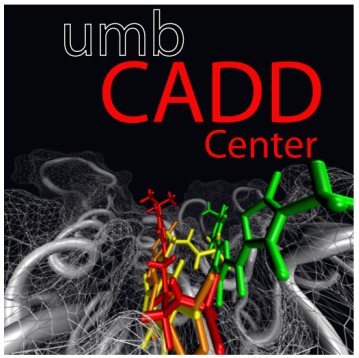
The CHARMM Force Field

Alexander D. MacKerell, Jr.
University of Maryland, Baltimore
School of Pharmacy

CECAM Workshop: Advances in Biomolecular Modelling and Simulations using CHARMM
University College Dublin, Ireland
June 5-8, 2012



UNIVERSITY of MARYLAND
SCHOOL OF PHARMACY

Outline

- Part I: the CHARMM Force Field
  - Introduction: molecular mechanics
  - The CHARMM potential energy function
  - CHARMM recent developments & future prospects
- Part II: Parametrization tutorial
  - The CHARMM General Force Field (CGenFF)
  - Parametrization procedure
  - Some afterthoughts…

Molecular Mechanics

- Simulate reality in atomistic detail
  - "Ultimate microscope"
  - Study behavior of proteins in their native environment
  - Nanotech,…
- Microscopic Reality = Quantum Mechanics (QM)
  - Can be simulated, but not at biologically relevant system sizes and time scales

  $\Rightarrow$ Need for simplified model
- Molecular mechanics
  - Approximates reality using classical mechanics
  - Different models = Force Fields

Elements of a force field

- Potential energy function = mathematical equation

$$V = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{\substack{improper \\ dihedrals}} K_\varphi (\varphi - \varphi_0)^2 + \sum_{dihedrals} \sum_{n=1}^{6} K_{\phi,n} (1 + \cos(n\phi - \delta_n))$$

$$+ \sum_{\substack{nonbonded \\ pairs\ ij}} \frac{q_i q_j}{4\pi D r_{ij}} + \sum_{\substack{nonbonded \\ pairs\ ij}} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right]$$

- Many parameters: $K_b, b_0, \ldots, K_{\phi,n}, n, \delta_n, q_i, \varepsilon, R_{min}$
- Force field
  = Potential energy function + parameter set
  – Relates chemical structure and conformation to energy

Structure – energy relation

- Structure: n atoms
- Each atom has an x, y and z coordinate
  $\Rightarrow 3n$ coordinates

$$V = f\left(x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_n, y_n, z_n\right)$$

$$\vec{F}_1 = \frac{\partial V}{\partial x_1}\vec{1}_x + \frac{\partial V}{\partial y_1}\vec{1}_y + \frac{\partial V}{\partial z_1}\vec{1}_z$$

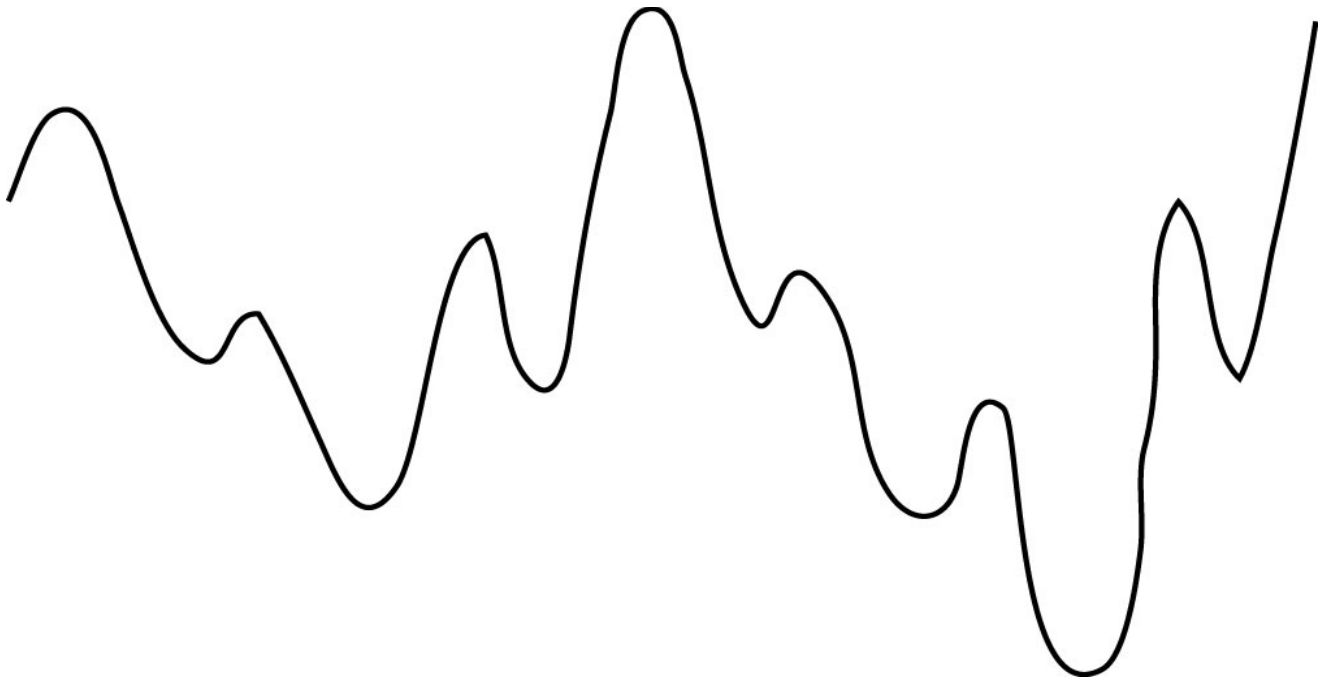
$$\vdots$$

$$\vec{F}_n = \frac{\partial V}{\partial x_n}\vec{1}_x + \frac{\partial V}{\partial y_n}\vec{1}_y + \frac{\partial V}{\partial z_n}\vec{1}_z$$

- Force field gives energy and forces as a simple, analytical* function in $3n$ dimensions

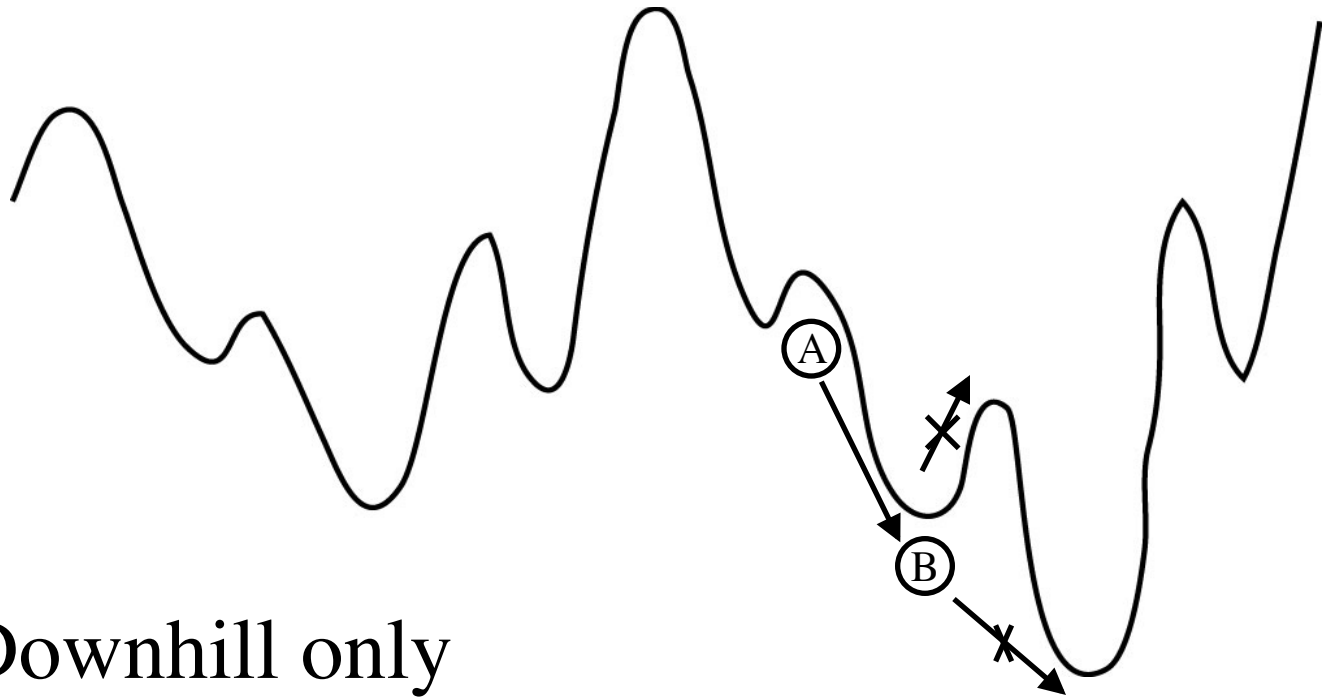
* For additive force fields.

2D representation of (3*n*)D potential energy surface



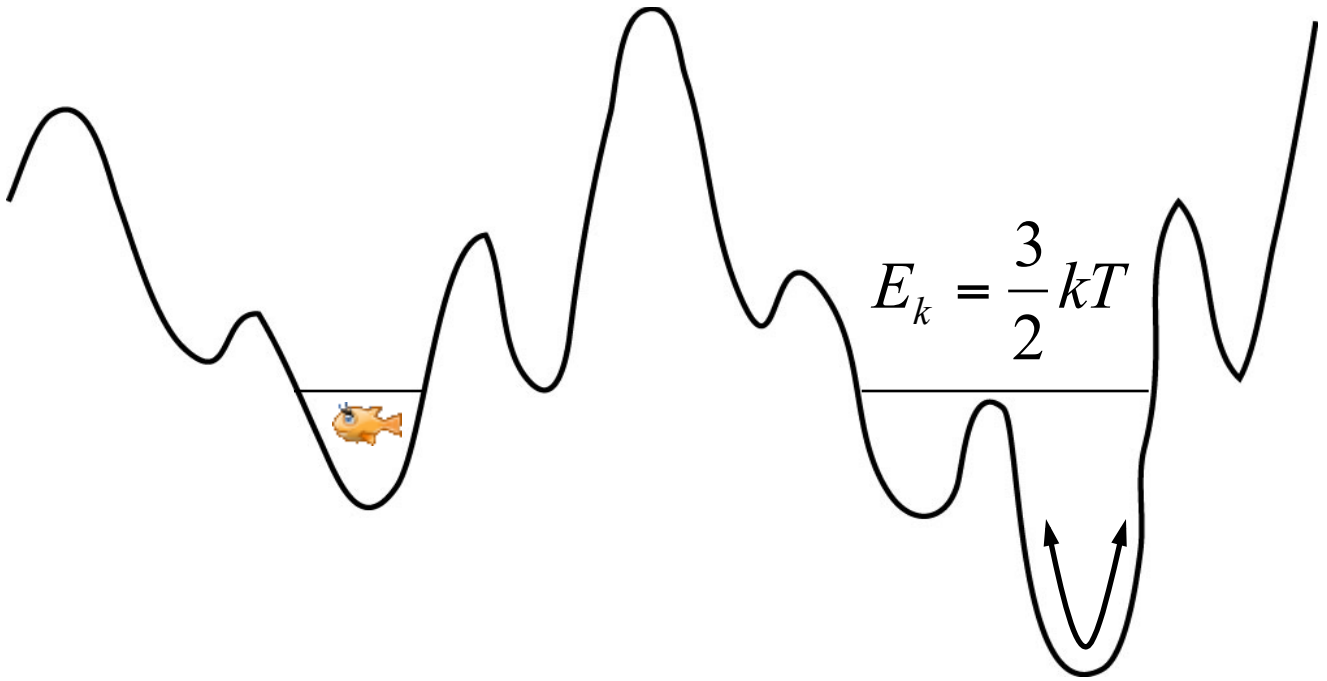
What can we do with it?

Energy minimization

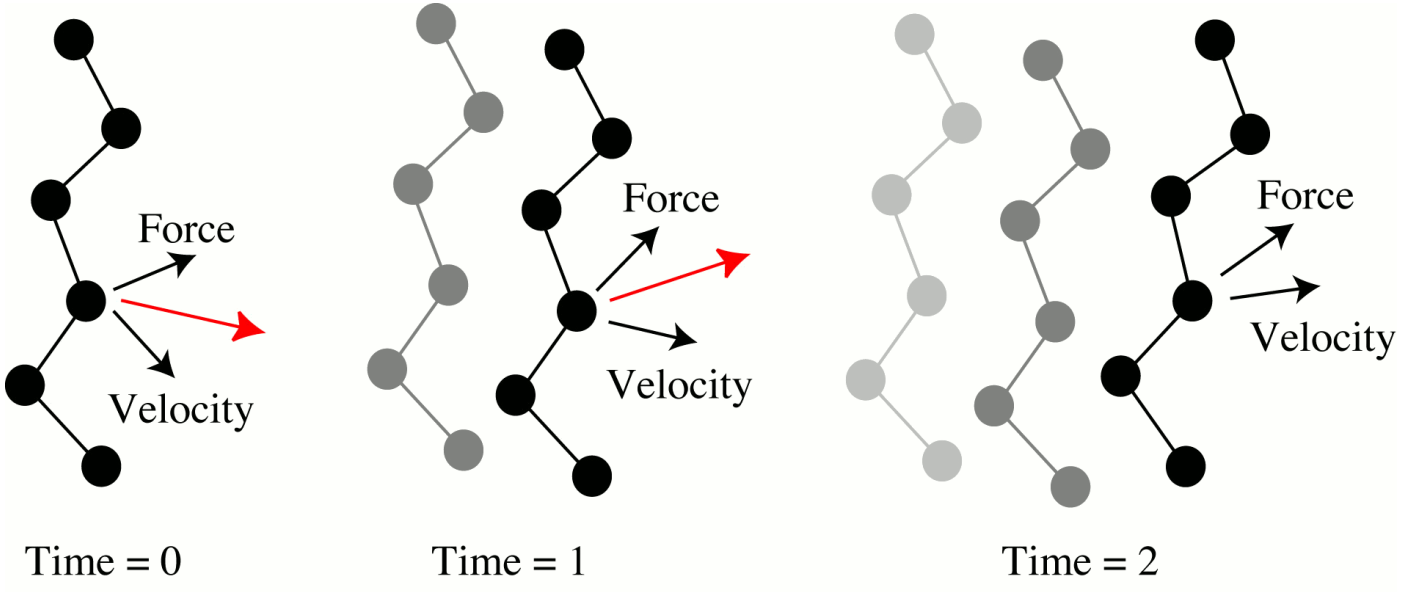


- Downhill only
- No barrier crossing

Molecular Dynamics (MD) simulation



Molecular Dynamics (MD) simulation



Common “additive” empirical force fields

- **Class I**
  - CHARMM
  - CHARMm (Accelrys)
  - AMBER
  - OPLS/AMBER/Schrödinger
  - ECEPP (free energy force field)
  - GROMOS

- **Class II**
  - CFF95 (Accelrys)
  - MM3
  - MMFF94 (CHARMM, Macromodel, MOE, elsewhere)
  - UFF, DREIDING

The additive CHARMM force field

- Mostly aimed at biomolecular simulations (in a broad sense)
- Multiple files:
  - Proteins “CHARMM22/CMAP”
    - C36 manuscript in review
  - Nucleic acids “CHARMM36”
  - Lipids “CHARMM36”
  - Carbohydrates “CHARMM35 (36)”
  - General organic molecules “CGenFF 2b7”

The potential energy function

- $V_{total} = V_{internal} + V_{external}$
- Bonded (aka. internal or intramolecular) energy
  $V_{internal} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{impropers} + \ldots$
- Nonbonded (aka. external or intermolecular) energy
  $V_{external} = V_{van\ der\ Waals} + V_{electrostatic}$

CHARMM bonded energy function and corresponding force field parameters

$$\sum_{bonds} K_b\left(b - b_0\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_0\right)^2 + \sum_{dihedrals} \sum_{n=1}^{6} K_{\phi,n}\left(1 + \cos\left(n\phi - \delta_n\right)\right)$$

$$+ \sum_{impropers} K_\varphi\left(\varphi - \varphi_0\right)^2 + \sum_{Urey-Bradley} K_{UB}\left(r_{1,3} - r_{1,3;0}\right)^2 + \sum_{\phi,\psi} V_{CMAP}$$

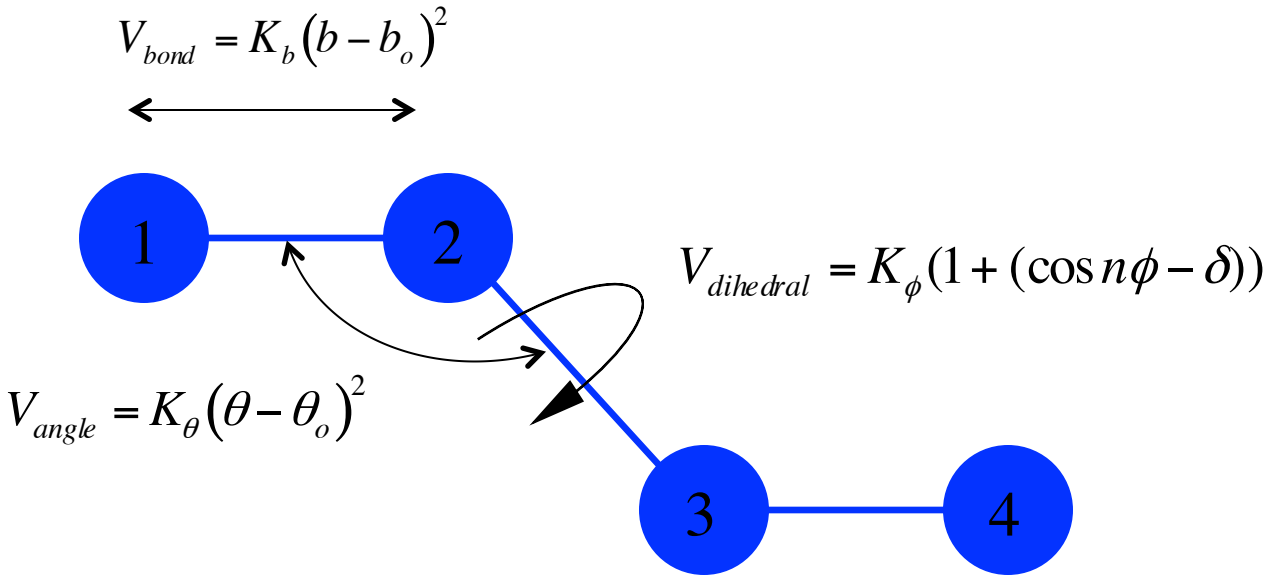
Equilibrium terms

- b_0: bonds
- θ_0: angles
- δ_n: dihedral phase
- φ_0: impropers
- $r_{1,3;0}$: Urey-Bradley

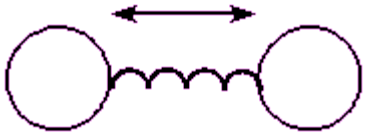
Force constants

- K_b: bonds
- K_θ: angles
- K_ϕ: dihedral amplitude
- K_ω: impropers
- K_{UB}: Urey-Bradley

