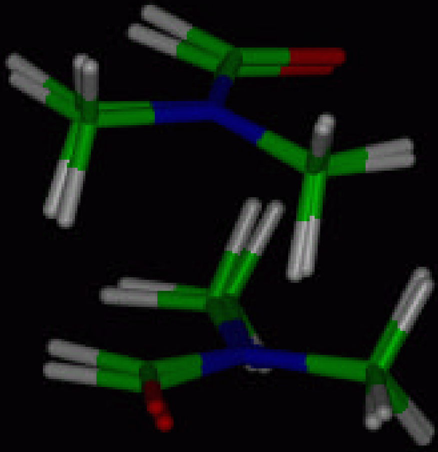
E = -5.37 kcal/mol
***rms* = 0.09Å**

C



E = -7.80 kcal/mol
***rms* = 0.24Å**

D



E = -8.79 kcal/mol
***rms* = 0.15Å**

Dimethylformamide Dimer Structure and Energy

	A	**B**	**C**	**D**
MP2/DZP+diffuse [#]	**-6.61**	**-5.51**	**-10.32**	**-10.98**
BSSE Corrected	**-4.07**	**-3.31**	**-5.86**	**-6.36**
MP2/aug-cc-pVTZ [#]	**-6.95**	**-5.82**	**-11.41**	**-12.11**
BSSE Corrected	**-5.35**	**-4.14**	**-8.34**	**-8.90**
AMOEBA (single)	**-4.94**	**-5.03**	**-7.37**	**-8.60**
AMOEBA (opt)	**-4.97**	**-5.37**	**-7.80**	**-8.79**
RMS	***0.08***	***0.09***	***0.24***	***0.15***
OPLS-AA (single)	**-1.68**	**-0.60**	**-3.41**	**-2.48**
OPLS-AA (opt)	**-3.45**	**-3.54**	**-5.42**	**-5.18**
RMS	***0.25***	***0.64***	***0.41***	***0.69***

[#] QM Results from Vargas, *et al.*, *JACS*, ***122***, 4750-4755 (2000)