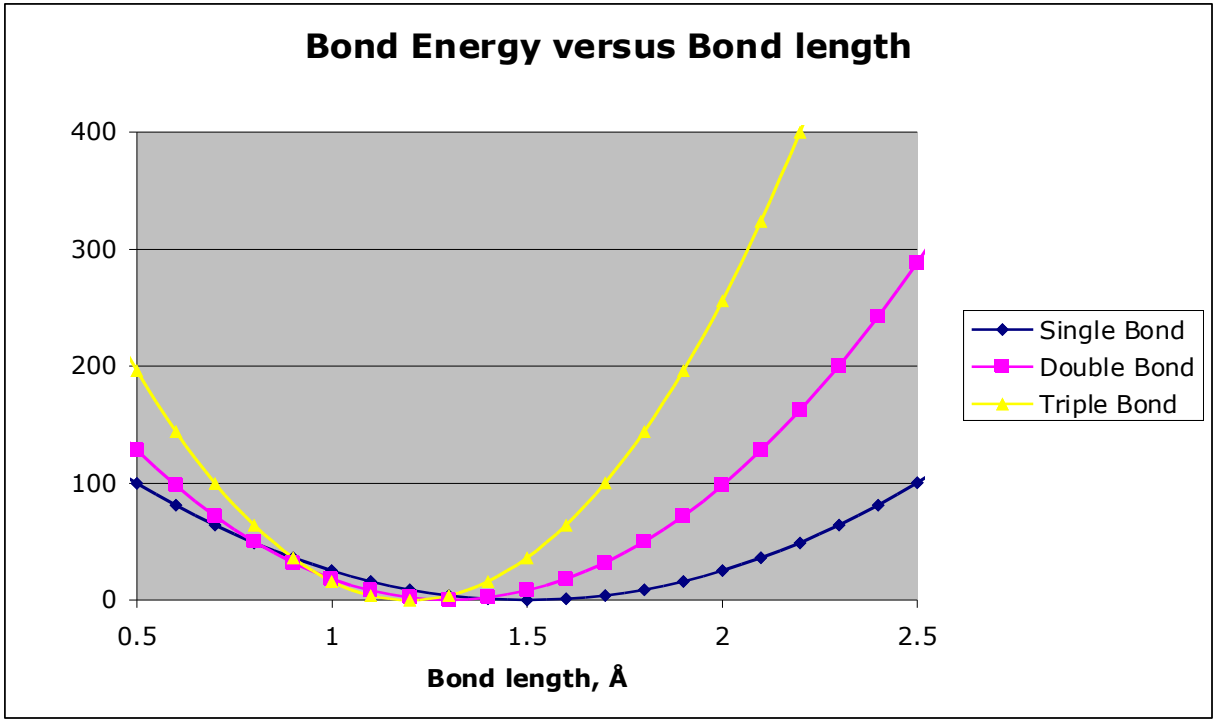
Diagram of intramolecular energy terms



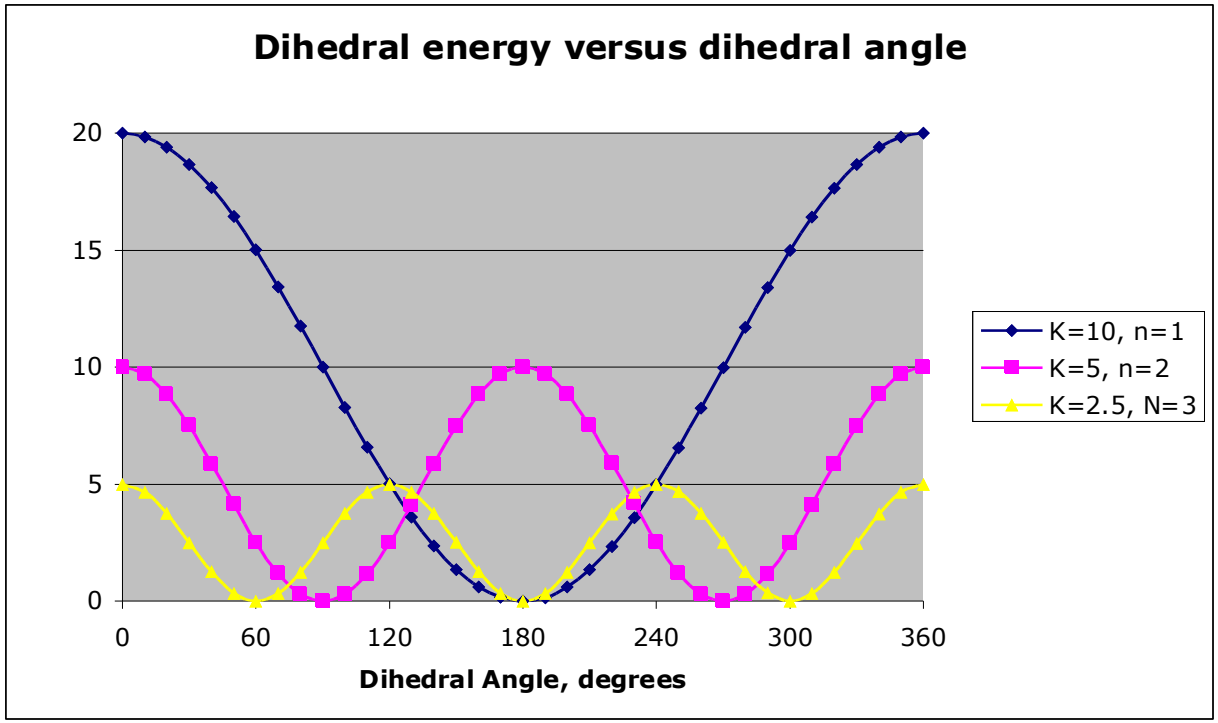
$$V_{bond} = K_b \left(b - b_o \right)^2$$



Chemical type	K_{bond}	b_o
C-C	100 kcal/mole/Å^2	1.5 Å
C=C	200 kcal/mole/Å^2	1.3 Å
C=-C	400 kcal/mole/Å^2	1.2 Å

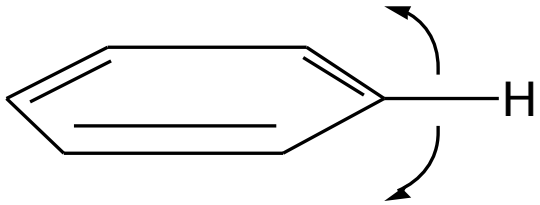


$$V_{dihedral} = K_{\phi}(1 + (\cos n\phi - \delta))$$

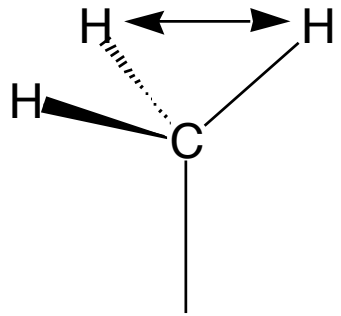


$\delta = 0^{\circ}$

Note use of a Fourier series for a dihedral

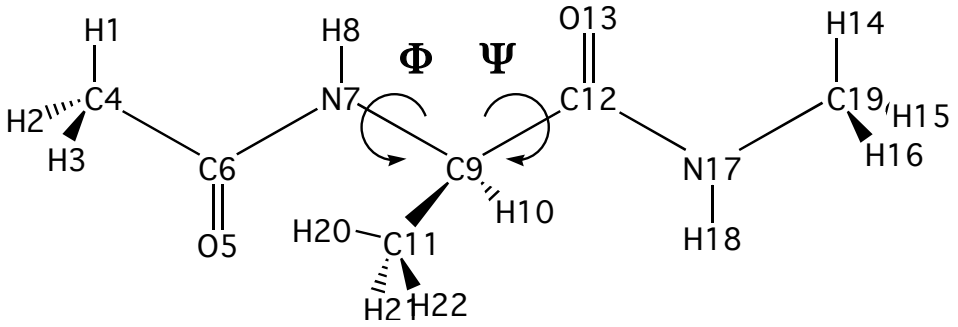


$$V_{improper} = K_{\varphi}\left(\varphi - \varphi_o\right)^2$$



$$V_{Urey-Bradley} = K_{UB}\left(r_{1,3} - r_{1,3o}\right)^2$$

2D dihedral energy correction map to the CHARMM 22 ϕ,ψ backbone (CMAP)

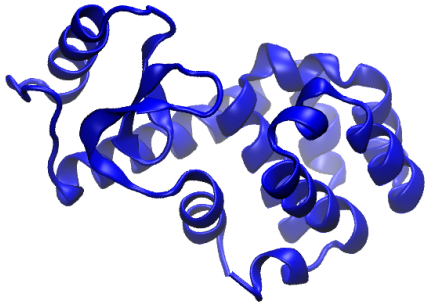


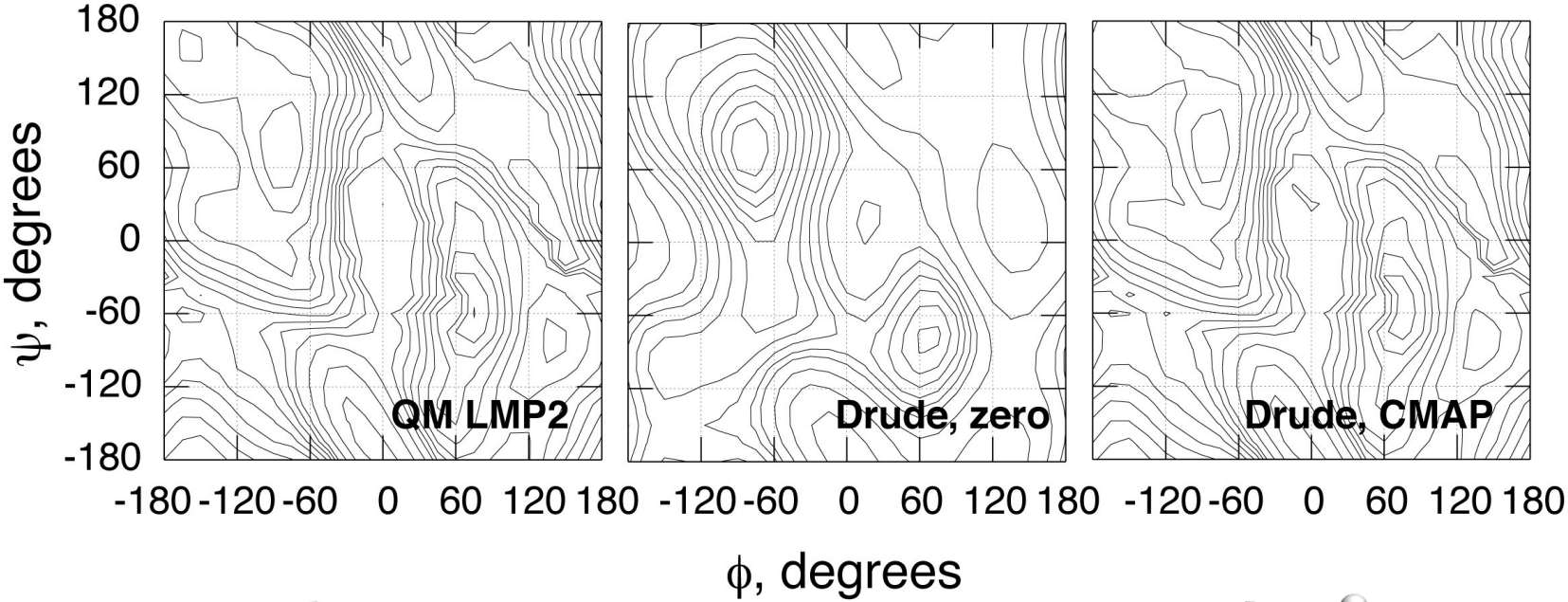
- ϕ,ψ grid-based energy correction

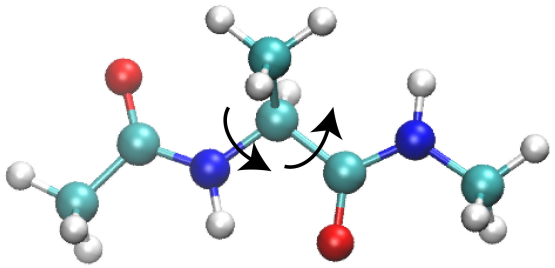
$$V_{CMAP} = f(\phi,\psi) = \sum_{i=0}^{3}\sum_{j=0}^{3} c_{ij}\left(\frac{\phi-\phi_L}{\Delta_\phi}\right)^i\left(\frac{\psi-\psi_L}{\Delta_\psi}\right)^j$$

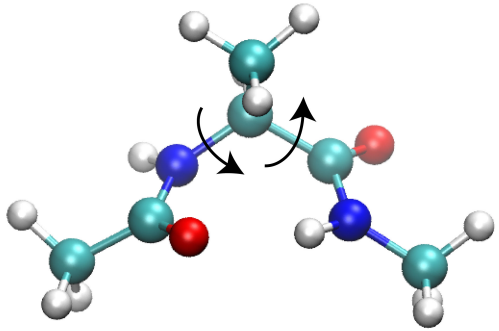
- c_{ij} determined from energies at grid points by bicubic spline interpolation ⇒ smooth first derivatives and continuous second derivatives

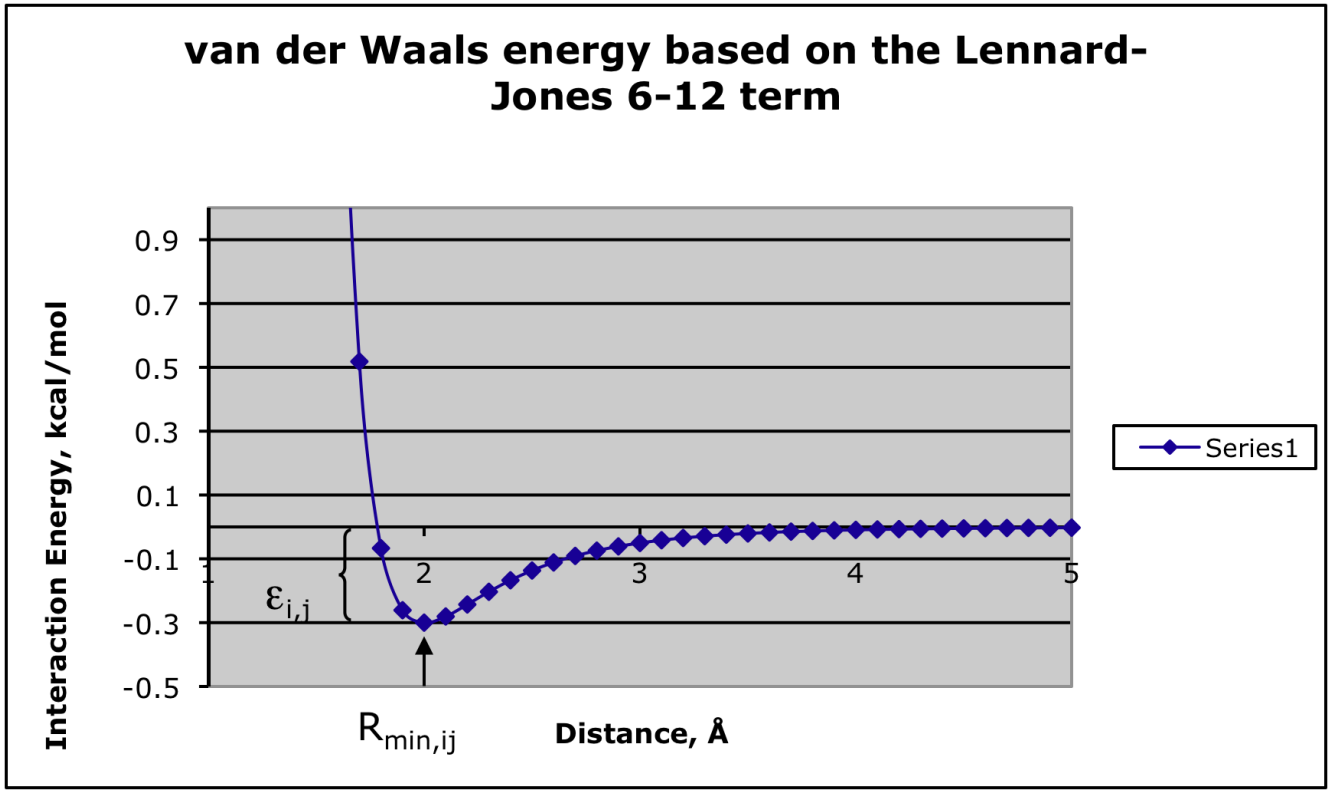
Towards proteins:
Alanine dipeptide ϕ,ψ surfaces



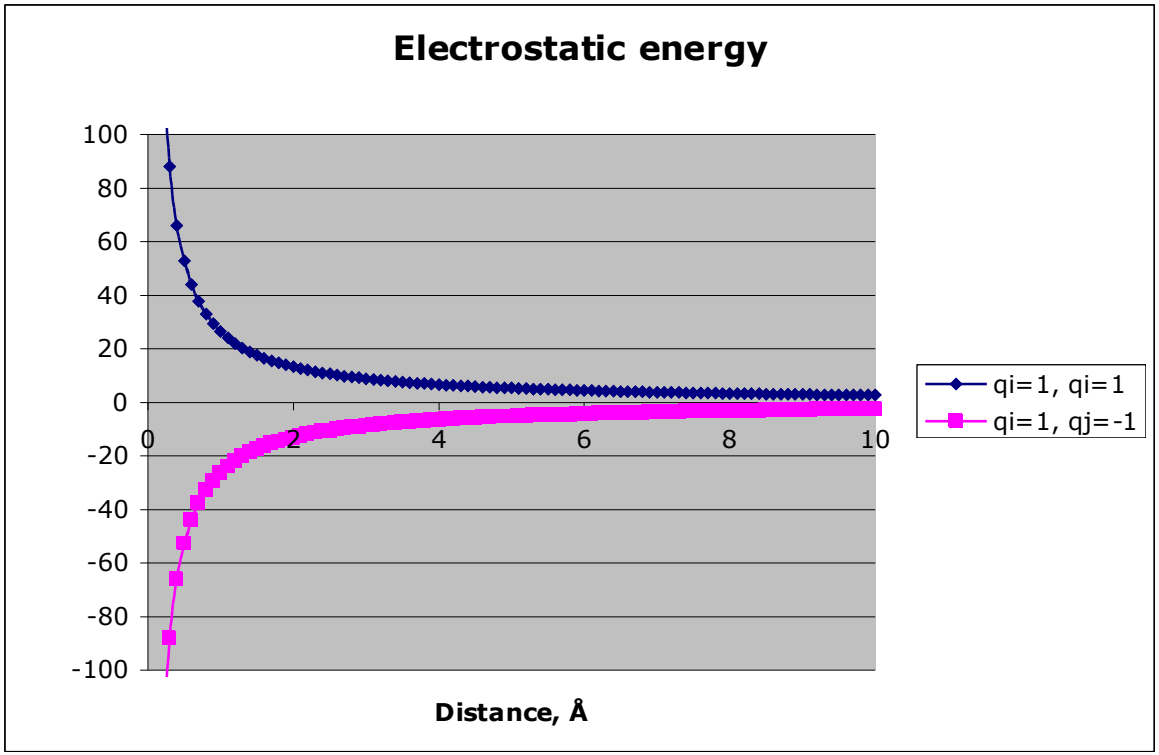




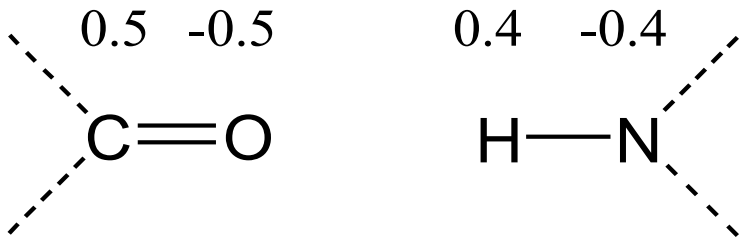




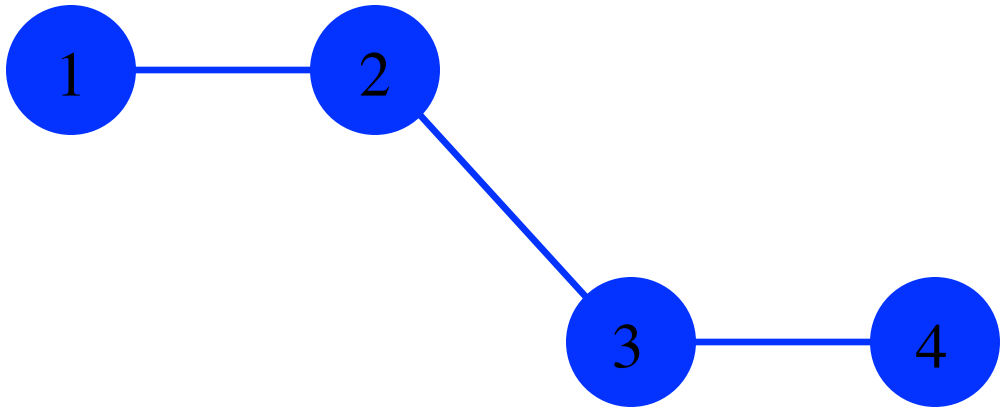
$$\varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$



Treatment of hydrogen bonds: partial atomic charges yield appropriate interaction energies



Example of nonbond exclusions



nonbonded (intermolecular) interactions between bonded atoms are treated with special rules

1,2 interactions: excluded from energy calculation

1,3 interactions: excluded

1,4 interactions: included or scaled

> 1,4 interactions: included

Alternate treatment of nonbond interactions and future of force fields

Variations on treatment of nonbond terms in additive force fields to the explicit treatment of electronic polarizability

Example Class II energy function

$$\sum_{bonds}\left[K_{b,2}(b-b_o)^2+K_{b,3}(b-b_o)^3+K_{b,4}(b-b_o)^4\right]$$

$$+\sum_{angles}\left[K_{\theta,2}(\theta-\theta_o)^2+K_{\theta,3}(\theta-\theta_o)^3+K_{\theta,4}(\theta-\theta_o)^4\right]$$

$$+\sum_{dihedrals}\left[K_{\phi,1}(1-\cos\phi)+K_{\phi,2}(1-\cos 2\phi)+K_{\phi,3}(1-\cos 3\phi)\right]$$

$$+\sum_{impropers}K_{\chi}\chi^2$$

$$+\sum_{bonds}\sum_{bonds'}K_{bb'}(b-b_o)(b'-b_o')+\sum_{angles}\sum_{angles'}K_{\theta\theta'}(\theta-\theta_o)(\theta'-\theta_o')$$

$$+\sum_{bonds}\sum_{angles}K_{b\theta}(b-b_o)(\theta-\theta_o)$$

$$+\sum_{bonds}\sum_{dihedrals}(b-b_o)\left[K_{\phi,b1}\cos\phi+K_{\phi,b2}\cos 2\phi+K_{\phi,b3}\cos 3\phi\right]$$

$$+\sum_{bonds'}\sum_{dihedrals}(b'-b_o')\left[K_{\phi,b'1}\cos\phi+K_{\phi,b'2}\cos 2\phi+K_{\phi,b'3}\cos 3\phi\right]$$

$$+\sum_{angles}\sum_{dihedrals}(\theta-\theta_o)\left[K_{\phi,\theta1}\cos\phi+K_{\phi,\theta2}\cos 2\phi+K_{\phi,\theta3}\cos 3\phi\right]$$

$$+\sum_{angles}\sum_{angles'}\sum_{dihedrals}(\theta-\theta_o)(\theta'-\theta_o')\cos\phi$$

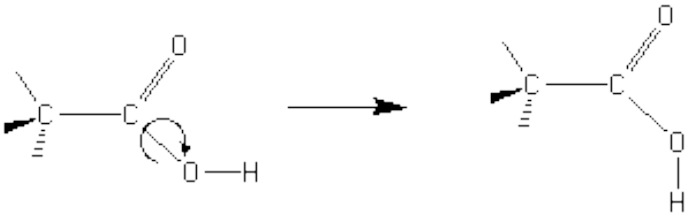
Alternate intermolecular terms for the electrostatic (additive) or vdW interactions

$$V_{Hbond} = \sum_{Hbonds} \varepsilon_{HB} \left[\left(\frac{R_{HB,A-H}}{r_{A-H}} \right)^{12} - \left(\frac{R_{HB,A-H}}{r_{A-H}} \right)^{10} \right] * \cos\left(\theta_{A-H-D}\right)$$

$$V_{vdw} = \sum_{vdw} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{9} - \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right]$$

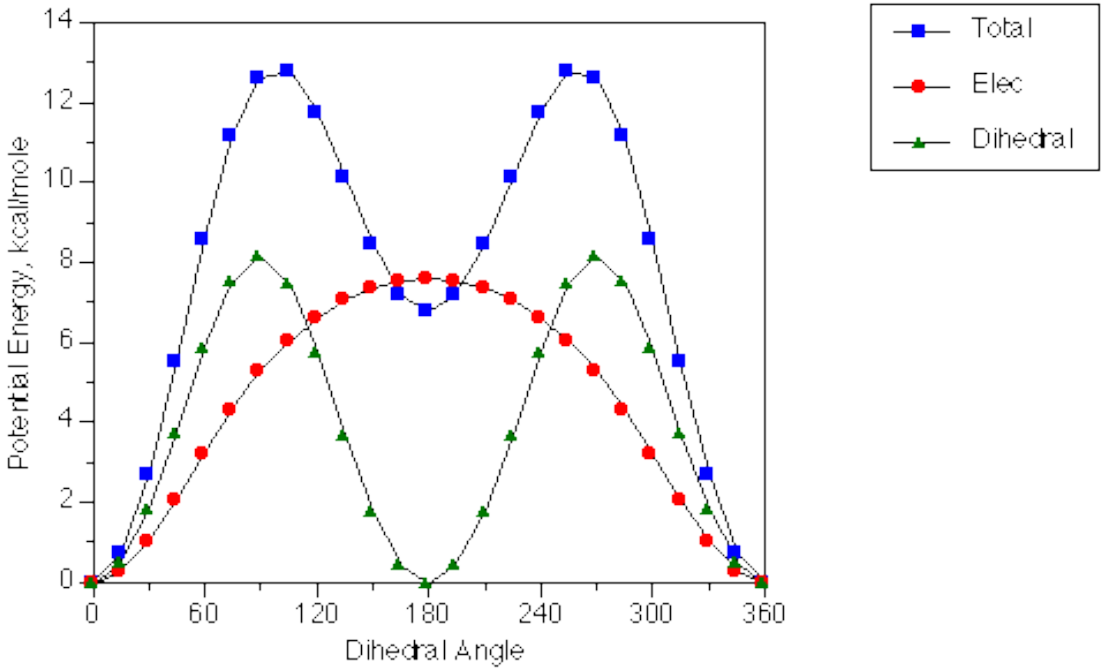
$$V_{vdw} = \sum_{vdw} \varepsilon_{ij} \left(e^{\frac{-aR_{\min,ij}}{r_{ij}}} - \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right)$$

Total potential energy for rotation of O-C-O-H dihedral in acetic acid



cis, O-C-O-H = 0.0

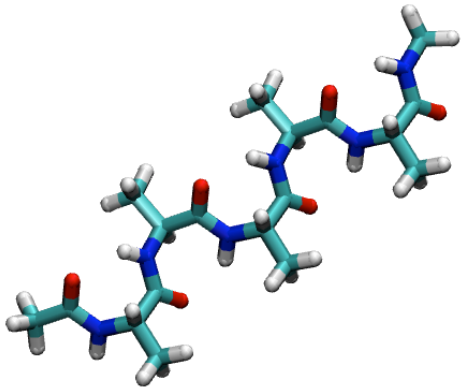
trans, O-C-O-H = 180.0



Summary of recent developments and future prospects

- Recent developments:
  - Proteins: “CHARMM36”
  - Nucleic Acids “CHARMM36”
  - Lipids : Alex Sodt tomorrow!
  - Carbohydrates “CHARMM36”
  - General organic molecules “CGenFF 2b7”

- Beyond the class I additive Potential Energy Function: alternate treatment of nonbond interactions and future of force fields.



C36 all-atom protein FF

Correct for conformational equilibrium of unstructured peptides in solution and improve treatment of sidechain conformational sampling

X^{2*} values from comparison of experimental NMR J-coupling constants

	Ala_3	**Ala_5**	**Ala_7**	**Val_3**	**Gly_3**	**GPGG**
C36	0.61	0.74	0.43	2.33	2.33	2.78
C22/CMAP		1.73			3.68	13.37
Amber ff99SB		1.4			3.04	2.39
OPLS-AA		1.35				
Gromos 53a6		1.68				

* Normalized sum of squared deviations between observed and theoretical NMR J-coupling constants

% Secondary structure and comparison with NMR data for Ala_5 and $(AAQAA)_3$

Peptide	**Property**	**ff99SB**	**ff99SB***	**C22/CMAP**	**C36**
Ala_5					
	% α_+	15.7	22.5	41.5	13.1
	% β	37.8	34.5	25.3	30.8
	% ppII	42.3	39.8	26.2	51.9
	% α-helix	0.0	0.6	3.7	0.1
	$\chi^2(J)$ [Hz^2]	1.7	1.7	2.0	1.2
Ac-$(AAQAA)_3$-NH_2					
	% α_+	26.9	48.5	98.7	44.0
	% β	32.4	22.7	0.46	19.1
	% ppII	31.8	21.8	0.41	29.5
	% α-helix	1.8	14.2	95.3	21.0
	$\chi^2(\delta_{C'})$ [ppm^2]	0.5	0.06	2.6	0.04

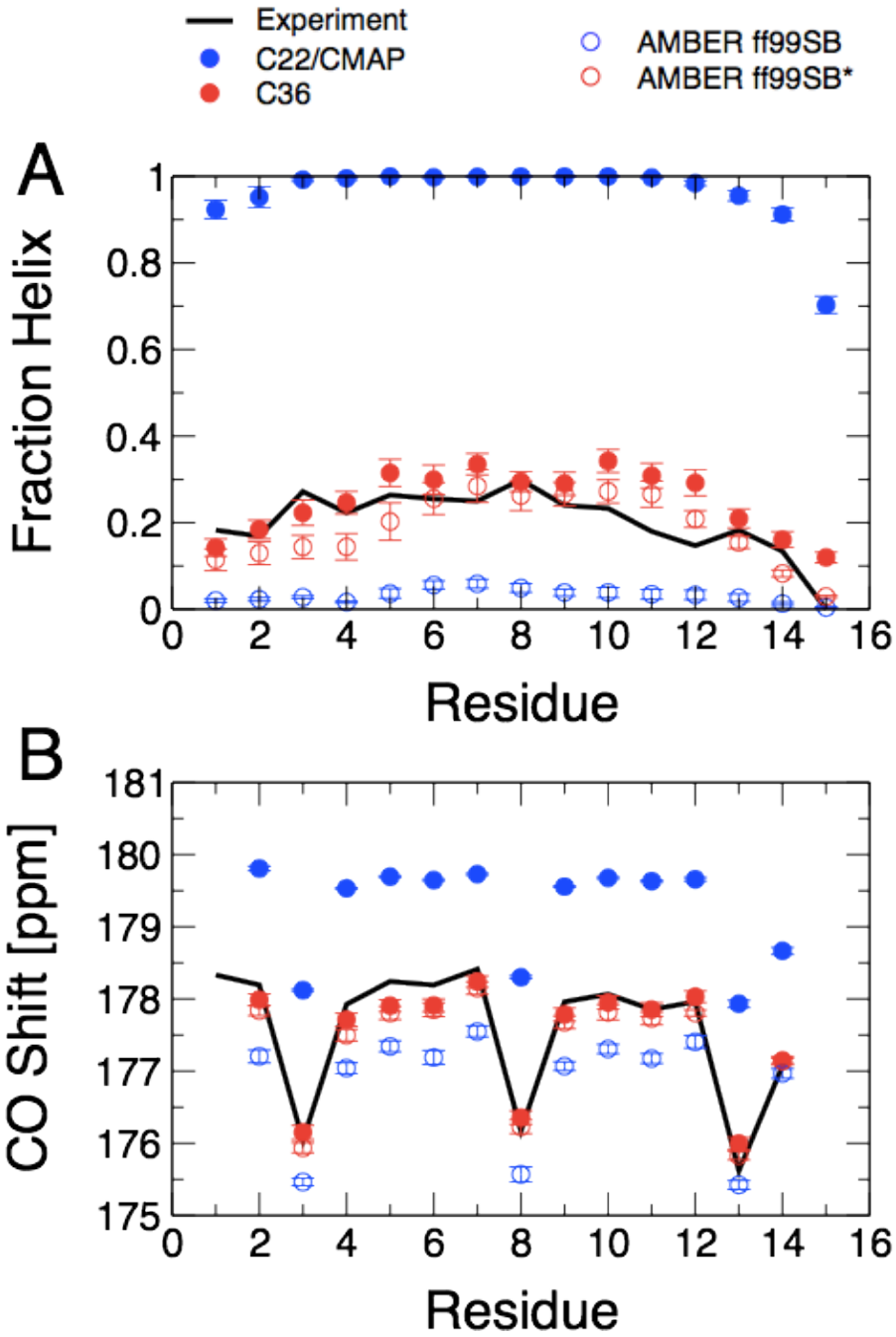
Ala_5 vs. J-coupling, $(AAQAA)_3$ vs. chemical shift (Sparta)

Comparison with side-chain NMR *J*-coupling data for folded and unfolded proteins

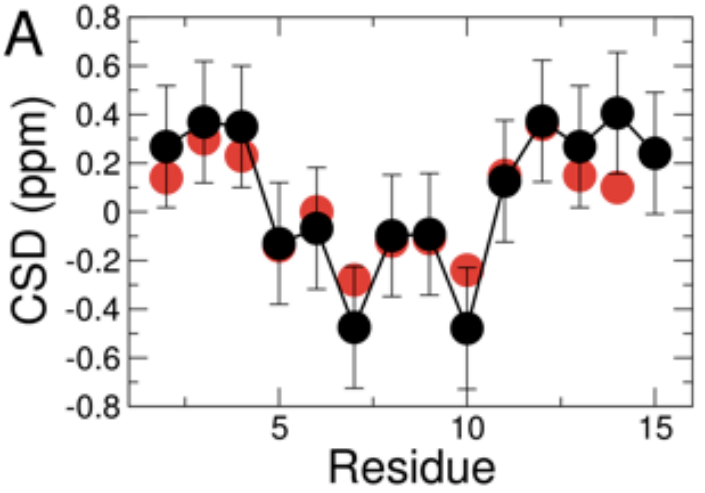
X^2 vs. NMR	Unfolded	Folded	All
Amber ff99SB-ILDN	9.7	8.9	9.3
C22/CMAP	9.8	9.2	9.5
C36	5.2	8.5	6.9

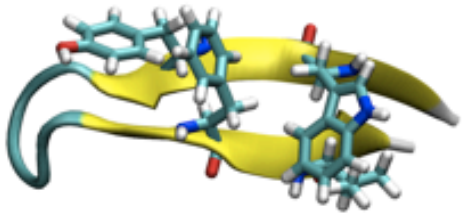
Unfolded proteins are ubiquitin and GB1 in urea and folded proteins are ubiquitin, GB3, BPTI and hen lysozyme. Results represent the average X^2 values over all amino acids.

Helix formation in Ac-$(AAQAA)_3$-NH_2.

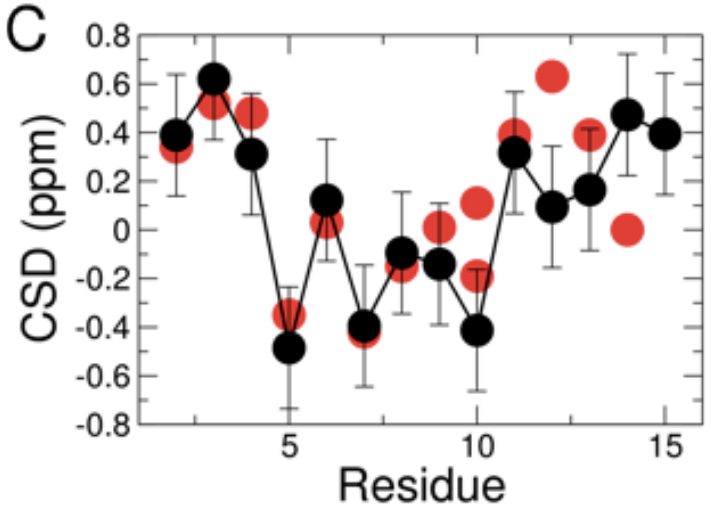


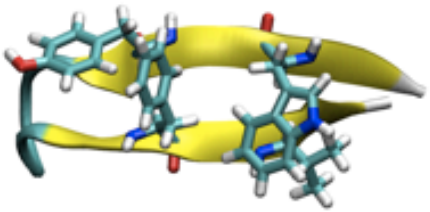
β-hairpin folding test. H^{α} chemical shifts from experiment (red) and calculated from simulation using SPARTA+ (black)





GB1 Hairpin

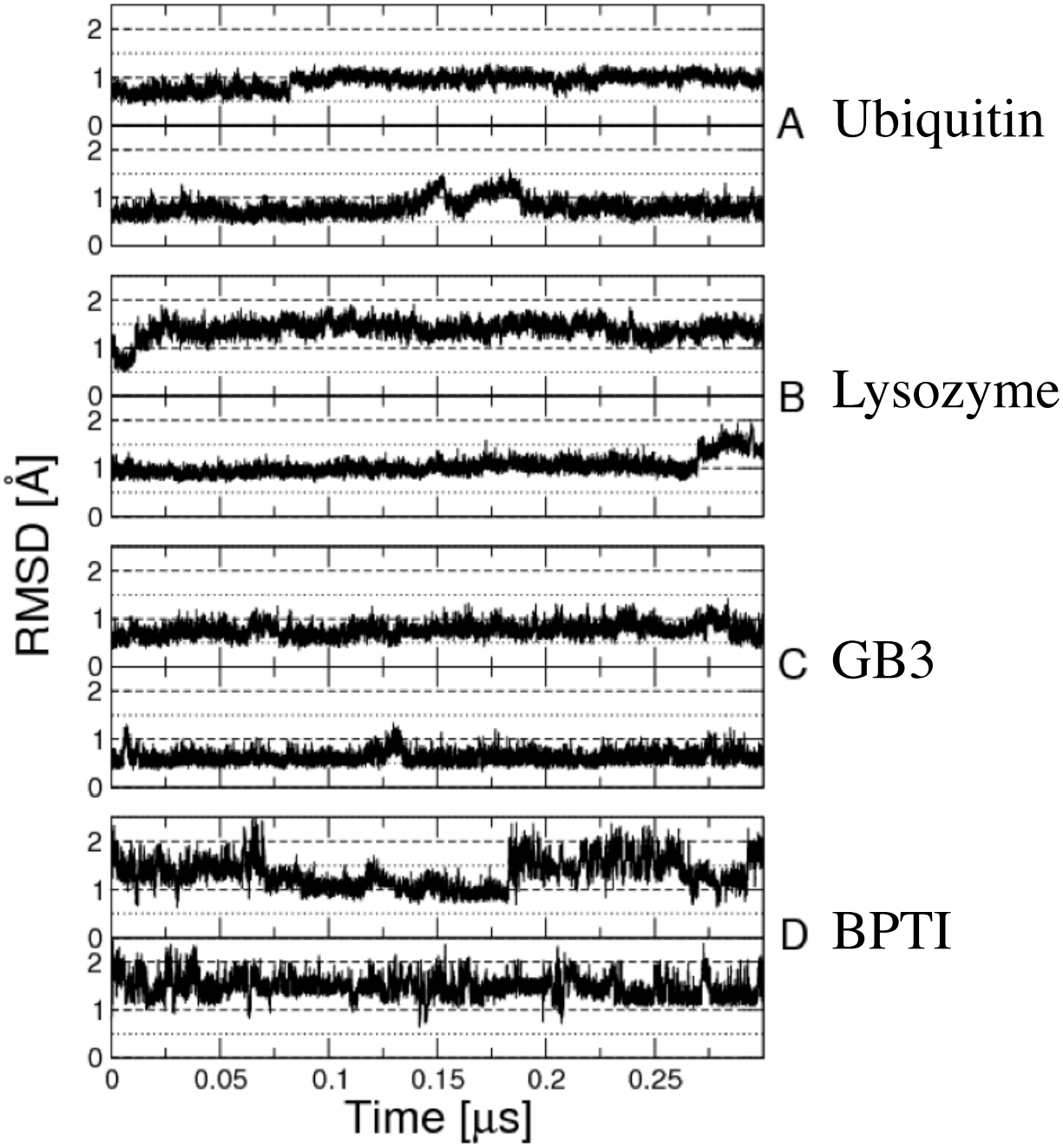


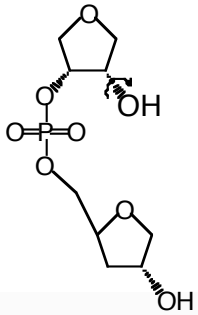


GB1m3 Hairpin

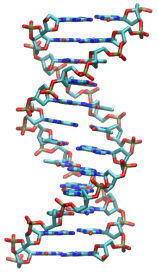
RMSD from experimental structures for folded proteins

Upper: C22/CMAP
Lower: C36





CHARMM36 Nucleic Acid FF



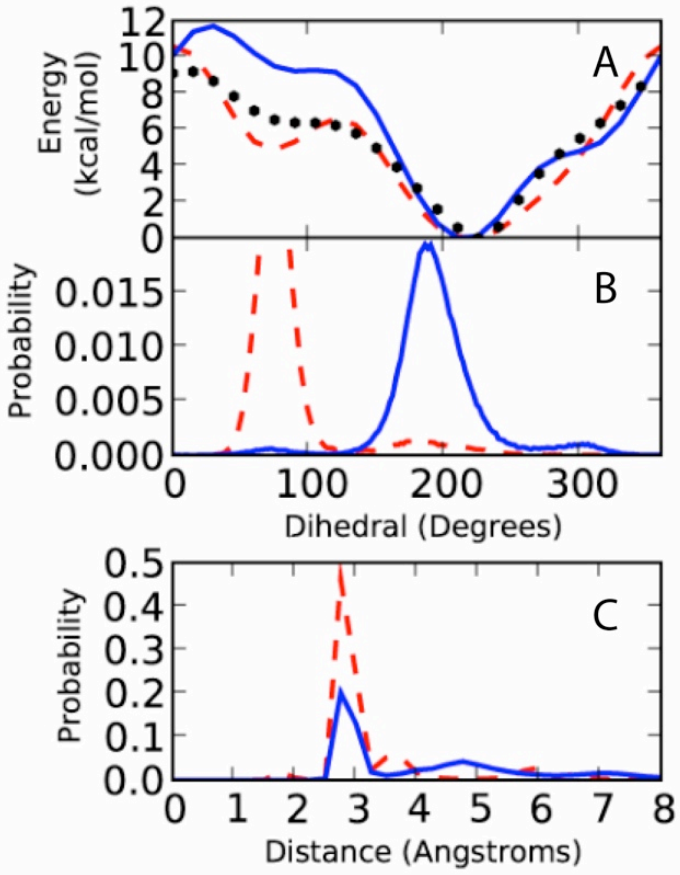


Table 1) Relative energy of B_{II} to B_I conformation for model compound E and percentage of B_{II} conformations from experimental and MD simulations of selected DNA duplexes.

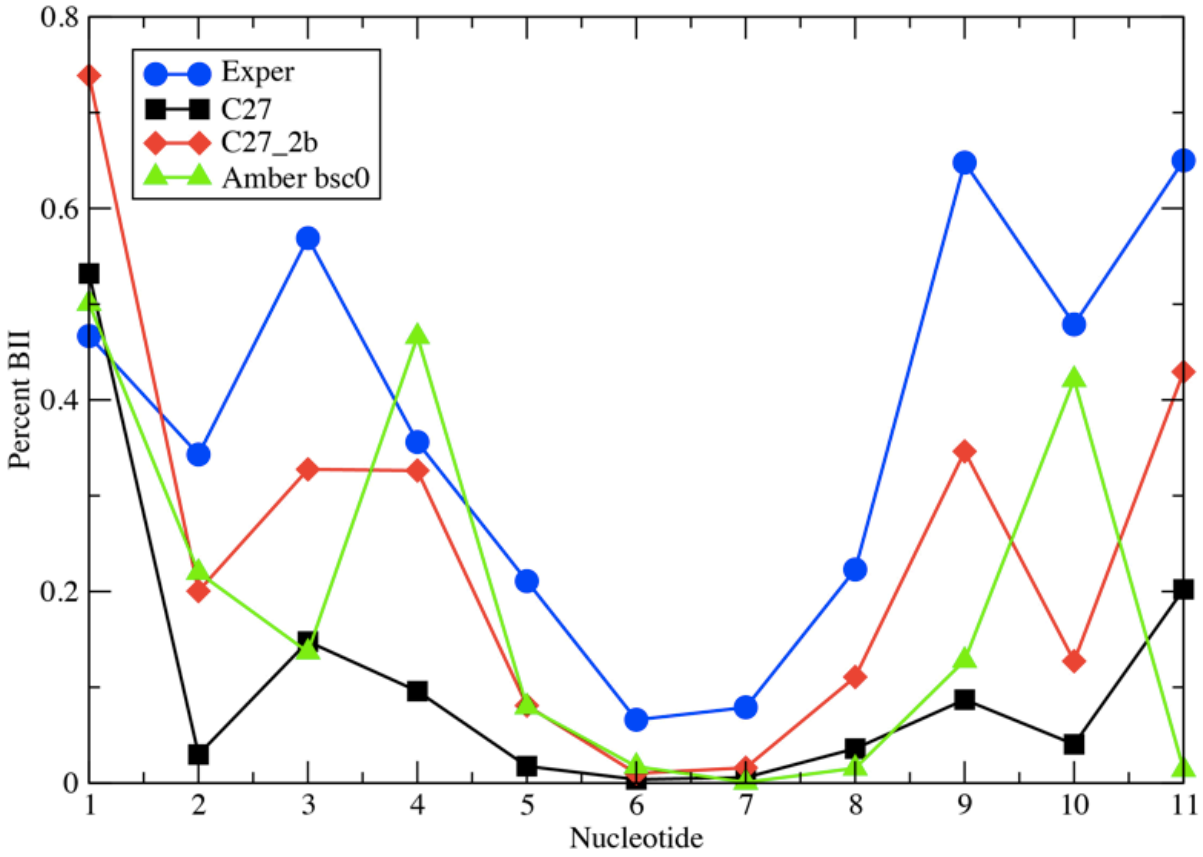
Cmpd E	QM	C27	C27r1	C27r2
ΔEnergy	2.16	2.78	1.91	1.59
%BII	Exp.	C27	C27r1	C27r2
bdj025	NA	17	32	40
EcoR1	37	10	24	28
1axp	NA	8	25	26
JunFos	30	7	24	26

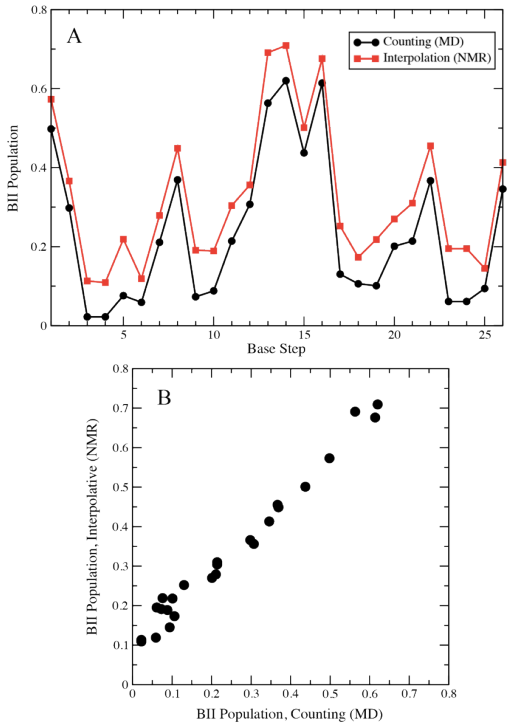
QM at the RIMP2/cc-pVTZ//MP2/6-31+G(d) level. Energies in kcal/mole. NA: not available.