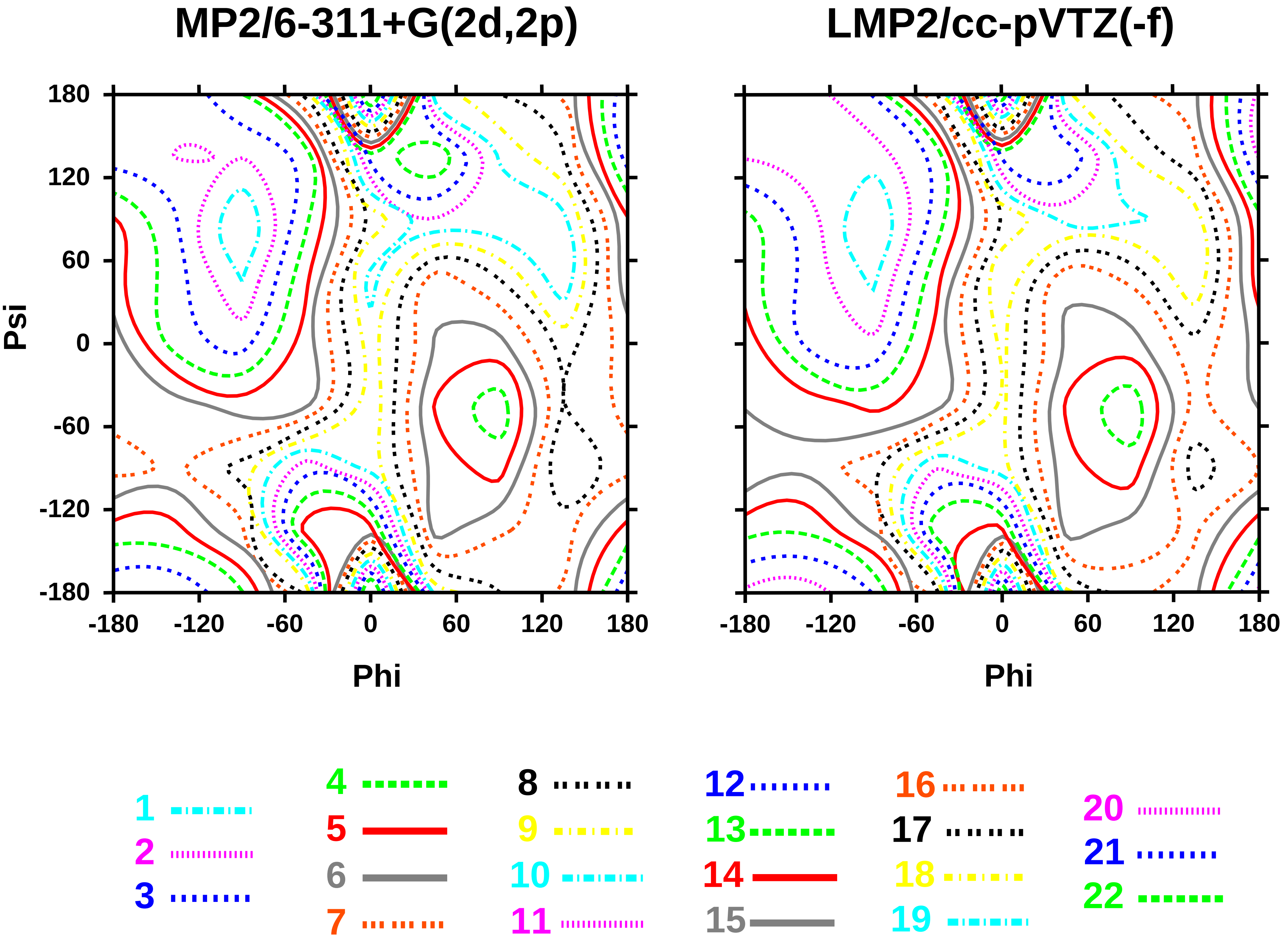
Parameterization for Polypeptides

- **vdW parameters and atomic polarizabilities transfered from small molecules**

- **Atomic multipole parameters**
  - -> **from small molecule fragments (?)**
  - -> **from capped amino acids (?)**

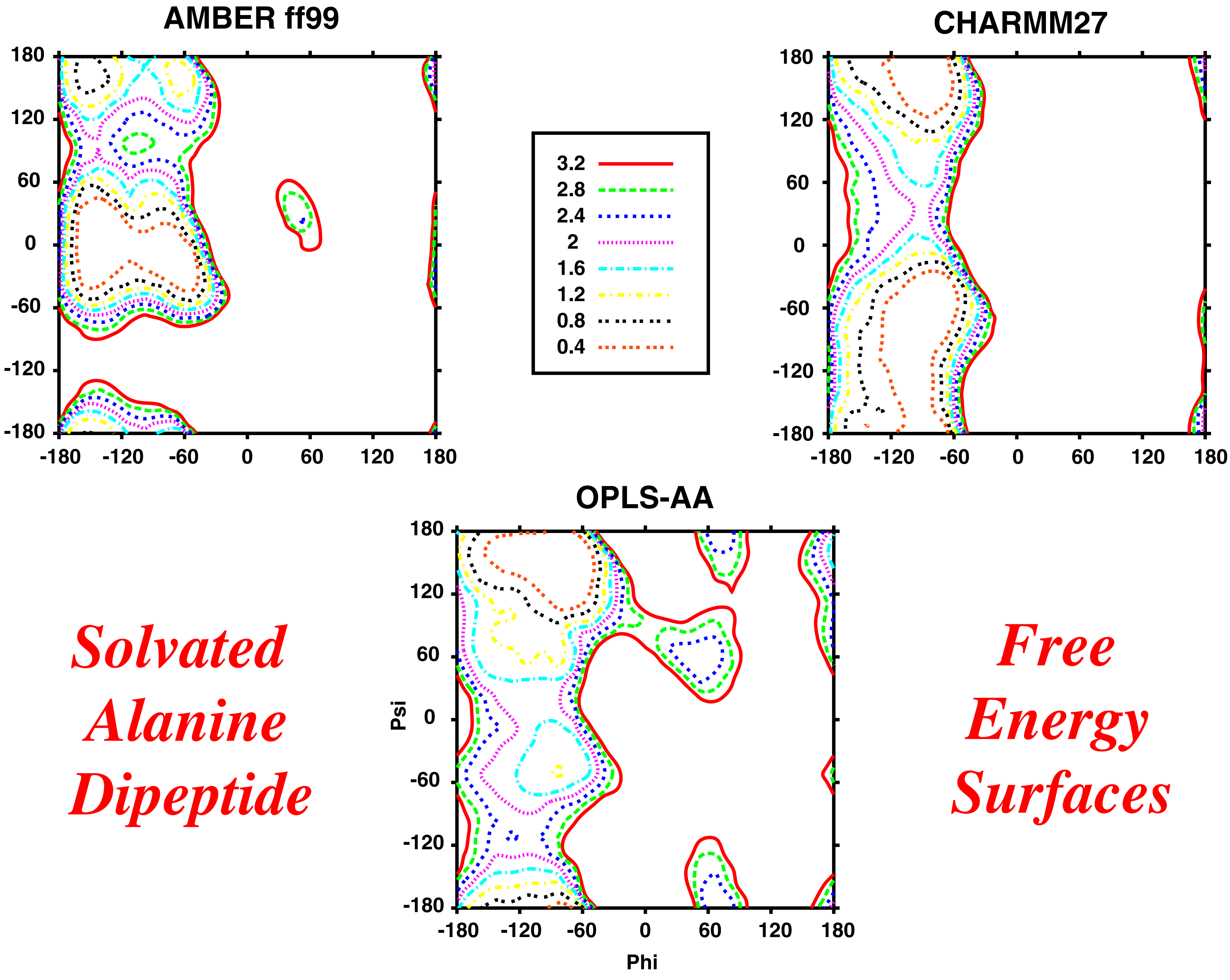
conformational dependence via intramolecular polarization