QM, C27, C27r

Corrected for
RNA: base pair opening
DNA: underestimation of BII conformation

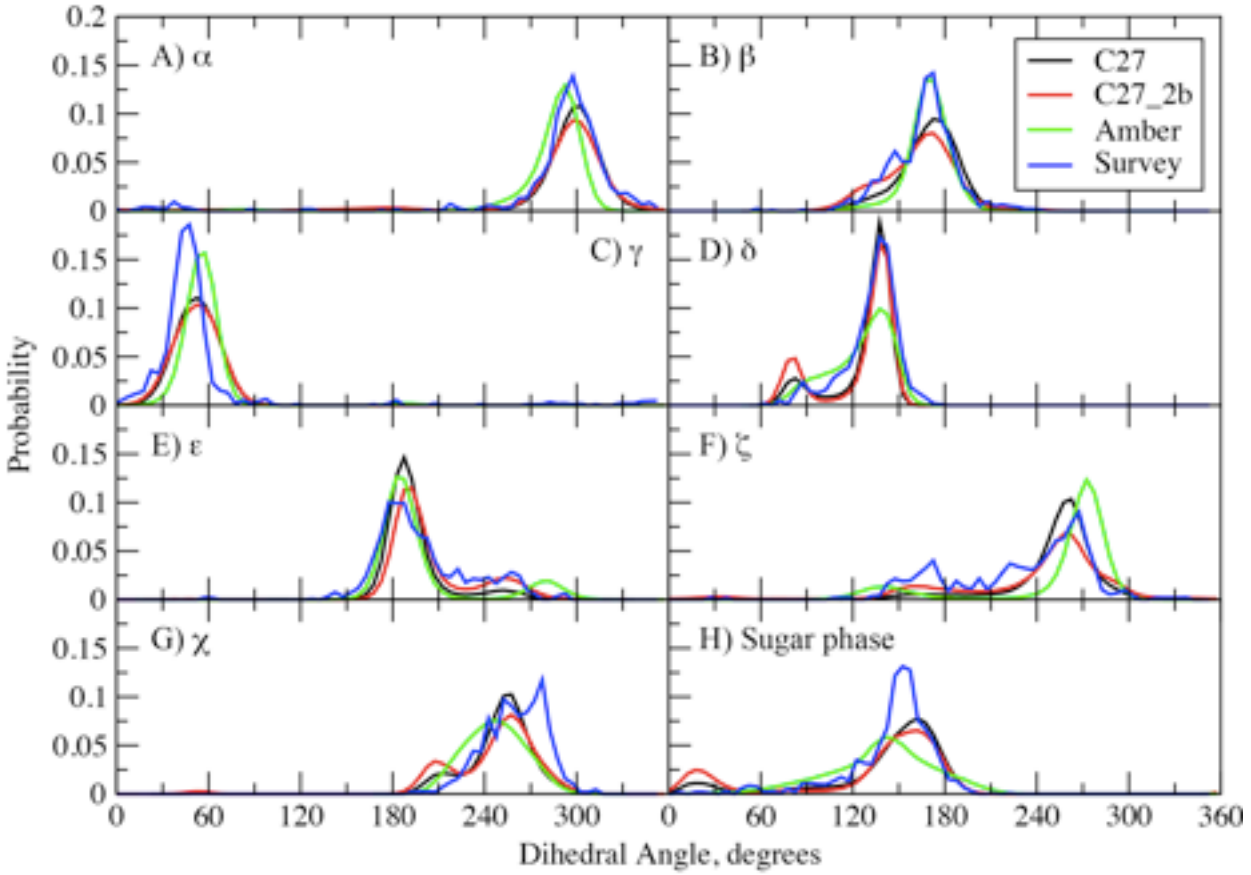
% BII conformation as a function of nucleotide for the EcoRI dodecamer



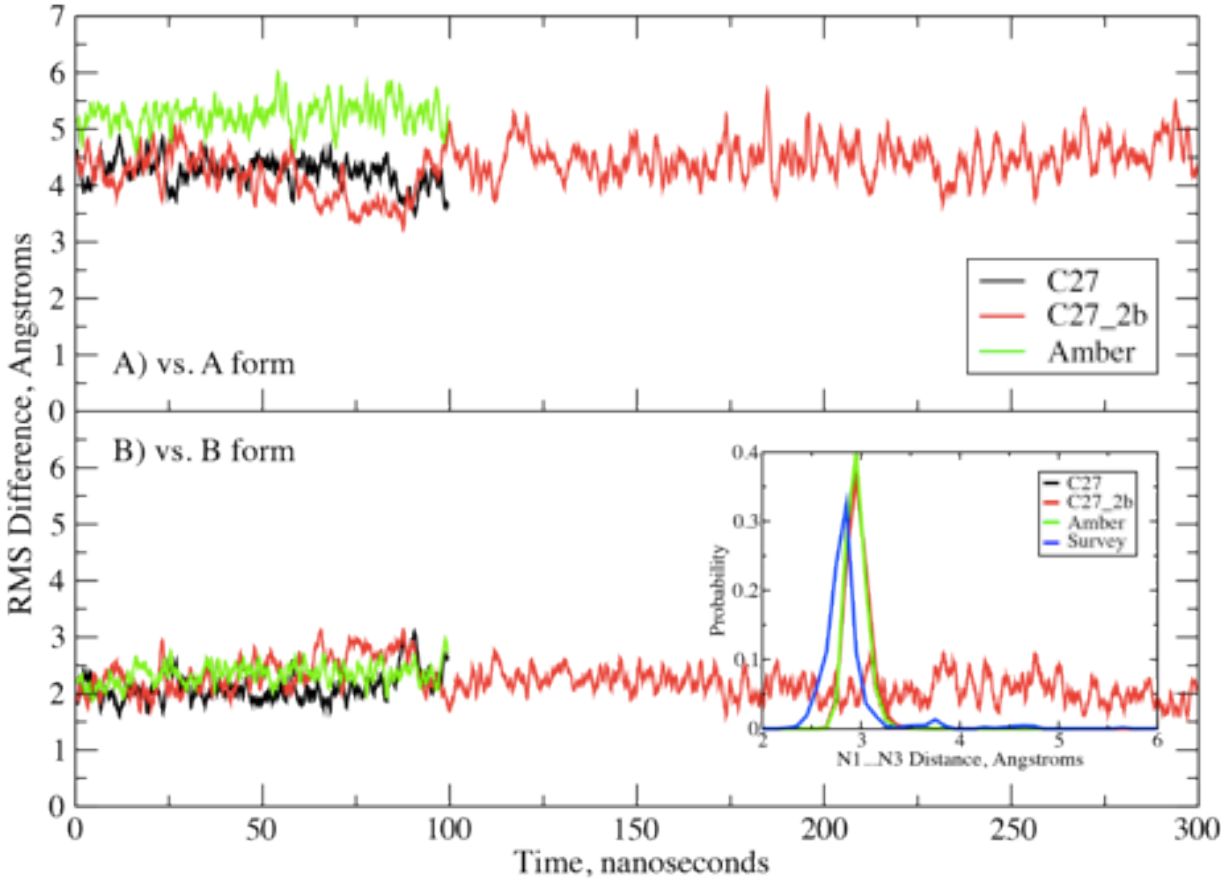


Lower C36 (C7_2b) value due to systematic overestimation of % BII by the experimental method do extract BII from NMR data

Dihedral angle and pseudorotation angle probability distributions for EcoRI using the C27, C27_2b and AMBER Parm99bsc0 (green) along with NDB survey of all B form structures with a resolution ≤ 2.5 Å.



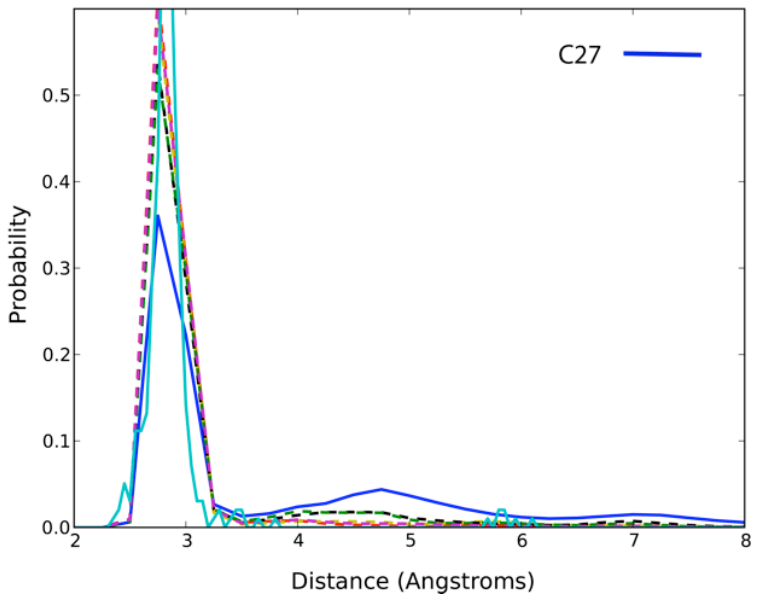
RMS Difference vs. time for the EcoR1 dodecamer.



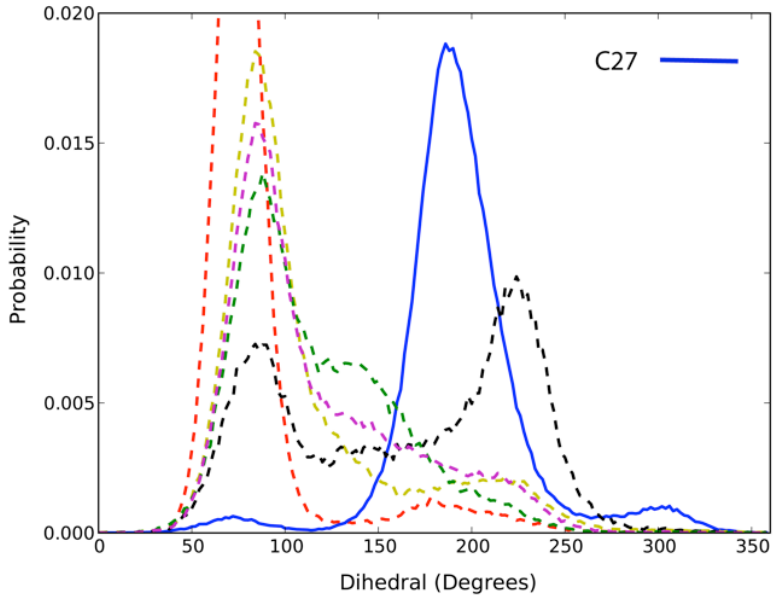
K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, A. D. MacKerell Jr., *Journal of Chemical Theory and Computation*, **8**:348–362 (2012)

CHARMM27 RNA FF: Overestimated WC base pair opening due to incorrect sampling of the 2’ OH orientation

N1-N3 WC distance



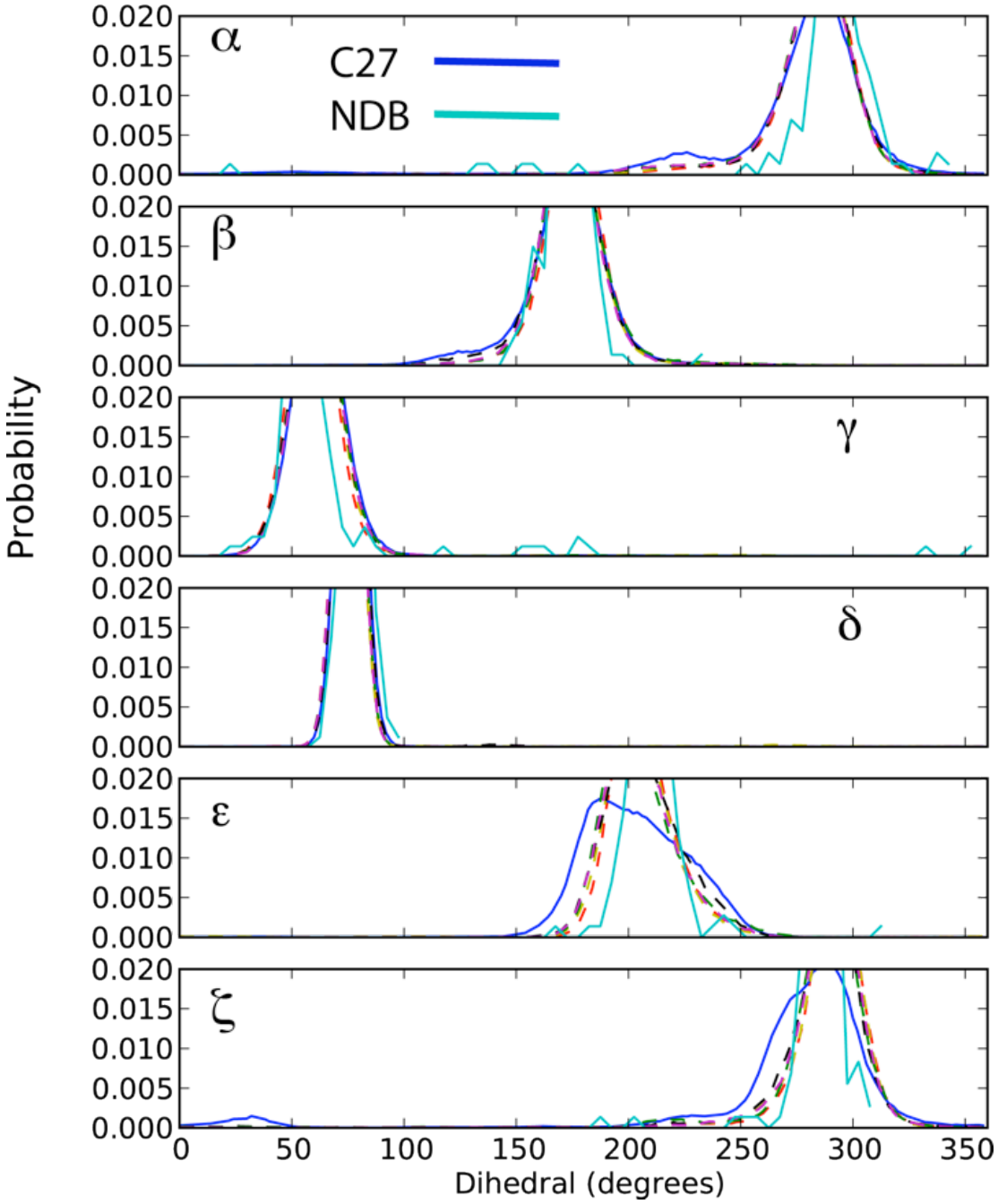
2’ OH orientation



base o3’ o4’

Cyan: survey of Nucleic acid DataBase (NDB)
Dashed: different FF interations

Probability distributions obtained for the phosphodiester backbone dihedral angles

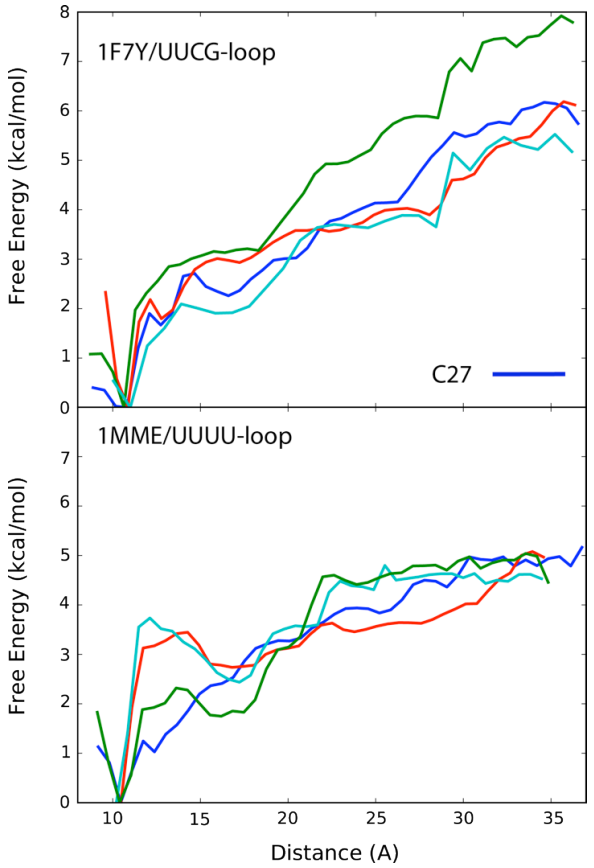


Folding free energy values for folding of RNA hairpins based on the PMF calculations

RNA	Exp.	C27	C36	Amber FF99
ΔG_{fold}: UUCG	-3.7±0.2	-3.3±0.3	-2.5±0.4	-13.5±2.0
ΔG_{fold}: UUUU	-1.5±0.1	-2.2±0.2	-2.0±0.2	-10.0±0.7
$\Delta\Delta G_{fold}$: UUCG-UUUU	-2.2	-1.1	-0.5	3.5

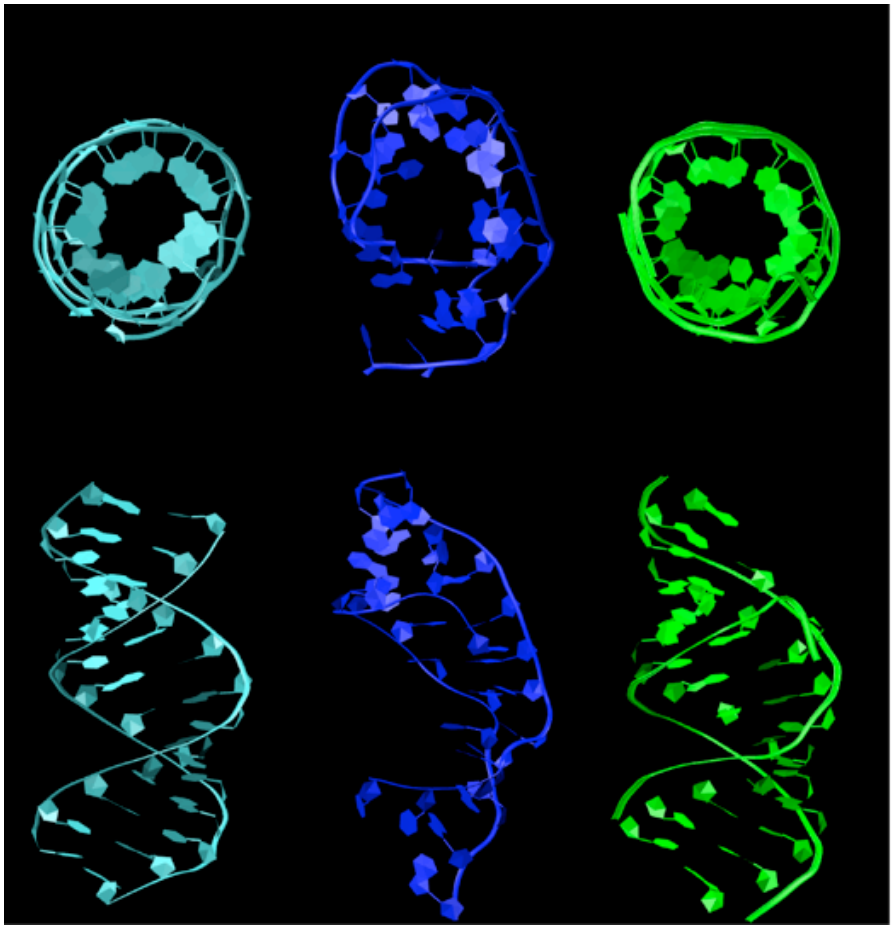
GCCUUCGGGC

CGCUUUUGCG

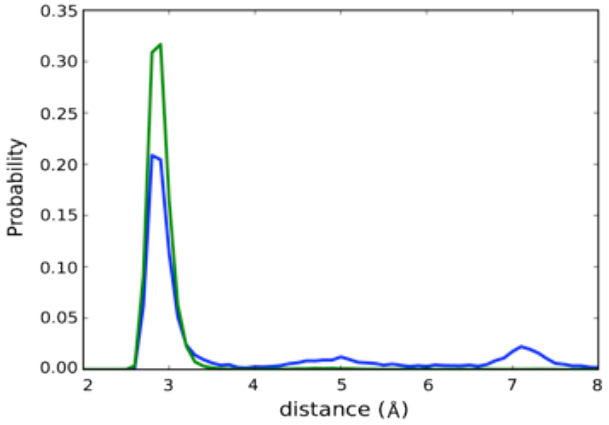


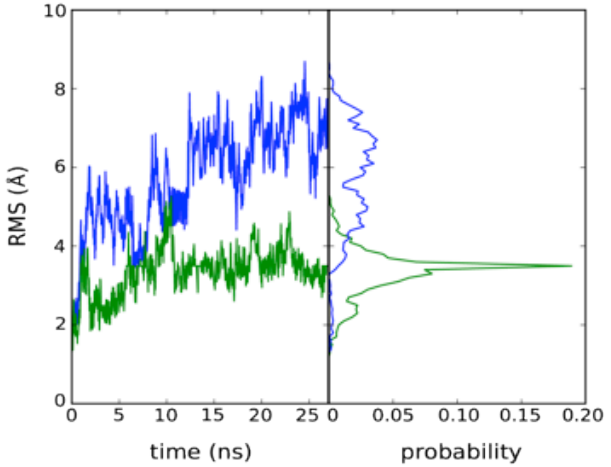
N. Deng and P. Cieplak, Biophys. J., **98**, 627-636 (2010)