- **Torsional parameters obtained by fitting to conformational energy surfaces**


MP2/6-311+G(2d,2p)
LMP2/cc-pVTZ(-f)
Psi
180
120
60
0
-60
-120
-180
Phi
-180
-120
-60
0
60
120
180
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Solvated Alanine Dipeptide

Free Energy Surfaces

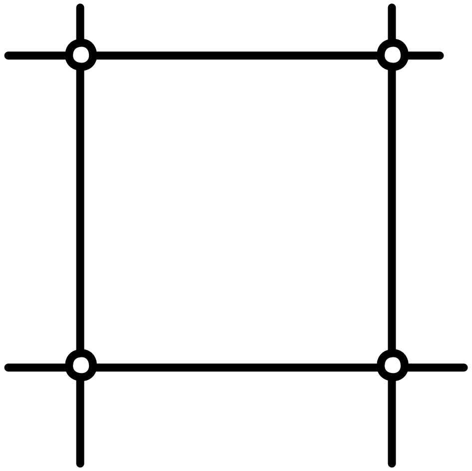


Torsional Energy Functional Forms

- ***Fourier series***

$$E_{tors} = k_1 [1+\cos(\phi)] + k_2 [1-\cos(2\phi)] + k_3 [1+\cos(3\phi)] + ...$$

- ***Bicubic spline***



Input:

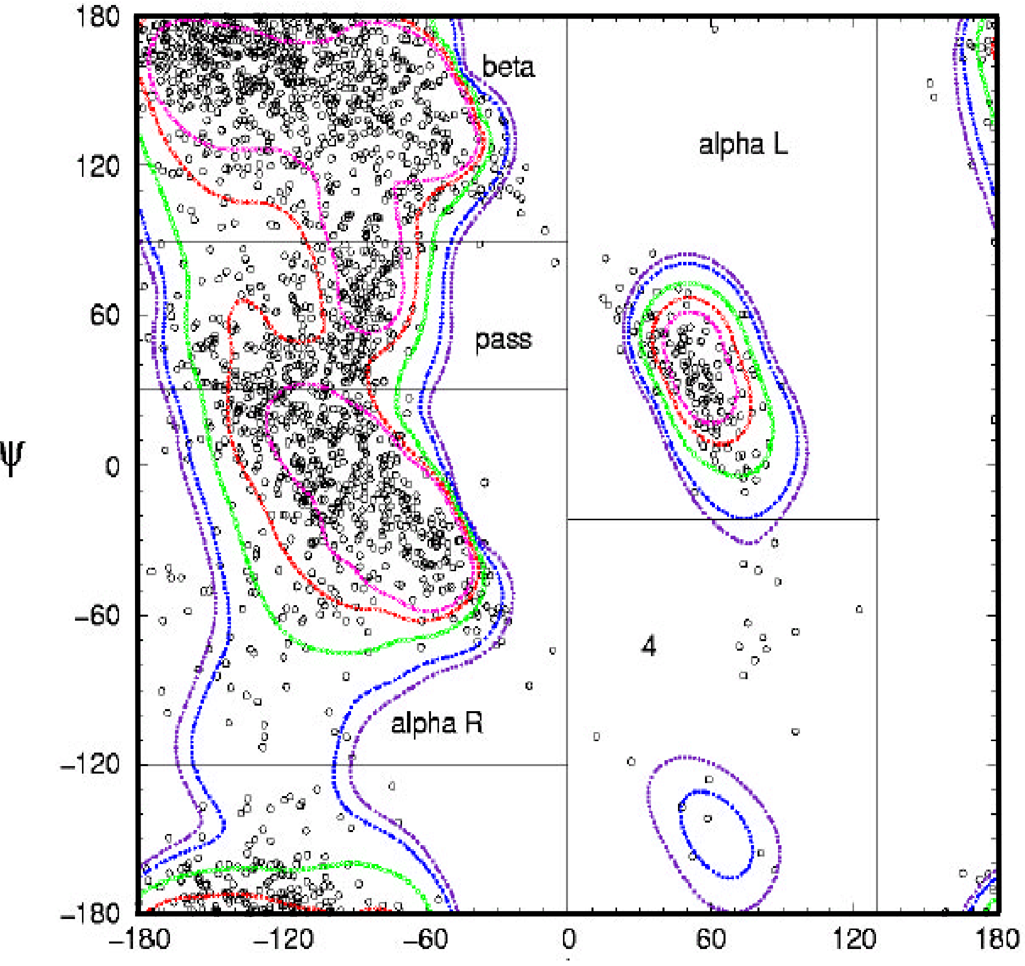
- z
- $\partial z/\partial \mathbf{x}$
- $\partial z/\partial \mathbf{y}$
- $\partial z^2/\partial \mathbf{x}\partial \mathbf{y}$

Output:

- $z(\mathbf{x}_i,\mathbf{y}_i)$
- **Smooth first derivative**
- **Continuous second derivatives**

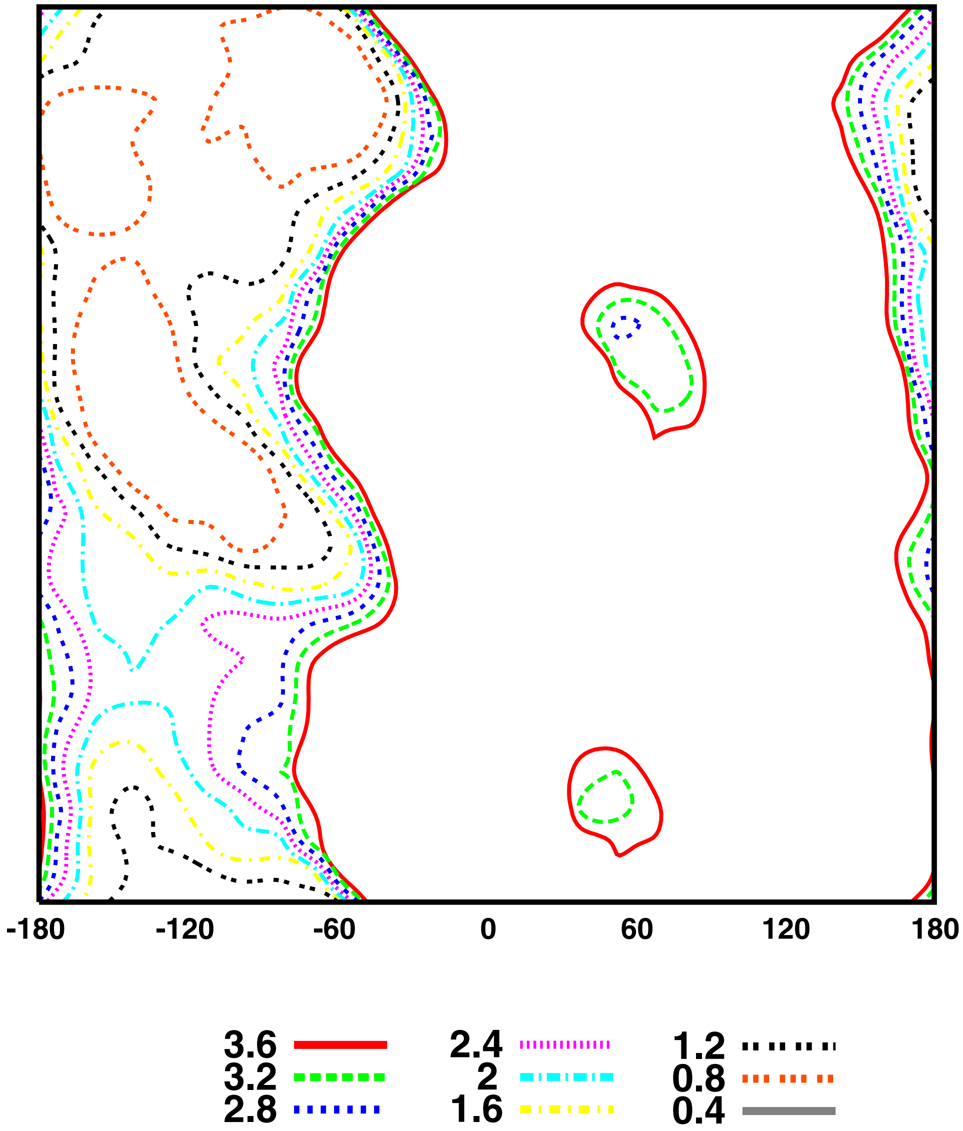
Comparison of QM/MM, PDB and AMOEBA Results

QM/MM *vs.* PDB



AMOEBA

(Fixed Charge Water)