AU base-pair RNA duplex structure (PDB: 1RNA)



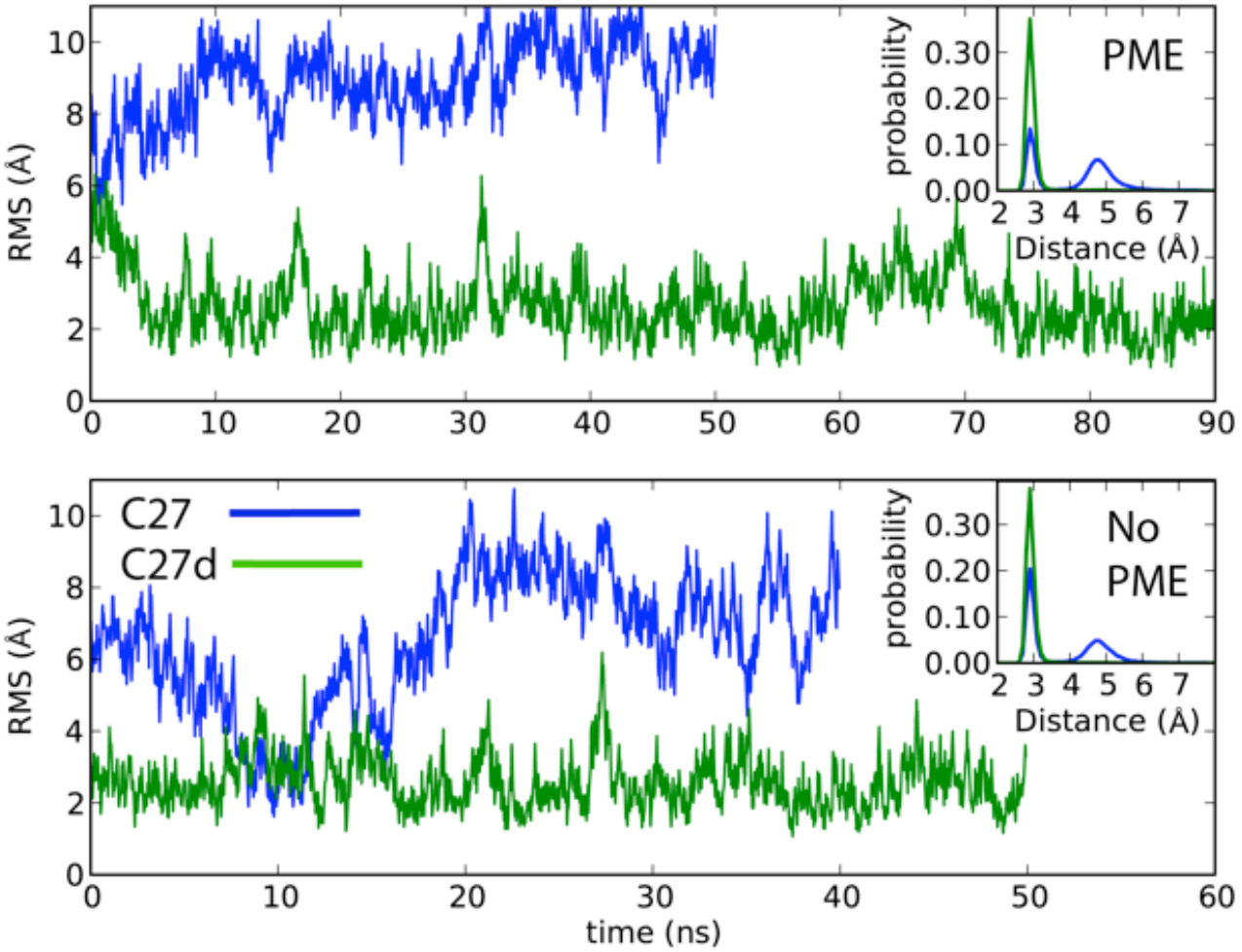
Xtal C27 C27d





18mer RNA duplex

GGCGCGCACCACGCGCGG



Limitation of additive force fields

The use of Coulomb's law with fixed atomic charges to treat the electrostatic interactions is a major simplification in current force fields. It is well known that the electron distribution of a molecule (and, thus, the atomic charges) changes as a function of the electrostatic field around the molecule. This is ignored in additive force fields. To compensate for this omission, the atomic charges are "enhanced" to mimic the polarization of molecules that occurs in a polar, condensed phase environment (e.g. aqueous solution, TIP3P water model dipole moment = 2.35 versus gas phase value of 1.85). This approximation has worked well in the current additive force fields; however, in many cases these models fail. To overcome this, next generation force fields are being developed that explicitly treat electronic polarization.

Methods to include electronic polarization in force fields

CHARMM
- Drude (MacKerell, Roux and coworkers)
- PIPF (Gao and coworkers, induced dipole)
- Cheq (Brooks, Patel and coworkers, fluctuating charge)

AMBER (induced dipole)
Friesner/Berne et al. (induced dipole/fluctuating charge)
TINKER/AMEOBA (induced dipole/multipole expansion)

All methods require that the perturbation of the electronic distribution due to the surrounding electrostatic field be optimized in an iterative fashion. This is due to the change in the "charge distribution" of a system leading to a new electrostatic field which then requires additional re-adjustment of the charge distribution (SCF: self-consistent field calculation). Matrix diagonalization may also be used, but is frequently inaccessible due to the large number of atoms in biological systems. In the end the need to perform an SCF calculation leads to a large increase in computational demands. Special methods to minimize this limitation in MD simulations have been developed (see below).

Fluctuating Charge Model (CHEQ)

Polarization is based on the movement of charge, q, between bonded atoms i and j in response to the surrounding electrostatic field. The extent of charge movement is based on the relative electronegativity, χ, and hardness, J, of the bonded atoms. The electrostatic energy is then obtained from the Coulombic interactions between the relaxed charges.

$$V(q_{ij}) = \chi_{ij} q_{ij} + \frac{1}{2} J_{ij} q_{ij}^2$$

$$\chi_{ij} = \chi_i^{'} + \chi_j^{'} \qquad\qquad J_{ij} = J_i^{'} + J_j^{'} + 2J_{ij}^{'}$$

Electronegativity: attraction of an atom for electrons
Hardness: work needed to transfer charge (resistance to charge movement)

Induced Dipole Model

Each atom, i, carries a charge, q_i, and a dipole moment, μ_i, such that electrostatic interactions between atoms i and j include:

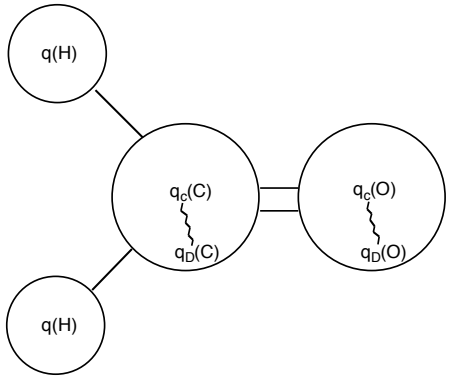
charge-charge interactions: $1/r_{ij}$
charge-dipole interactions: $1/r_{ij}^2$
dipole-dipole interactions: $1/r_{ij}^3$

Polarization included via relaxation of dipole moments in the electrostatic field, E_i, where α_i is the polarizability of atom i

$$\mu_i = \alpha_i\left(E_i^0 + E_i^{induced}\right) = \alpha_i\left(E_i^0 + \sum_{i \neq j} T_{ij}\mu_j\right)$$

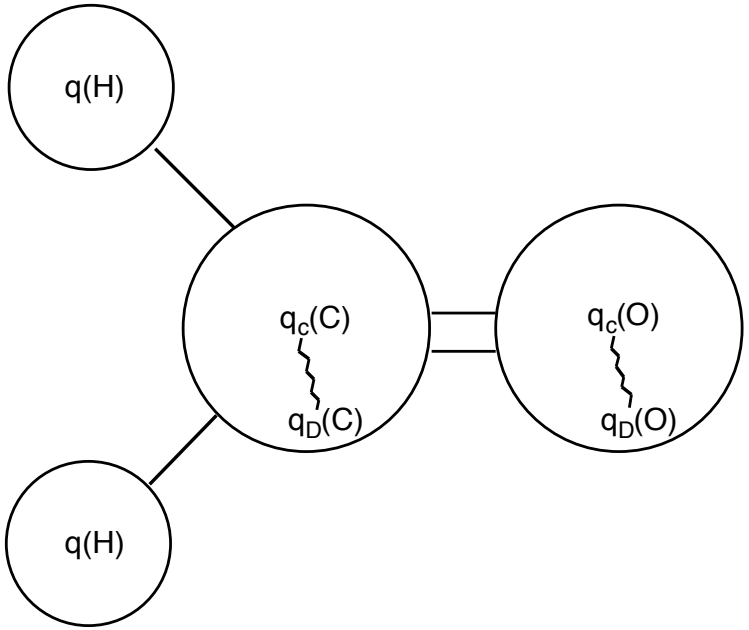
Amoeba: charges(monopoles), dipoles, quadrapoles and induced dipoles

Classical Drude Oscillator



To each atom, i, add a virtual particle (Drude) attached to the atomic core via a harmonic spring and place a charge, q_D, on the Drude. The Drudes then relax their positions with respect the surrounding electrostatic field with the relative positions of the Drudes with respect to their parent atom along with the respective charges of each yielding an induced dipole moment on each atom. The electrostatic energy is then obtained from the Coulombic interactions between the atomic and Drude charges.

Classical Drude oscillator



$$\alpha(A) = q_{\mathrm{D}}^2(A) \Big/ k_{\mathrm{D}}$$

$$q_D(A) = \sqrt{\alpha(A) \Big/ k_D}$$

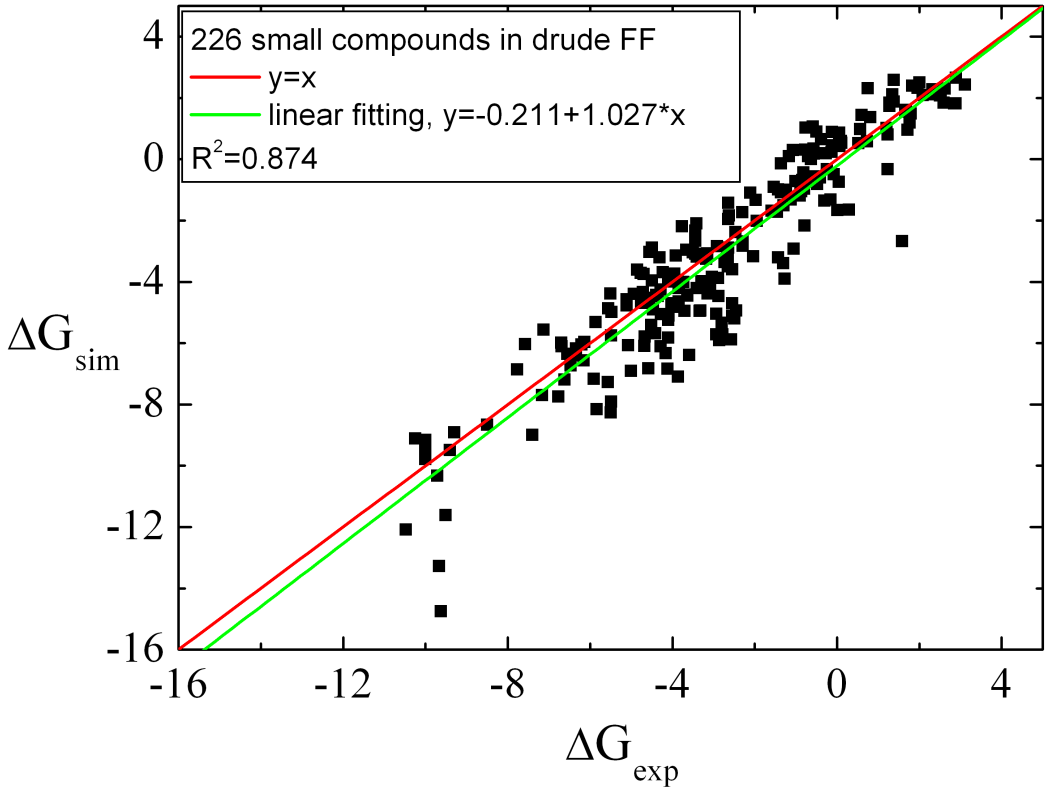
$$q_c(A) = q(A) - q_D(A)$$

$$U_{Drude} = \sum_{A<B}^{N,N_{\mathrm{D}}} \frac{q_{\mathrm{D}}(A)\cdot q_c(B)}{\left|\mathbf{r}_{\mathrm{D}}(A) - \mathbf{r}(B)\right|} + \sum_{A<B}^{N_{\mathrm{D}}} \frac{q_{\mathrm{D}}(A)\cdot q_{\mathrm{D}}(B)}{\left|\mathbf{r}_{\mathrm{D}}(A) - \mathbf{r}_{\mathrm{D}}(B)\right|} + \frac{1}{2}\sum_{A}^{N_{\mathrm{D}}} k_{\mathrm{D}}\left|\mathbf{r}_{\mathrm{D}}(A) - \mathbf{r}(A)\right|^2$$

MD Simulations with polarizable force fields: Extended-Langrangian

SCF calculation of induced dipole moments are computationally to demanding for MD simulations. As an alternative the polarization is treated as a dynamic variable that is propagated during the MD trajectory. This is done such that the electronic degrees of freedom being propagated in the MD simulation stay close to the Born-Oppenheimer approximation (e.g. equivalent to the SCF result). For example, in the Drude model, the Drude particle is assigned part of the mass of the parent atom (e.g. 0.5 amu) and then the Drude is propagated as an atom at each step of the MD simulation with the relative momentum of the Drude with respect to the parent atom "cooled" to 0 K, thereby approaching the Born-Oppenheimer approximation.

Free Energies of aqueous following automated optimization of 226 molecules using the Drude model



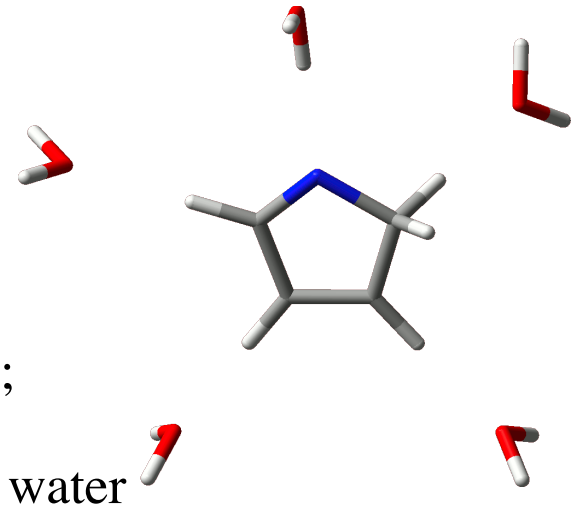
Part II

CHARMM General Force Field (CGenFF) parametrization tutorial

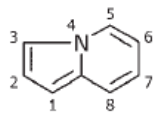
CGenFF: CHARMM General Force Field…

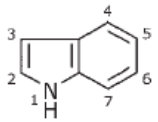
- "…for drug-like molecules"; mostly aimed towards Computer-Aided Drug Design (CADD)
- Why? Studying drug-target interactions with purely biomolecular force field problematic
  - Poor support for general organic molecules
  - Represent the drug with a different force field? NO! Different force fields are generally incompatible.
- Secondary role: coenzymes, prosthetic groups, metabolites,..

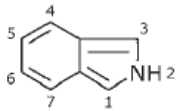
Points of focus

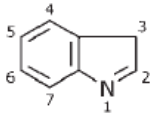


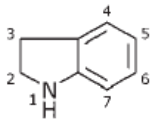
- Nonbonded parameters
  - LJ parameters: experimental density and ΔH_{vap}; assume transferability
  - Charges: scaled HF/6-31G(d) interactions with water
- Bonded parameters: MP2/6-31G(d) geometry, vibrational analysis and Potential Energy Surface (PES)
- Broad chemical space
  - Heterocyclic scaffolds
    - often central position in a drug's structure
    - "hinging effect" ⇒ geometry & conformational energetics
  - Functional groups
    - linkers, specific interactions with target
    - the more, the merrier

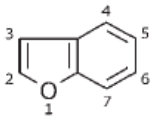


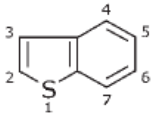


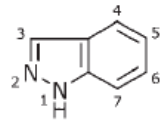


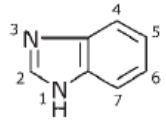


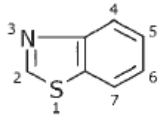


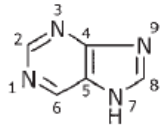


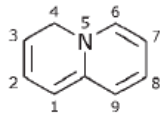


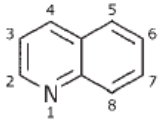


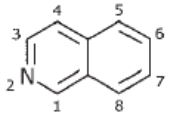


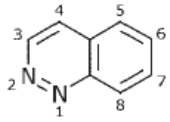


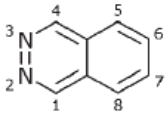


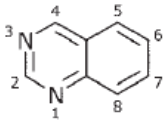


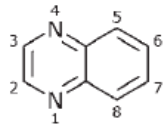


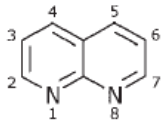


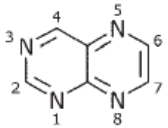


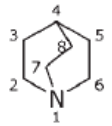


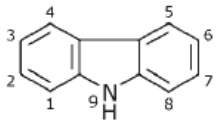


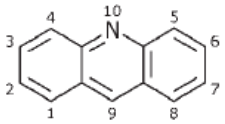


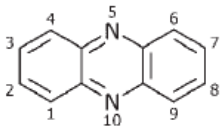


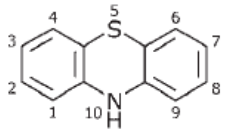


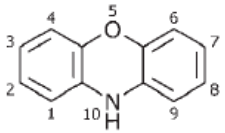


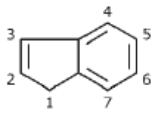


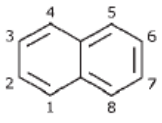


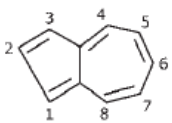


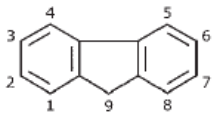


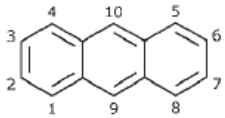


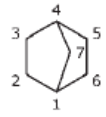


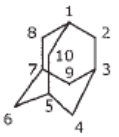












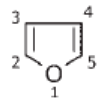
Numbering of Multiple Ring Systems

MAYBRIDGE

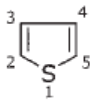
Maybridge, Trevillett, Tintagel, Cornwall PL34 0HW, England

Telephone +44 (0)1840 770453 Fax +44 (0)1840 770111

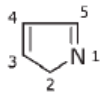
Email enquiries@maybridge.com Web www.maybridge.com