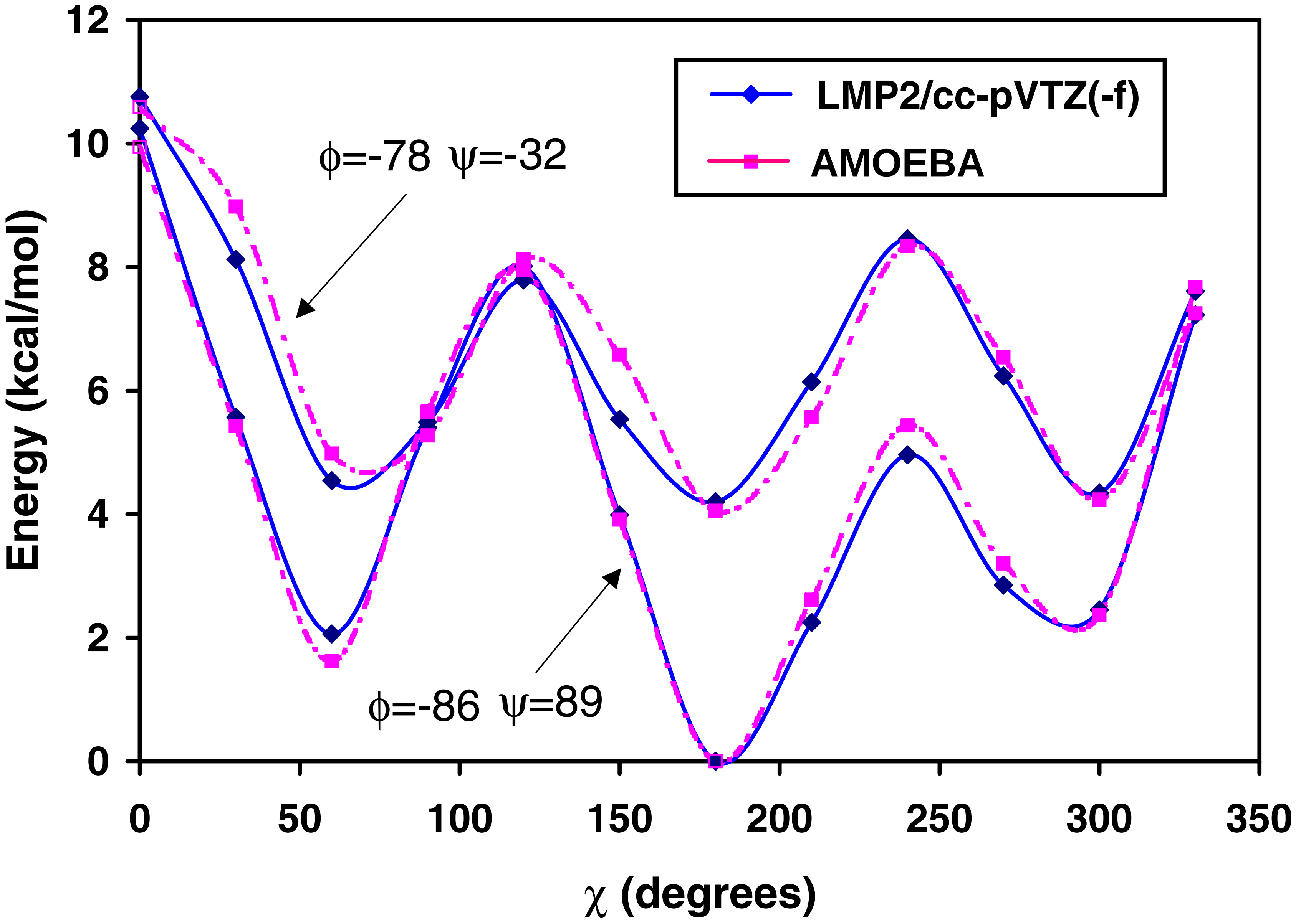


QM/MM, Jan Hermans, UNC
PDB, Jane Richardson, Duke

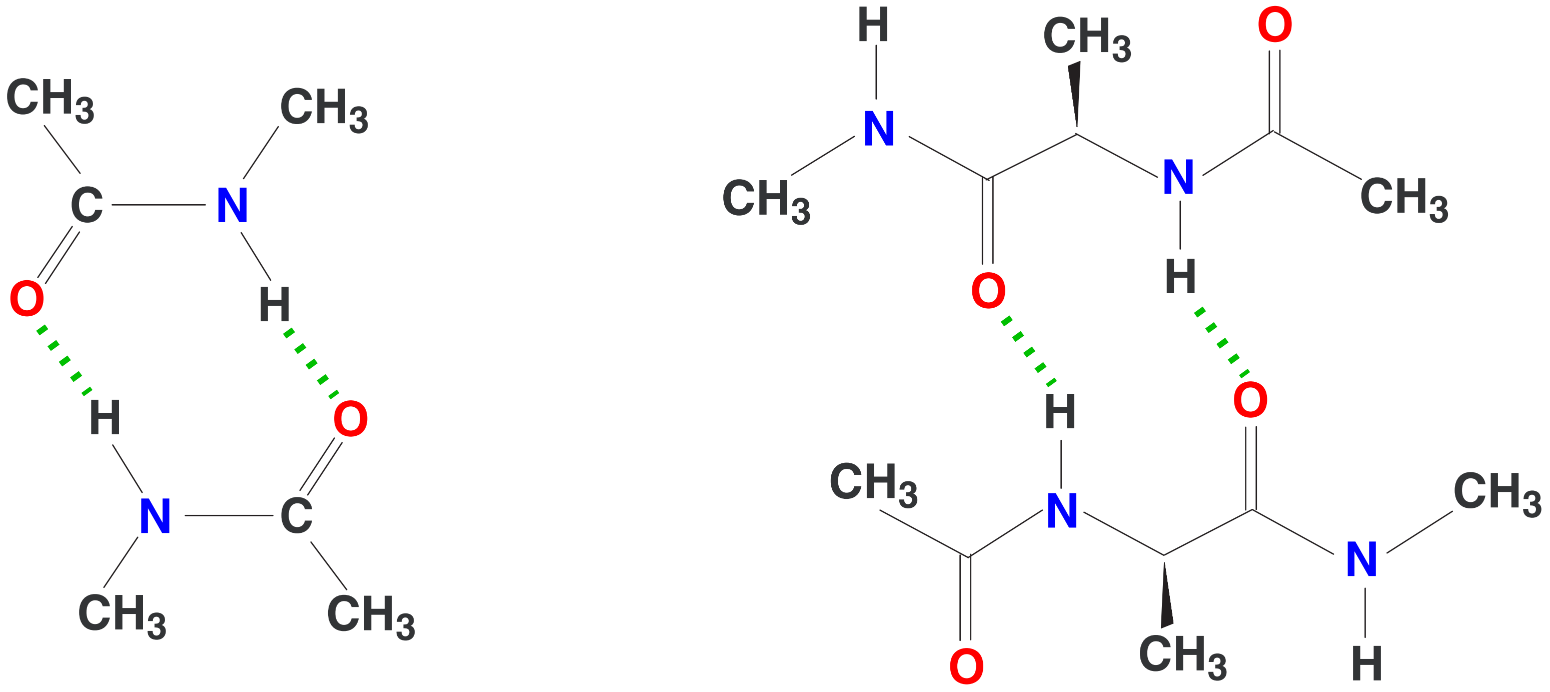
Conformational Populations

	Alpha	*Pass*	*Beta*	*Other*
Amber ff94	68	5	26	1
Amber ff99	77	10	13	1
CHARMM27	46	2	52	0
OPLS-AA	13	9	75	3
OPLS-AA/L	23	8	65	4
SCCDFTB (Amber)	27	16	48	9
SCCDFTB (CHARMM)	33	14	48	4
SCCDFTB (CEDAR)	27	12	61	0
AMOEBA (Polar Water)	29	16	54	1
AMOEBA (Fixed Water)	32	13	54	1