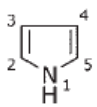
FURAN



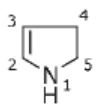
THIOPHENE



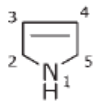
2H-PYRROLE



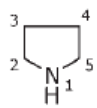
PYRROLE



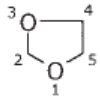
2-PYRROLINE



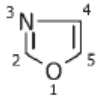
3-PYRROLINE



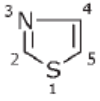
PYRROLIDINE



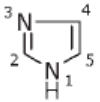
1,3-DIOXOLANE



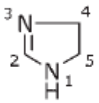
OXAZOLE



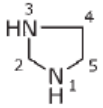
THIAZOLE



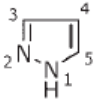
IMIDAZOLE



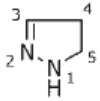
2-IMIDAZOLINE



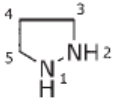
IMIDAZOLIDINE



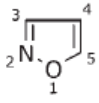
PYRAZOLE



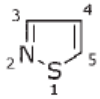
2-PYRAZOLINE



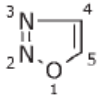
PYRAZOLIDINE



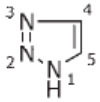
ISOXAZOLE



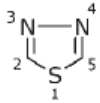
ISOTHIAZOLE



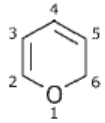
1,2,3-OXADIAZOLE



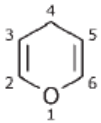
1,2,3-TRIAZOLE



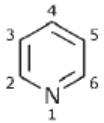
1,3,4-THIADIAZOLE



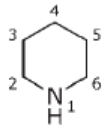
2H-PYRAN



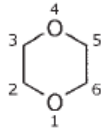
4H-PYRAN



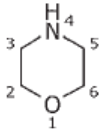
PYRIDINE



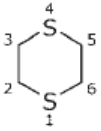
PIPERIDINE



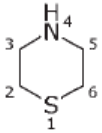
1,4-DIOXANE



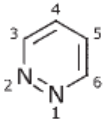
MORPHOLINE



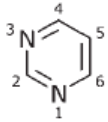
1,4-DITHIANE



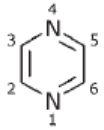
THIOMORPHOLINE



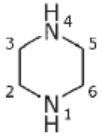
PYRIDAZINE



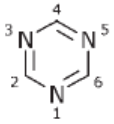
PYRIMIDINE



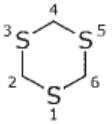
PYRAZINE



PIPERAZINE



1,3,5-TRIAZINE



1,3,5-TRITHIANE

Numbering of Multiple Ring Systems

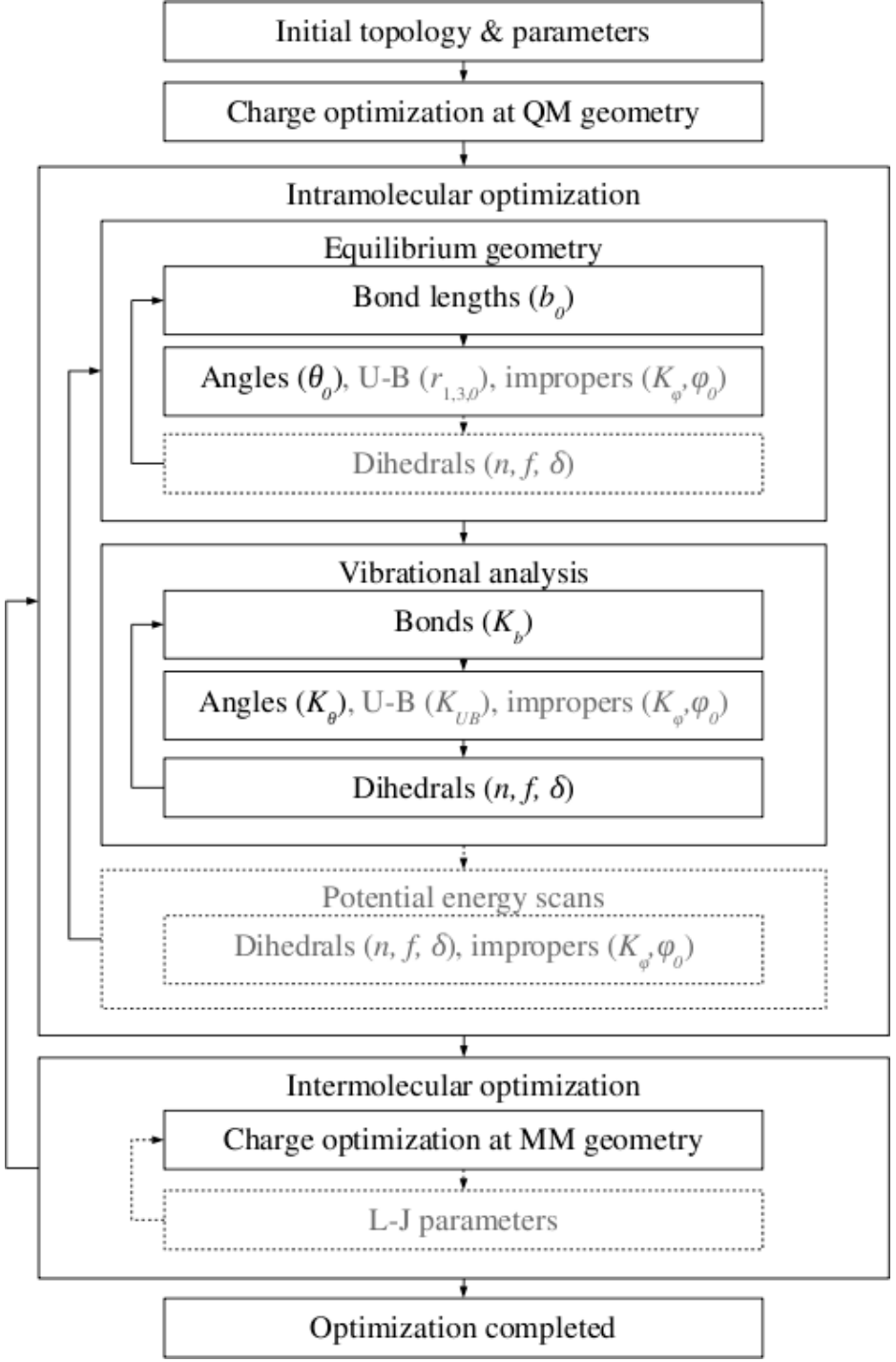
MAYBRIDGE

Maybridge, Trevillett, Tintagel, Cornwall PL34 0HW, England

Telephone +44 (0)1840 770453 Fax +44 (0)1840 770111

Email enquiries@maybridge.com Web www.maybridge.com

Parametrization flow diagram



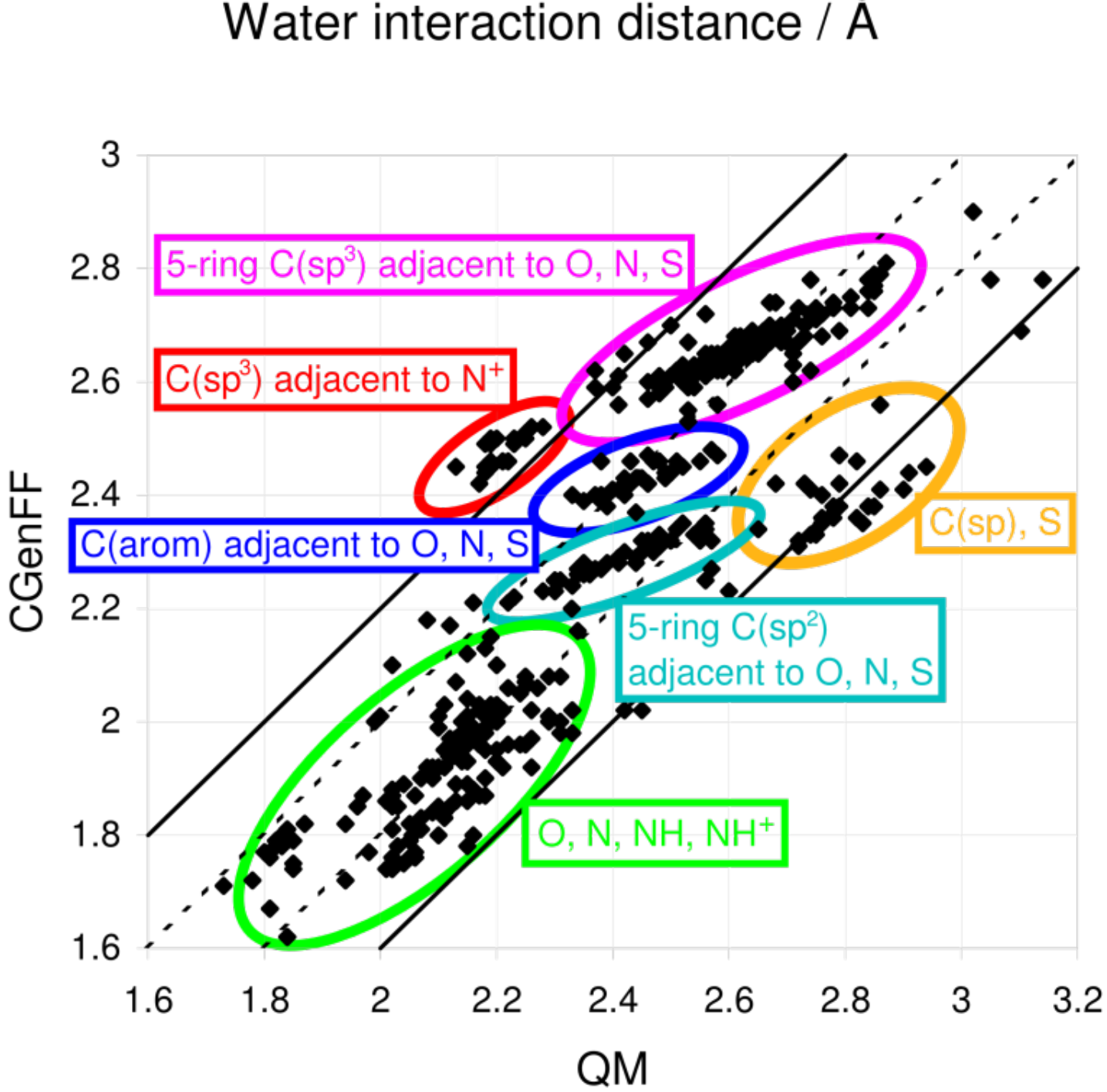
Statistics

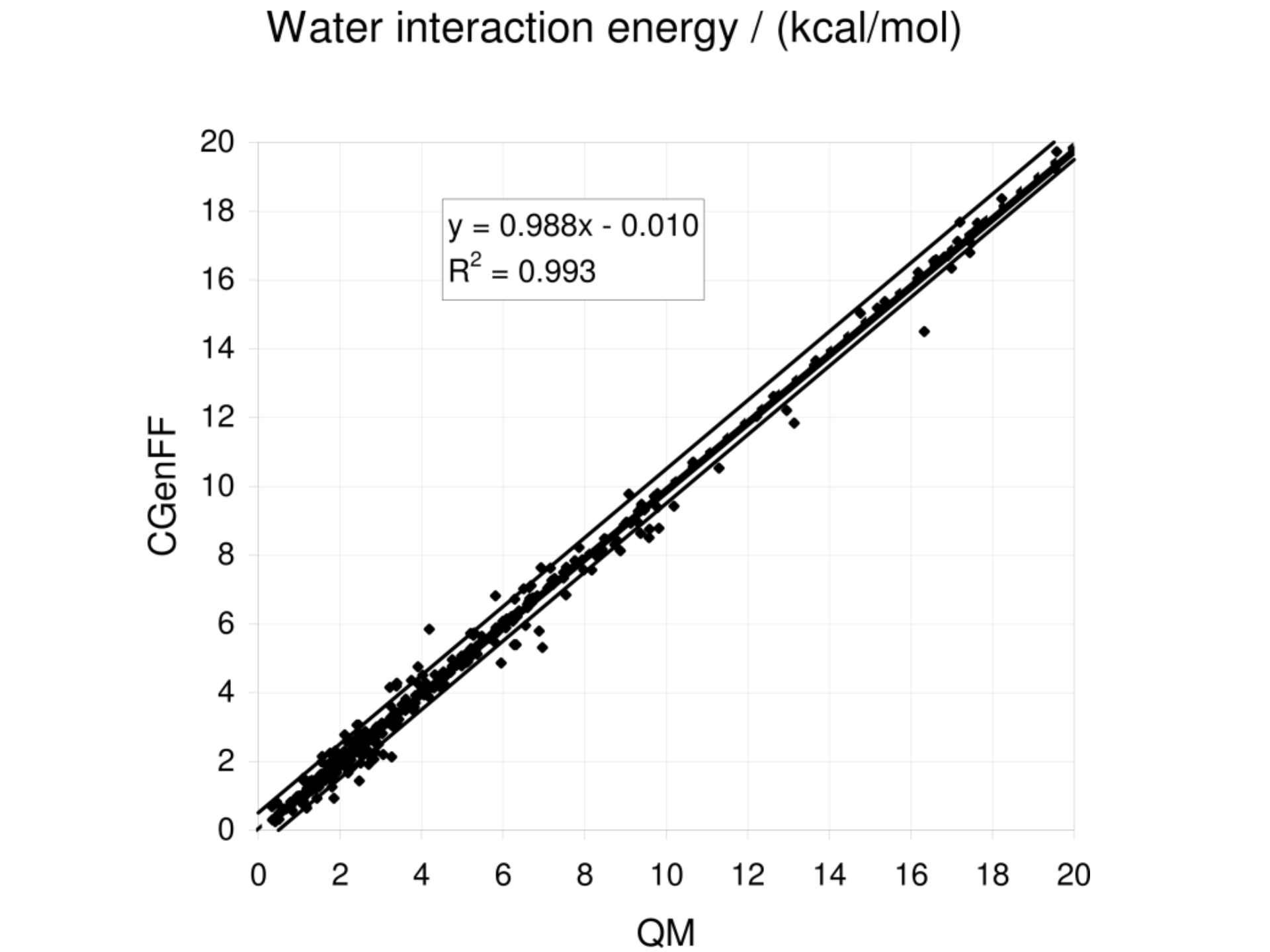
- 445 model compounds, 139 atom types, 5129 bonded parameters
  - Bonded parameters

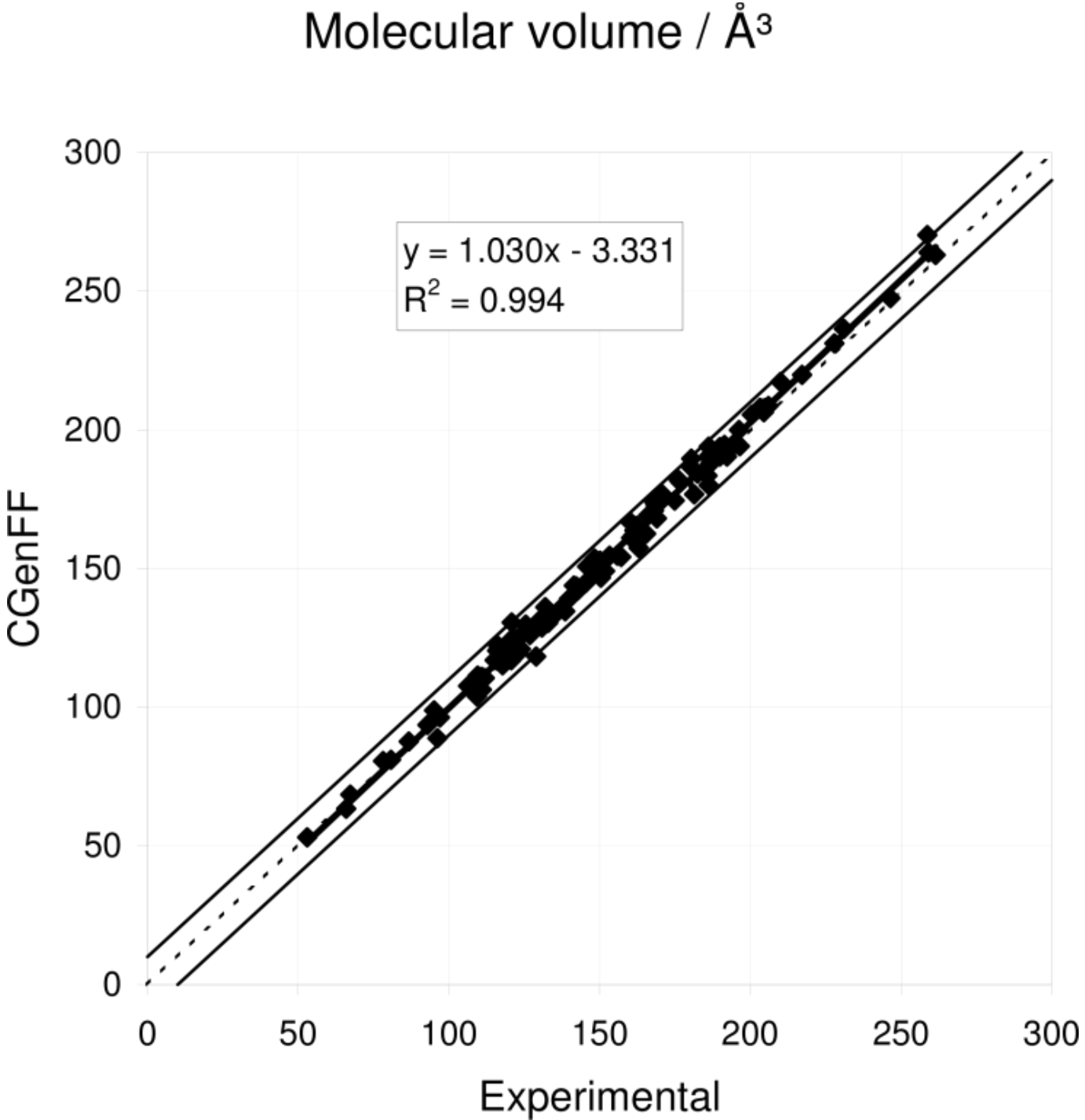
	data points	AD	RMSD	AAD
Bond lengths (A)	899	-0.0003	0.016	0.012
Valence angles (deg)	1420	-0.09	1.51	1.07
Dihedrals (deg)	137	-1.0	7.3	3.5
Vibrational frequencies	2619	3.6%	19.7%	6.4%

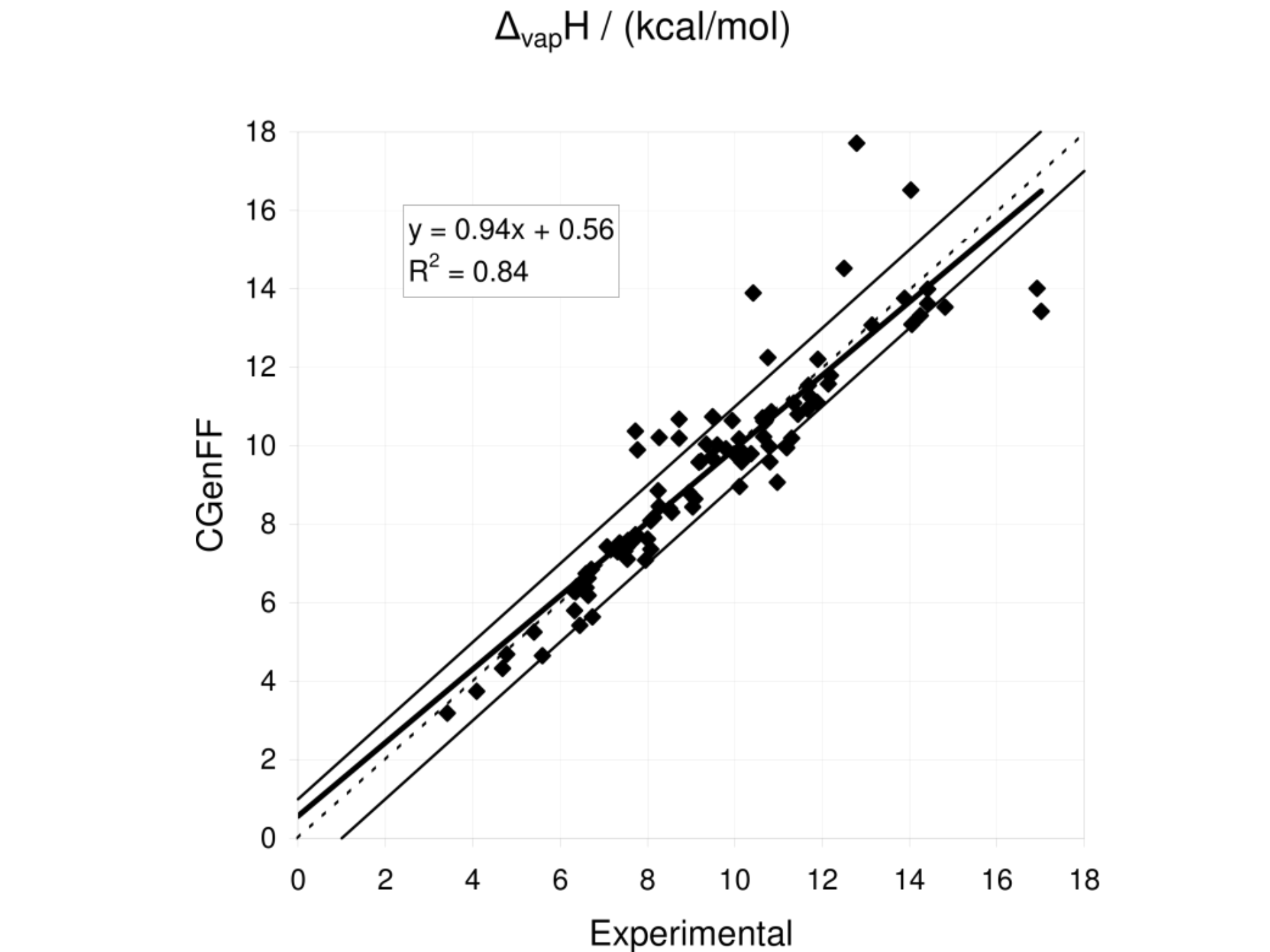
  - Nonbonded

		data points	AD	RMSD	AAD
‖μ‖	compared to MP2	78	30%	37%	32%
	compared to HF	65	27%	35%	30%
μ direction	compared to MP2	78	5.1°	8.5°	5.1°
	compared to HF	65	5.0°	8.1°	5.0°
Water interaction energy (kcal/mol)		437	0.07	0.34	0.20
Water interaction distance (Å)		437	0.11	0.20	0.16
Molecular volume (Å³)		111	0.6%	2.6%	2.1%
Heat of vaporization (kcal/mol)		95	-0.3%	10.6%	7.0%


Water interaction distance / Å
CGenFF
QM
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.6
1.8
2
2.2
2.4
2.6
2.8
3
3.2
5-ring C(sp3) adjacent to O, N, S
C(sp3) adjacent to N+
C(arom) adjacent to O, N, S
C(sp), S
5-ring C(sp2) adjacent to O, N, S
O, N, NH, NH+

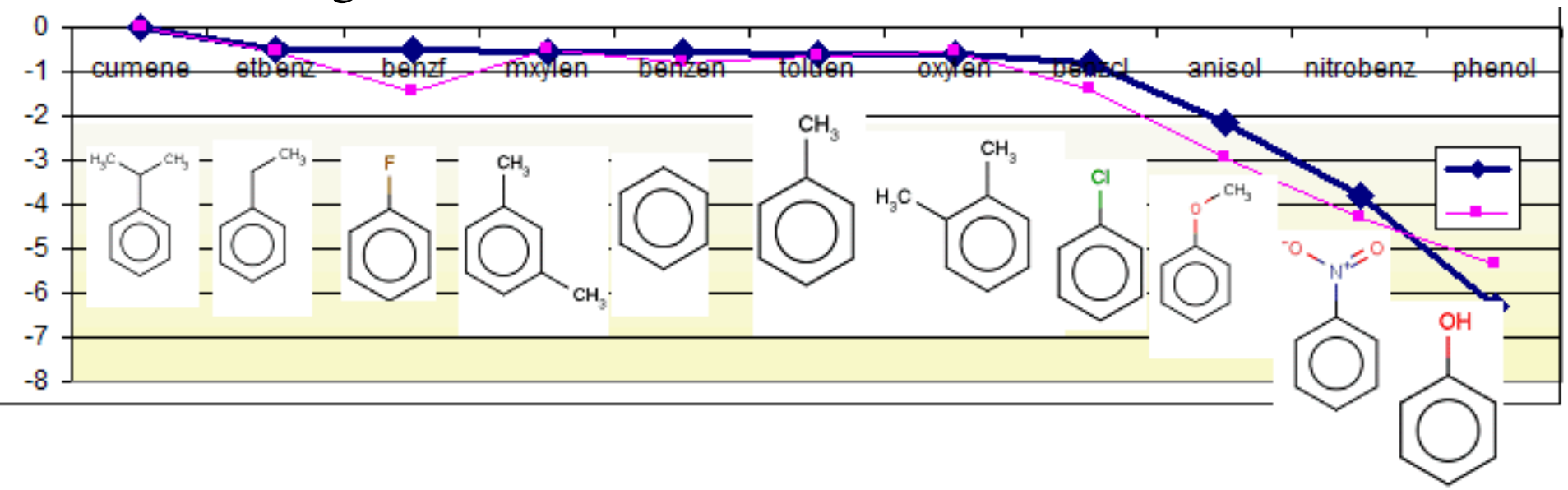

Water interaction energy / (kcal/mol)
y = 0.988x - 0.010
R2 = 0.993
CGenFF
QM
0
2
4
6
8
10
12
14
16
18
20


Molecular volume / Å3
y = 1.030x - 3.331
R2 = 0.994
CGenFF
Experimental
0
50
100
150
200
250
300


$\Delta_{vap}H$ / (kcal/mol)
y = 0.94x + 0.56
R^2 = 0.84
CGenFF
Experimental
0
2
4
6
8
10
12
14
16
18

Validation

Free energies of solvation:



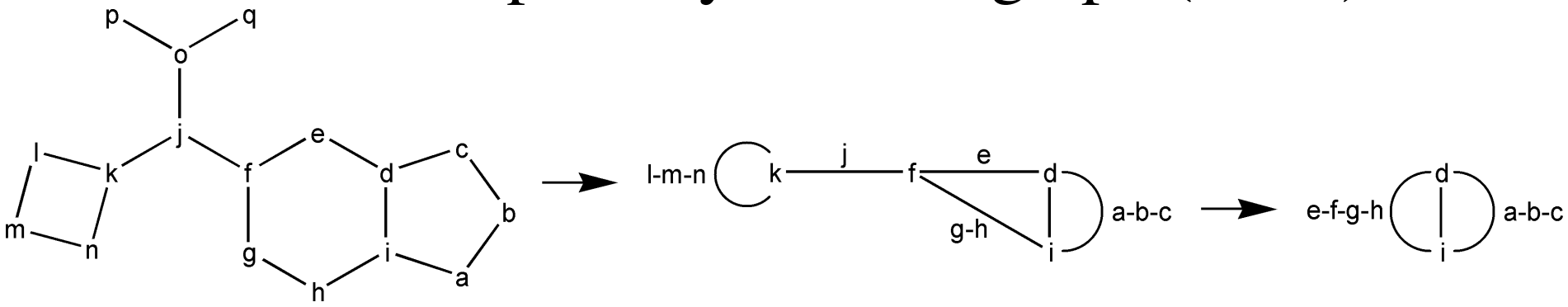
Willian C. Swope, IBM, Almaden Research Center
Daniel J. Price, GlaxoSmithKline, Research Triangle Park

CGenFF program

- Input: mol2 file
- Ring perception
- Resolution of of resonance
- Aromaticity perception
- Actual atom typing
- Assignment of parameters by analogy
- Assignment of charges
- Output: CHARMM toppar stream file

Ring perception: HRG

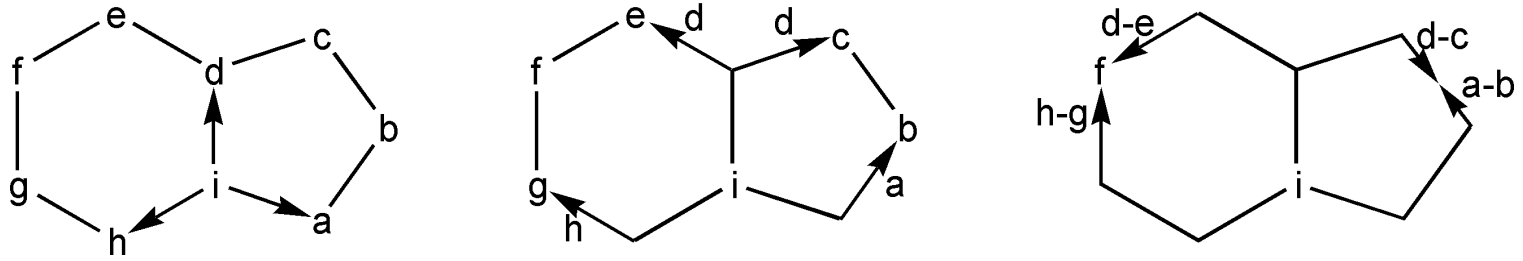
- Homeomorphically reduced graph (HRG):



- Pros: conceptually simple, fast
- Cons: does not actually identify rings. Generalizing it to do so is more complicated and/or may suffer combinatorial explosion on systems with many conjugated rings.

Ring perception: Figueras' algorithm

- Variation on breadth-first message-passing algorithm:



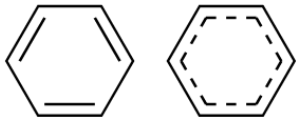
- Pros: fast, polynomial scaling
- Cons:
  - Presented as a Smallest Sets of Smallest Rings (SSSR) algorithm
  - may waste time iterating through straight chains

⇒ Ring perception in CGenFF atom typer:

  - Stage 1: reduction to HRG
  - Stage 2: variation on Figueras' algorithm that identifies the 3 smallest rings of which each atom is a member

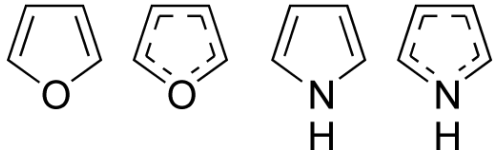
Problems with resonance

- bonds in input may or may not be marked as aromatic

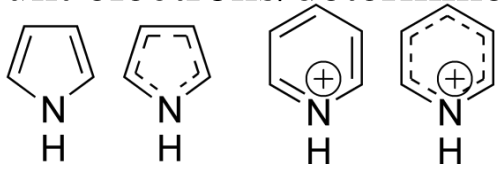


Problems:

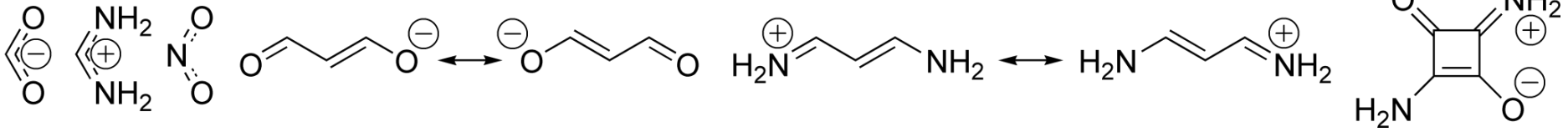
  - Mark 5-membered rings as aromatic?



  - If yes, how to reliably count electrons/determine charge?



  - Mark other cases of 50/50 resonance as aromatic or not?



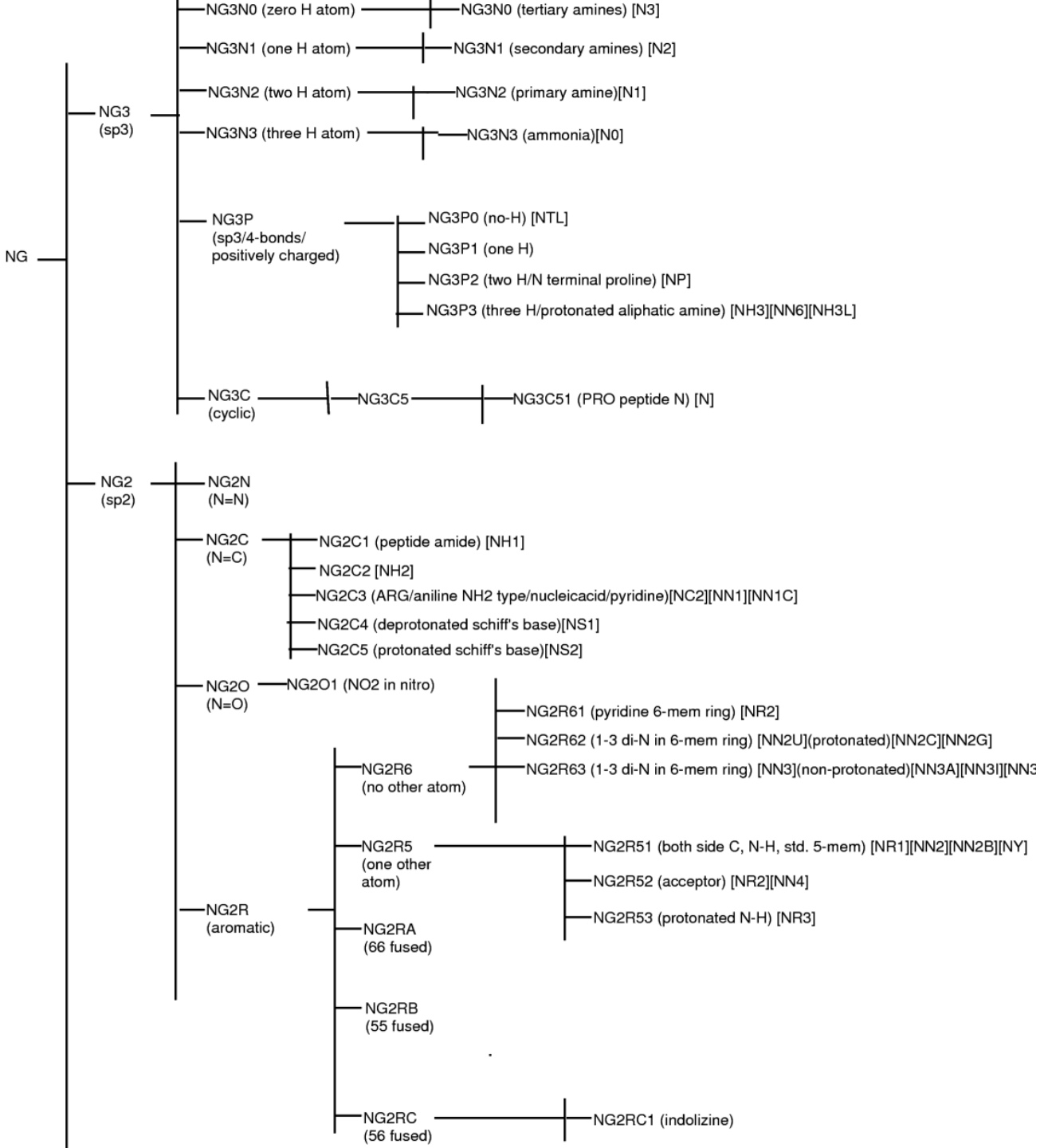
- Solution:
  - Accept all possibilities; resolve resonance and identify aromaticity
  - Perform actual atom typing on a single resonance structure

CGenFF aromaticity perception

- We mark a ring as aromatic if:
  - it's 5 or 6-membered
  - all carbons in the ring are sp^2
  - it has 6 in-ring π-electrons, where
    - an in-ring double bond counts for 2e
    - a heteroatom with a Lone Pair that is not part of another aromatic ring or a double bond counts for 2e
    - any atom that is part of another aromatic ring counts for 1e

$\Rightarrow$ iterative procedure

Heirarchical Atom typing



Atom typer: rule file

```
cat NG_
typ NG1T1  : nb 4 ne ( bo 3 el C ) err "nitrilium ion"
typ NG1T1  : ne ( bo 3 el C )
typ NG1T1  : ne ( bo 3 ) err "triple bonded non-nitrile nitrogen"
sub NG_2   : ne ( bo 2 )
sub NG_3P  : nb 4 charge 1
sub NG_N   : nb 3
typ NG301  : err "negatively charged single bonded nitrogen"
end

cat NG_N
#Amidinium and guanidinium
typ NG2R52 : ring 5 ne ( bo 1 el C ne ( bo 2 el N nb 4 ) )
typ NG2P1  : ne ( bo 1 el C ne ( bo 2 el N nb 4 ) )
#Single bonded neutral nitrogens that are sp2
typ NG2RC0 : arom 5 arom 6
typ NG2RC0 : ring2 5 arom 6 # warn "aromatic 6-ring fused to all-sp2 non-aromatic 5-ring"
typ NG2R51 : arom 5
typ NG2R51 : ring2 5
typ NG2R61 : arom 6
sub NG_AM  : ne ( bo 1 el C ne ( bo 2 elos ) )
typ NG2S3  : ne ( arom 6 ) ( el H ) ( el H ) impr
typ NG2S3  : ne ( el P ne ( bo 2 el O ) )
typ NG2S3  : ne ( el P ne ( bo 1 el O nb 1 ) )
#Single bonded neutral nitrogens that are sp3
typ NG3C51 : ring 5
typ NG3N1  : ne ( el N )
typ NG331  : ne ( el H ) ( el H ) ( el H )
typ NG321  : ne ( el H ) ( el H )
typ NG311  : ne ( el H )
typ NG301  :
end

cat NG_AM
#The next one is pretty specific because of exceptions like butyrolactam fused with pyrrole
typ NG2R53 : ! ( arom 5 ) ! ( arom 6 ) ! ( arom 7 ) ring23 5 ne ( bo 1 el C ring23 5 ne ( bo 2 elos ) )
typ NG2S2  : ne ( el H ) ( el H )
typ NG2S1  : ne ( el H )
typ NG2S0  :
end
```

Bonded parameters: algorithm

- Translate the atom typer rule file into a matrix containing the penalties of changing any atom into any other atom
- For a given molecule, generate a list of missing parameters
- In the case of a missing angle parameter A-B-C, the penalty for changing atom B is multiplied by a fixed number (eg. 10); same for the central to atoms in a dihedral parameter
- Perform substitutions of atom types that have the lowest possible penalty until we arrive at a combination of atom types that already exists

$\Rightarrow$ gradual degradation of quality; penalty score

Bonded parameters

- A parameter:

CG2O1 NG2S1 370.00 1.3450 ! Alanine Dipeptide, lk

- Rule file:

```
cat NG3
sub NG3P   : pri  0 alt NG3N 2 up 12
sub NG3N   : pri  5 alt NG3P 2 up 12
end

cat NG3P
typ NG3P2  : pri 0 alt NG3P1 1 alt NG3P3 3 alt NG3P0 4 up 8
typ NG3P3  : pri 1 alt NG3P2 1 alt NG3P1 2 alt NG3P0 4 up 8
typ NG3P1  : pri 3 alt NG3P2 1 alt NG3P0 3 alt NG3P3 4 up 8
typ NG3P0  : pri 4 alt NG3P1 1 alt NG3P2 2 alt NG3P3 4 up 8
end

cat NG3N
typ NG321  : pri 0   alt NG311 1   alt NG301 1.5 alt NG3N1 2.5 alt NG3C51 3   alt NG331 4   up 8
typ NG311  : pri 0.5 alt NG301 0.5 alt NG321 1   alt NG3N1 1.5 alt NG3C51 2   alt NG311 5   up 8
typ NG301  : pri 1   alt NG311 0.5 alt NG321 1.5 alt NG3N1 2   alt NG3C51 2.5 alt NG331 5.5 up 8
typ NG3N1  : pri 1.5 alt NG311 1.5 alt NG301 2   alt NG321 2.5 alt NG3C51 3.5 alt NG331 6.5 up 8
typ NG3C51 : pri 2.5 alt NG311 2   alt NG301 2.5 alt NG321 3   alt NG3C51 4   alt NG331 7   up 8
typ NG331  : pri 4   alt NG321 4   alt NG311 5   alt NG301 5.5 alt NG3N1  6.5 alt NG3C51 7  up 8
end
```

Charges

- Bond charge increments associated with bond parameters
- Example:

```
CG321 HGA2 309.00 1.1110 !<% 0.09 %>! PROT adm jr., 3/2/92
```

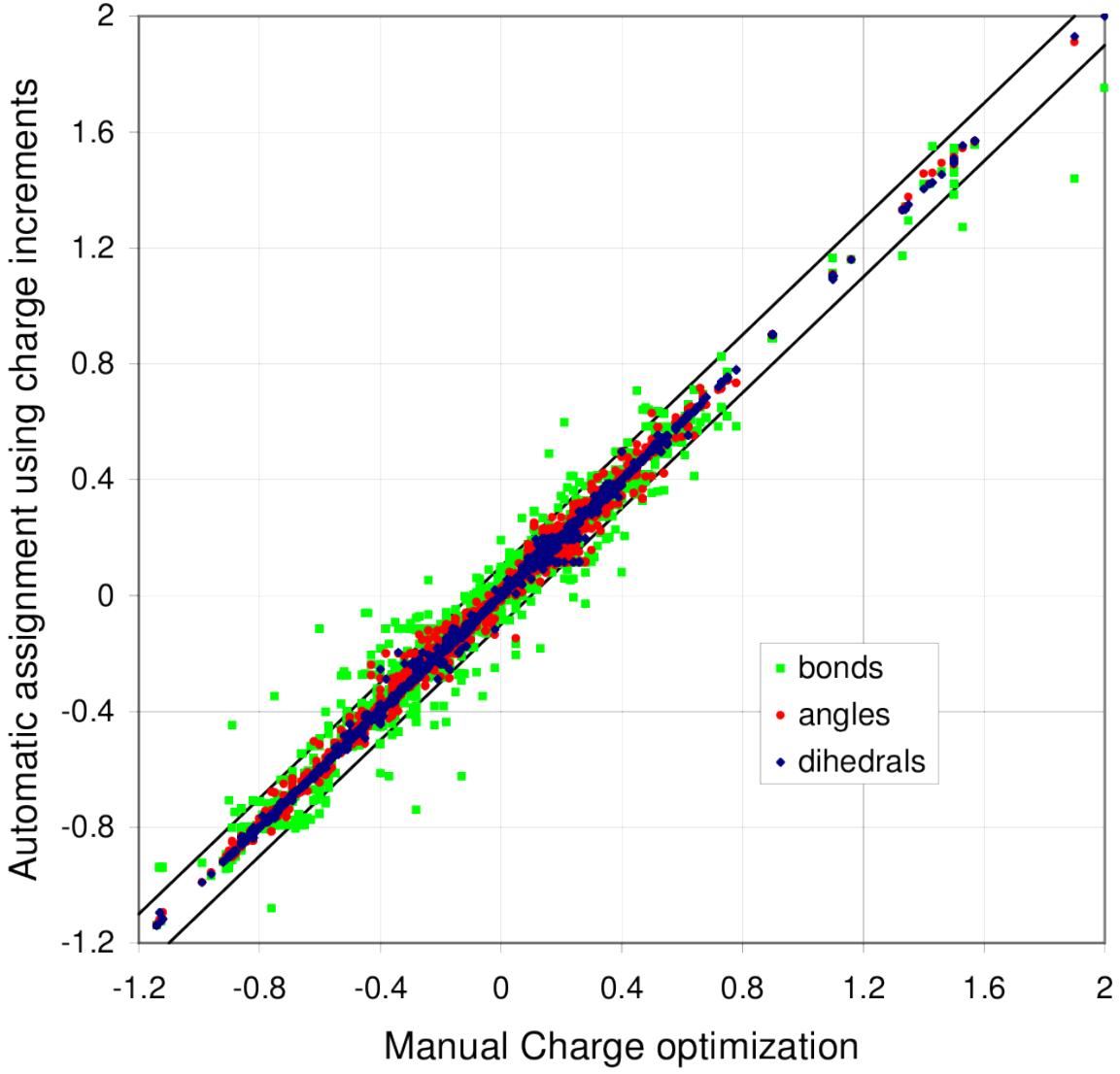
⇓

```
...
ATOM C1  CG321 -0.18
ATOM H11 HGA2   0.09
ATOM H12 HGA2   0.09
...
```

- 2 increments associated with angles and 3 with dihedrals ⇒ future possibility of capturing inductive and mesomeric effects
- Optimized to reproduce charges on existing compounds
- Novel molecules: increments “stick to” parameters by analogy

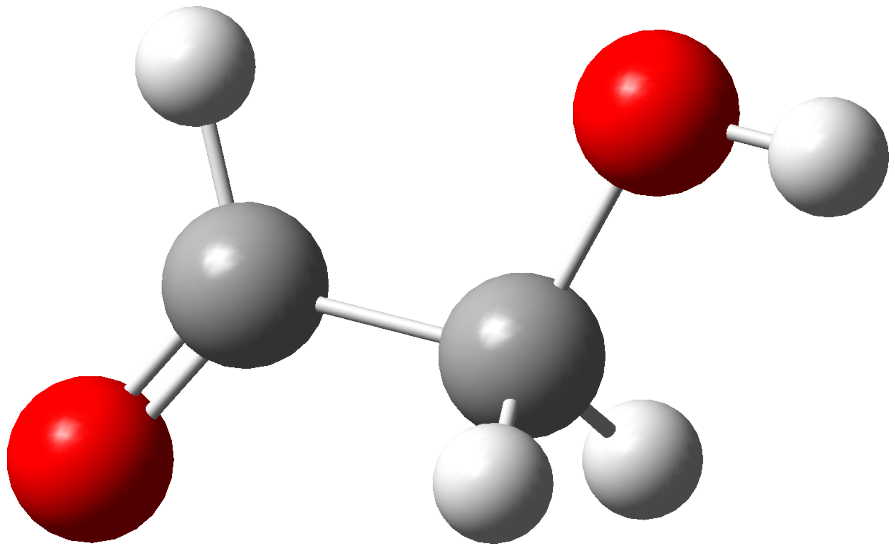
Advantage of higher-order increments

- Reproduction of 9681 manually optimized charges
- Bonds (1st order): RMSD = 0.0394 e
- Angles (2nd order): RMSD = 0.0174 e
- Dihedrals (3rd order): RMSD = 0.0081 e