Valine Sidechain Energetics



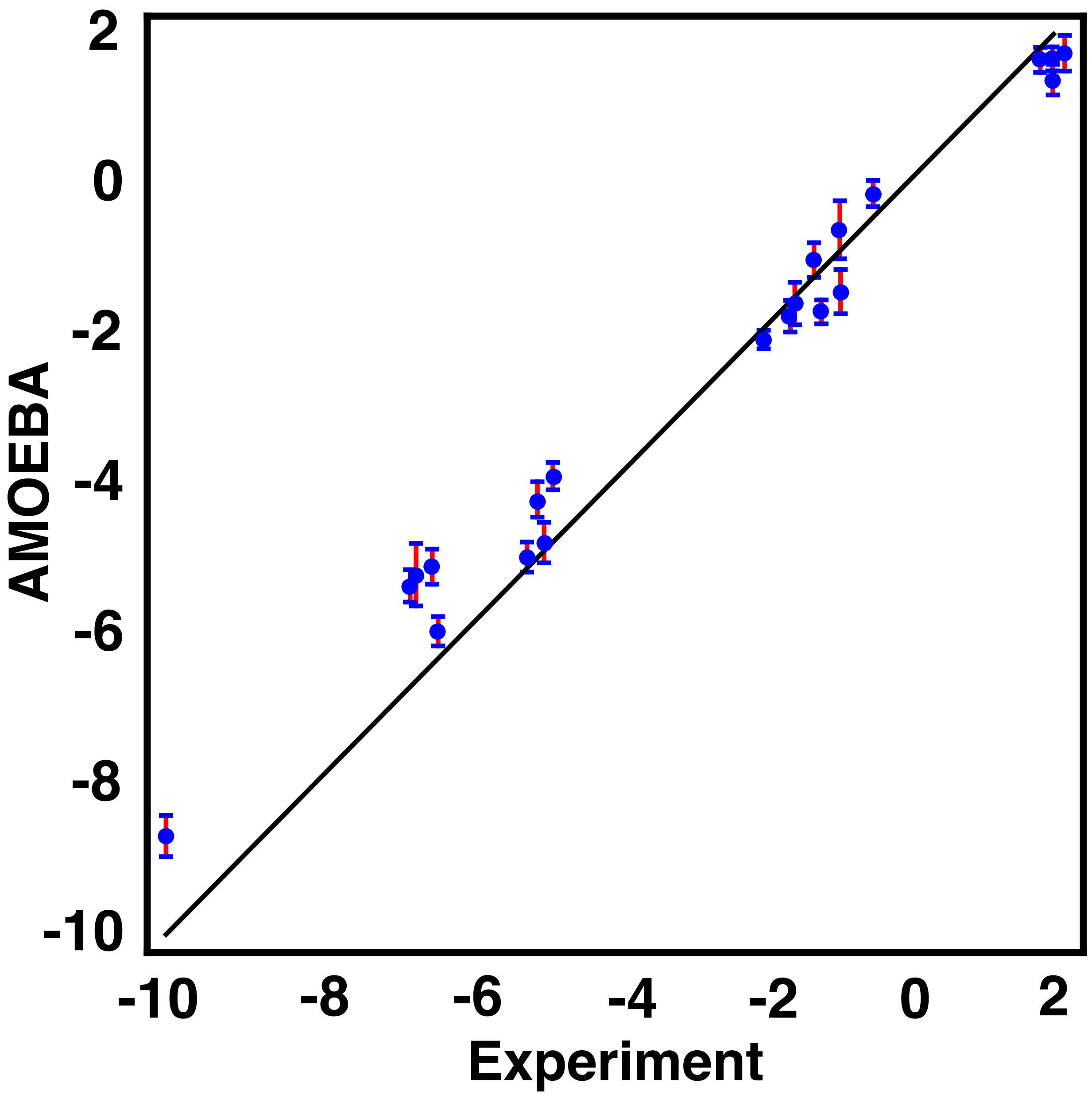
cis-N-Methylacetamide vs β-Sheet Model



	cis-NMA	β-Sheet	ΔE
MP2/(CEP)4-31G+(2d)	-20.5	-17.5	+3.0
BP/DZVP (BSSE)	-16.2	-8.4	+7.8
SIBFA	-18.7	-17.1	+1.6
TINKER	-17.3	-11.5	+5.8
AMBER94	-11.3	-14.8	-3.5
CHARMM27	-11.6	-16.9	-5.3
OPLS-AA	-11.5	-16.9	-5.4

[#] **QM and SIBFA data from Gresh, *et al.*, JACS, *121*, 7885-7894 (1999)**

Hydration Free Energies

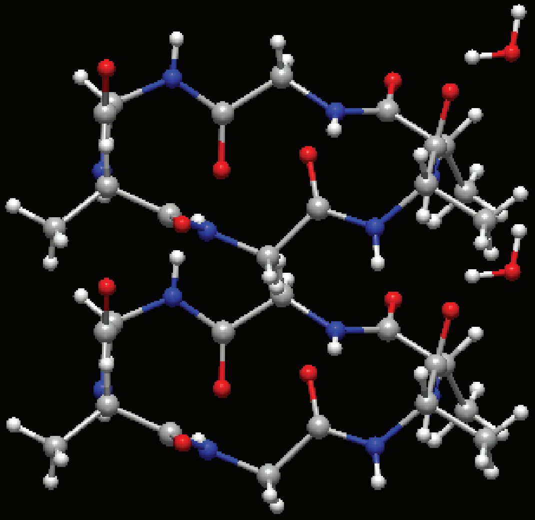


RMS Error = 0.65 kcal/mol
Median Unsigned Error = 0.50 kcal/mol

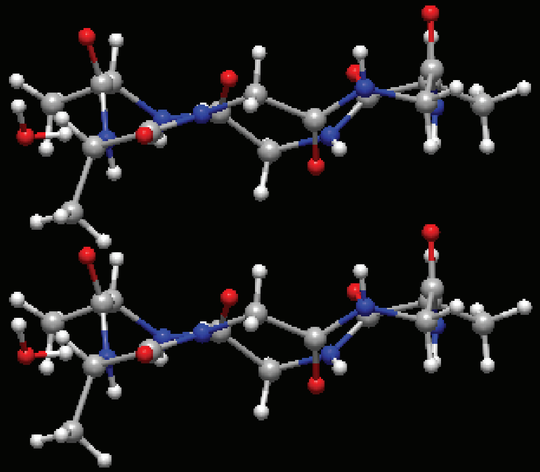
(Chuanjie Wu, Jay Ponder, John Chodera, David Mobley, Imran Haque, Vijay Pande)

MOLECULE	Experiment	AMOEBA	DELTA
Phenol	-6.62	-5.05	-1.57
Acetic Acid	-6.70	-5.63	-1.07
N-Methylacetamide	-10.0	-8.66	-1.34
p-Cresol	-6.40	-5.60	-0.80
Isopropanol	-4.77	-4.21	-0.56
Ethanol	-4.98	-4.69	-0.29
Water	-6.32	-5.86	-0.46
H_2S	-0.44	-0.41	-0.03
Toluene	-0.89	-1.53	0.64
Methanethiol	-1.24	-1.44	0.20
Methanol	-5.12	-4.79	-0.33
Propanol	-4.89	-4.85	-0.04
Ethylbenzene	-0.70	-0.80	0.10
Methylethylsulfide	-1.50	-1.98	0.44
DImethylether	-1.92	-2.22	0.30
Ethane	1.82	1.73	0.09
Dimethylsulfide	-1.56	-1.85	0.29
Methane	1.99	1.73	0.26
n-Butane	2.15	1.11	1.04
Benzene	-0.88	-1.23	0.35
Ethanethiol	-1.14	-1.74	0.60
Propane	1.99	1.69	0.30

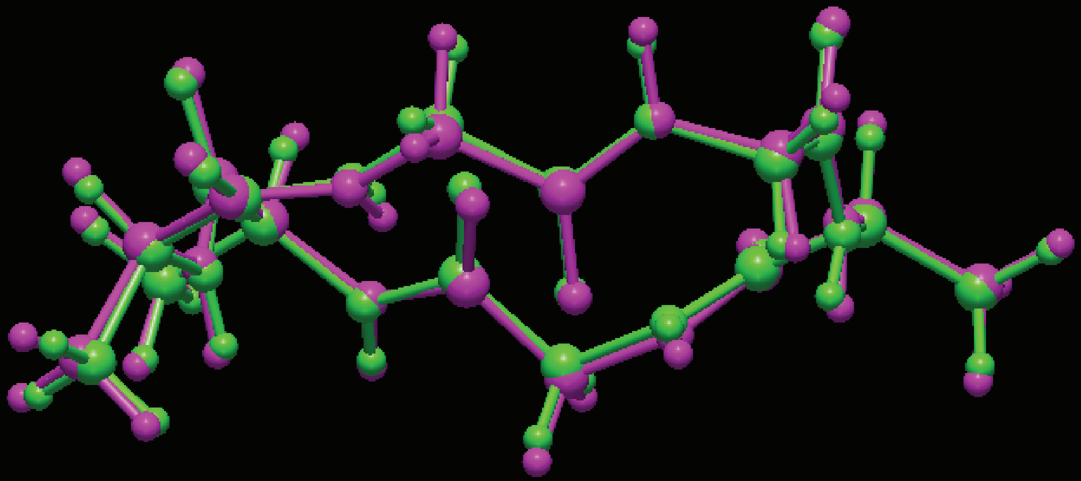
Peptide Crystals as Models for Protein Structure & Folding



Cyclic Ala-Ala-Gly-Gly-Ala-Gly-
+ 1 H_2O ($P2_1$)



AMOEBA *RMS* = 0.12 Ang

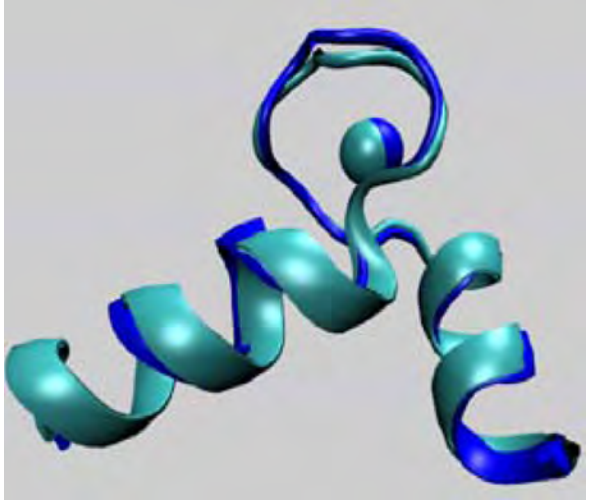


Intramolecular C=O...H-N
Hydrogen Bond Angle

X-Ray	**144 degrees**
OPLS-AA	**158 degrees**
AMOEBA	**142 degrees**

AMOEBA Binding Free Energies

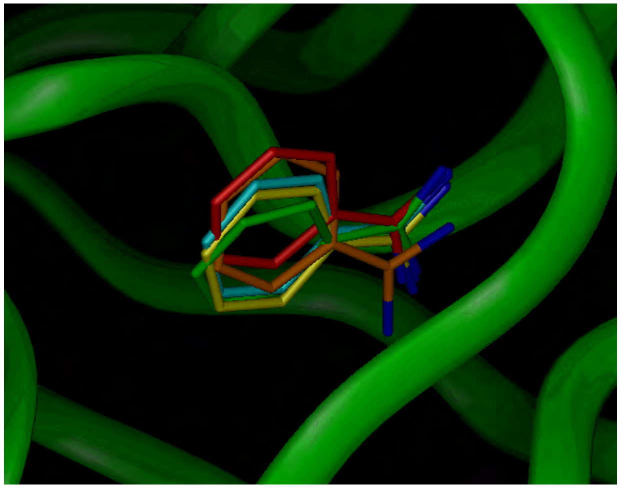
EF Hand: Relative Ca^{+2}/Mg^{+2} Binding Affinity



Wild Type: ~10^4x (expt)
6.6 kcal/mol (calc)

Glu -> Asp: ~10 x (expt)
1.3 kcal/mol (calc)

Trypsin-Benzamidine: Absolute Binding Affinity



6.3 to 7.3 kcal/mol (expt)

6.7 +/- 0.6 kcal/mol (calc)

Pengyu Ren, U Texas