Final result (1)

```
* Toppar stream file generated by
* CHARMM General Force Field (CGenFF) program version 0.9.1 beta
* For use with CGenFF version 2b6
*

read rtf card append
* Topologies
*

! "penalty" is the highest penalty score of the associated parameters.
! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.

RESI acyloin       0.000 ! param penalty=  42.900 ; charge penalty=   7.692
GROUP             ! CHARGE   CH_PENALTY
ATOM C1     CG2O4   0.197 !    5.743
ATOM O2     OG2D1  -0.395 !    4.755
ATOM H3     HGR52   0.090 !    2.145
ATOM C4     CG321   0.170 !    7.116
ATOM H5     HGA2    0.090 !    0.000
ATOM H6     HGA2    0.090 !    0.000
ATOM O7     OG311  -0.625 !    7.692
ATOM H8     HGP1    0.383 !    3.769

BOND C1   O2
BOND C1   H3
BOND C1   C4
BOND C4   H5
BOND C4   H6
BOND C4   O7
BOND O7   H8
IMPR C1    C4     O2     H3

END
```



Final result (2)

```
read param card flex append
* Parameters generated by analogy
*

ANGLES
CG2O4 CG321 OG311  112.00 122.50 ! acyloin, from CG2O5 CG311 OG311, penalty= 4.5

DIHEDRALS
OG2D1 CG2O4 CG321 OG311  0.0000 2   0.00 ! acyloin, from OG2D3 CG2O5 CG311 OG311, penalty= 14.5
HGR52 CG2O4 CG321 OG311  0.0000 3 180.00 ! acyloin, from HGR52 CG2O4 CG321 CG331, penalty= 42.9
CG2O4 CG321 OG311 HGP1   0.2200 1   0.00 ! acyloin, from CG2O5 CG311 OG311 HGP1, penalty= 4.5
CG2O4 CG321 OG311 HGP1   0.2300 2 180.00 ! acyloin, from CG2O5 CG311 OG311 HGP1, penalty= 4.5
CG2O4 CG321 OG311 HGP1   0.4200 3   0.00 ! acyloin, from CG2O5 CG311 OG311 HGP1, penalty= 4.5

END
```

Parametrization tutorial: Acetazolamide

/home/mackerell/tutorial.tgz

Log in / create account

WIKIPEDIA
The Free Encyclopedia

Navigation

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Article Talk Read Edit View history

Acetazolamide

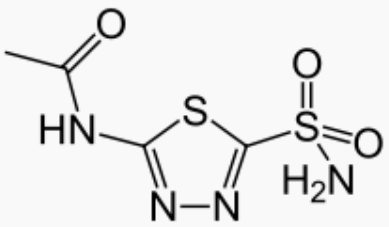
From Wikipedia, the free encyclopedia

Not to be confused with acetohexamide.

Acetazolamide, sold under the trade name **Diamox**, is a carbonic anhydrase inhibitor that is used to treat glaucoma, epileptic seizures, Idiopathic intracranial hypertension (pseudotumor cerebri), altitude sickness, cystinuria, and dural ectasia. Acetazolamide is available as a generic drug and is also a diuretic.

Contents

Acetazolamide



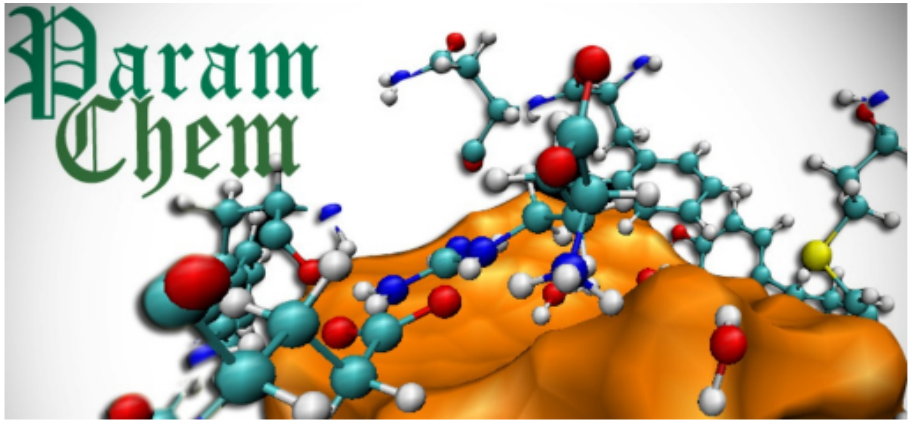
Systematic (IUPAC) name

ParamChem project outline https://www.paramchem.org/

- “Extensible Cyberenvironments for Empirical and Semiempirical Hamiltonian Parameter Optimization and Dissemination”
- Translation: “online parameterization engine and database for force fields and semi-empirical methods”
- Current functionality: CGenFF program (“initial guess”)
- Actual parametrization: under construction!

Submit to molecule to CGenFF program at paramchem.org

- Build mol2 in molecular editor
- Mind the bond orders and hydrogens!!!
  - Most commercial molecular editors (both QM and bio) get this right; free versions exist
  - Mind the quirks – look at the mol2 file (simple format)



Mission Statement
Preview
FAQs
User Account
Upload Molecule
Privacy Statement
Simulation, Force Field & Drug Design Links
Change Log

Output Data

filename: acazam.mol2

Output Information

click here to download/view the atomtypes, charges and parameters

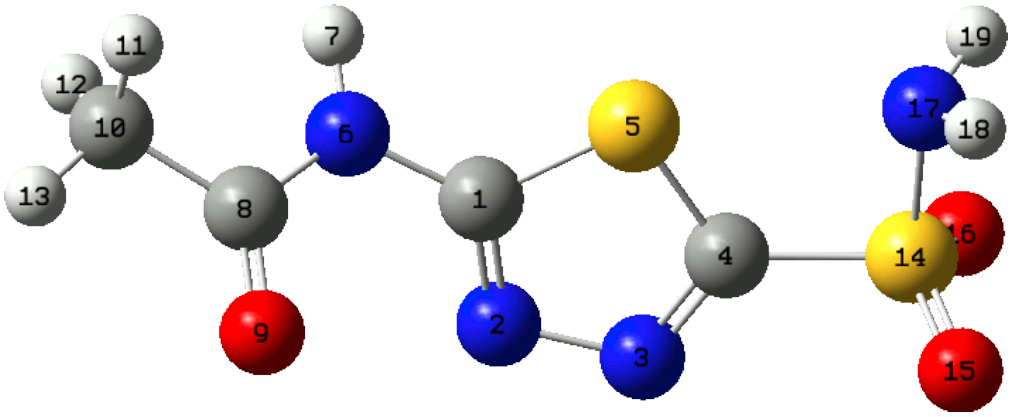
Upload File

click here to upload a new file

Summary of output data and its utilization

CGenFF output

```
RESI acazam        0.000 ! param penalty= 209.000 ; charge penalty= 121.856
GROUP              ! CHARGE   CH_PENALTY
ATOM C1     CG2R53  0.568 !  108.073
ATOM N2     NG2R50 -0.451 !   41.238
ATOM N3     NG2R50 -0.512 !   38.122
ATOM C4     CG2R53  0.801 !  118.976
ATOM S5     SG2R50 -0.096 !  121.856
ATOM N6     NG2S1  -0.424 !  109.486
ATOM H6     HGP1    0.322 !    4.985
ATOM C7     CG2O1   0.523 !    7.138
ATOM O7     OG2D1  -0.524 !    3.801
ATOM C8     CG331  -0.273 !    3.228
ATOM H81    HGA3    0.090 !    0.000
ATOM H82    HGA3    0.090 !    0.000
ATOM H83    HGA3    0.090 !    0.000
ATOM S9     SG3O2   0.640 !  102.754
ATOM O91    OG2P1  -0.419 !    3.561
ATOM O92    OG2P1  -0.419 !    3.561
ATOM N10    NG321  -0.766 !    5.281
ATOM H101   HGP1    0.380 !    0.425
ATOM H102   HGP1    0.380 !    0.425
```



...

```
ANGLES
NG2R50 CG2R53 NG2S1    65.00    127.80 ! acazam , from NG2R53 CG2R53 OG2D1, penalty= 57
NG2R50 CG2R53 SG3O2    70.00    109.00 ! acazam , from NG2R53 CG2R53 SG311, penalty= 40
NG2S1  CG2R53 SG2R50   30.00    118.00 ! acazam , from SG2R50 CG2R53 HGR52, penalty= 144
SG2R50 CG2R53 SG3O2    70.00    109.00 ! acazam , from NG2R53 CG2R53 SG311, penalty= 144
CG2R53 NG2R50 NG2R50  110.00    106.80 ! acazam , from CG2R51 NG2R50 NG2R50, penalty= 3
CG2O1  NG2S1  CG2R53   50.00    120.00 ! acazam , from CG2O1 NG2S1 CG2R61, penalty= 8.5
CG2R53 NG2S1  HGP1     34.00    117.00 ! acazam , from CG2R61 NG2S1 HGP1, penalty= 8.5
CG2R53 SG2R50 CG2R53  110.00     97.00 ! acazam , from CG2R51 SG2R50 CG2R53, penalty= 3
CG2R53 SG3O2  NG321    60.00     98.00 ! acazam , from CG2R61 SG3O2 NG321, penalty= 8.5
CG2R53 SG3O2  OG2P1    60.00    101.00 ! acazam , from CG2R61 SG3O2 OG2P1, penalty= 8.5
```

...

Strategy

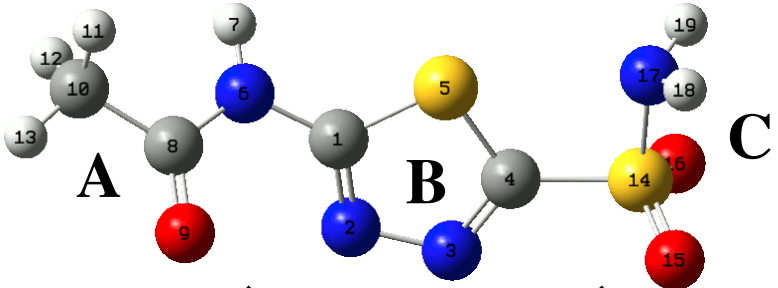
- All high penalties involve the central 1,3,4-thiadiazole ring and its linkages
- Common motif:
  - Individual chemical groups: often well-supported by CGenFF
  - Linkages between them: high penalties

  ⇒ **Divide-and-conquer**

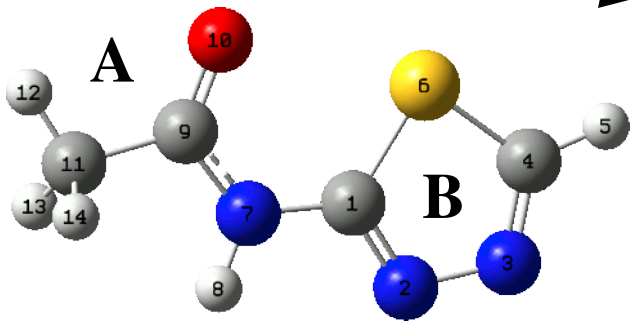
- Note in this case: central 1,3,4-thiadiazole ring poorly supported by CGenFF

- Typically only optimize new parameters not already in CGenFF

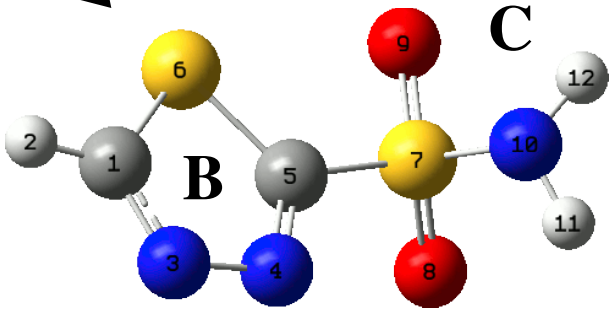
Divide-and-conquer (linkage-oriented!)

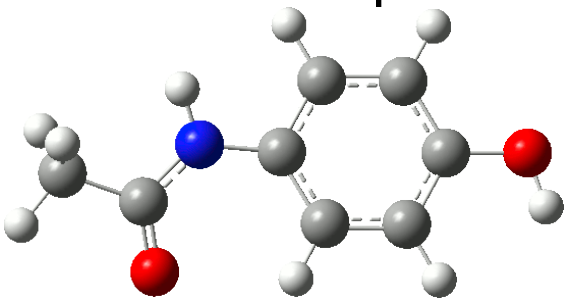


TDAZAM

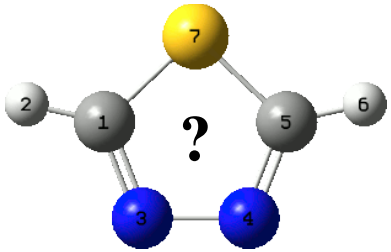


SULFAZ

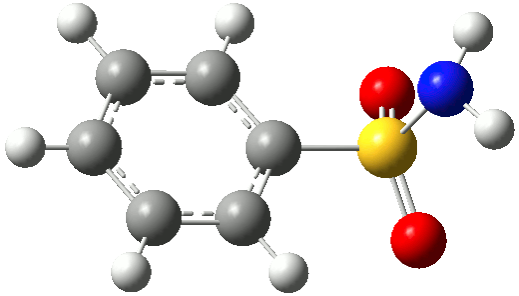




PACP



TDAZ



BSAM

Full parameter optimization

- Target data generation
  - MP2/6-31G* optimization and frequency calculation
  - HF/6-31G*† water interactions at MP2* geometry (only HF!!)
  - MP2/6-31G* Potential Energy Scans (PES)
    - Geometry/Frequency/PES may be at a higher level if accessible
- Optimization
  - Charges using water interactions at MP2 geometry
  - Hard degrees of freedom
    - Reference bonds and angles target MP2 geometry
    - Force constants (and amplitudes) target MP2 vibrational analysis
  - Soft degrees of freedom target MP2 PES
  - Revisit charges using water interactions at MM geometry
- Intrinsically iterative but above order often mitigates this

* MP2 calculations on anions require diffuse functions on non-hydrogen atoms (MP2/6-31+G*)

† – HF interaction energy scaled by 1.16 for non-ions

– Unscaled MP2 instead of HF for water interactions for 3rd row and heavier

Parametrization tutorial (/home/mackerell/tutorial.tgz)

Collection of scripts/charmm inputs to perform various steps of the optimization process.

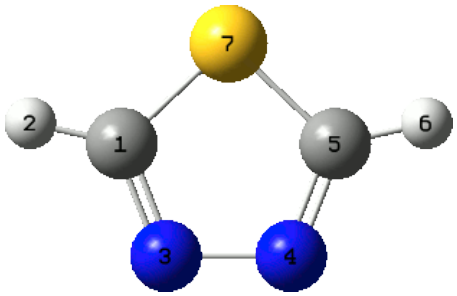
See tutorial.orig/README (may be streamed to automatically run all inputs)

Relevant inputs will be listed

Hard and soft degrees of freedom

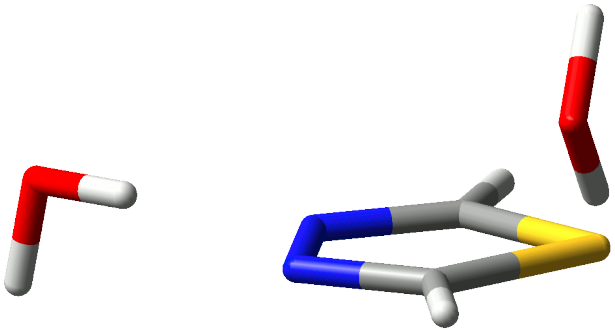
- Hard degrees of freedom: bonds, angles and rigid dihedrals and impropers
- Soft degrees of freedom: flexible dihedrals and impropers

- TDAZ: all degrees of freedom are hard
  $\Rightarrow$ no PES

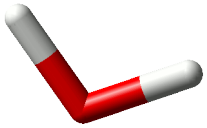


See generate.inp

TDAZ water interactions (1)



See water_constr.inp



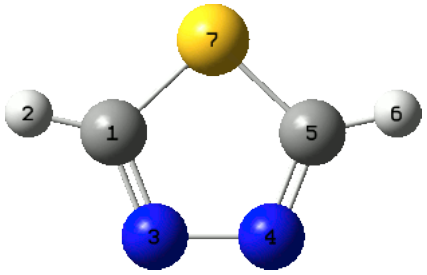
```
RESI tdaz          0.000
GROUP            ! CHARGE   CH_PENALTY
ATOM C1     CG2R53  0.224 !   24.449
ATOM H1     HGR52   0.129 !   12.902
ATOM N2     NG2R50 -0.341 !   30.278
ATOM N3     NG2R50 -0.341 !   30.278
ATOM C4     CG2R53  0.224 !   24.449
ATOM H4     HGR52   0.129 !   12.902
ATOM S5     SG2R50 -0.024 !   27.259
```

```
TDAZ WATER INTERACTION
######## USING MP2 GEOMETRY
QM DIPOLE (DEBYE)
MP2/6-31G*: X= -3.1938 Y= 1.2595 Z= 0.0011 TOT= 3.4332
HF/6-31G*: X= -3.4411 Y= 1.3570 Z= 0.0012 TOT= 3.6990
EMPIRICAL DIPOLE: -3.82723 1.50925 1.234034E-03 4.11406
1) C1H..OHH 90.0 DEGREES
A.I. -3.95 2.28
EMP. -3.44291 -3.46935 2.643415E-02 2.25
ENE DIFF: 0.50709 DIST DIFF: -3E-02
2) N2...HOH, 90. DEGREES
A.I. -5.33 2.19
EMP. -4.28255 -4.43091 0.148366 2.05
ENE DIFF: 1.04745 DIST DIFF: -0.14
3) S5...HOH, 90. DEGREES
A.I. -0.37 2.90
EMP. 0.24841 0.646316 -0.397906 2.87
ENE DIFF: 0.61841 DIST DIFF: -3E-02
4) S5...HOH LONE PAIR, HOH: 0. DEGREES
A.I. -0.44 2.84
EMP. 0.457146 0.866105 -0.408958 3
ENE DIFF: 0.897146 DIST DIFF: 0.16
AVEDIFF, RMSDIFF, AVERAGE ABSOLUTE ERROR
0.767525 0.21518 0.767525
```

```
RESI TDAZ          0.000 ! 1,3,4-thiadiazole, jon
GROUP                    !          H1
ATOM C1     CG2R53  0.53  !        ____/
ATOM H1     HGR52   0.08  !    N2----C1
ATOM N2     NG2R50 -0.48  !     |     \
ATOM N3     NG2R50 -0.48  !     |      S5
ATOM C4     CG2R53  0.53  !     |____ /
ATOM H4     HGR52   0.08  !    N3----C4
ATOM S5     SG2R50 -0.26  !          \
                 !          H4
```

```
TDAZ WATER INTERACTION
######## USING MP2 GEOMETRY
QM DIPOLE (DEBYE)
MP2/6-31G*: X= -3.1938 Y= 1.2595 Z= 0.0011 TOT= 3.4332
HF/6-31G*: X= -3.4411 Y= 1.3570 Z= 0.0012 TOT= 3.6990
EMPIRICAL DIPOLE: -3.93696 1.55252 1.196776E-03 4.23202
1) C1H..OHH 90.0 DEGREES
A.I. -3.95 2.28
EMP. -3.89637 -3.99037 9.400732E-02 2.23
ENE DIFF: 5.363E-02 DIST DIFF: -5E-02
2) N2...HOH, 90. DEGREES
A.I. -5.33 2.19
EMP. -5.24768 -5.64785 0.400169 2
ENE DIFF: 8.232E-02 DIST DIFF: -0.19
3) S5...HOH, 90. DEGREES
A.I. -0.37 2.90
EMP. -0.644116 -0.284353 -0.359763 2.5
ENE DIFF: -0.274116 DIST DIFF: -0.4
4) S5...HOH LONE PAIR, HOH: 0. DEGREES
A.I. -0.44 2.84
EMP. -0.365761 7.597964E-02 -0.44174 2.51
ENE DIFF: 7.4239E-02 DIST DIFF: -0.33
AVEDIFF, RMSDIFF, AVERAGE ABSOLUTE ERROR
-1.598175E-02 0.149401 0.121076
```

TDAZ reference angles (1)




```
ANGLES
CG2R53 NG2R50 NG2R50  110.00    106.80 ! ... penalty= 3
CG2R53 SG2R50 CG2R53  110.00     97.00 ! ... penalty= 3
```

```
ANGLES
CG2R53 NG2R50 NG2R50  110.00    104.50 ! TDAZ,
CG2R53 SG2R50 CG2R53  110.00     96.60 ! TDAZ
```

```
RESI TDAZ ring 5 C1 N2 N3 C4 S5 : sum = 545 (540)
Parameters involved:
../../toppar/tdaz_water1.str line 37: CG2R53 NG2R50 NG2R50 ang=106.80
../../toppar/tdaz_water1.str line 37: CG2R53 NG2R50 NG2R50 ang=106.80
../../toppar/par_all36_cgenff.prm line 1059: NG2R50 CG2R53 SG2R50 ang=117.20
../../toppar/tdaz_water1.str line 38: CG2R53 SG2R50 CG2R53 ang=97.00
../../toppar/par_all36_cgenff.prm line 1059: NG2R50 CG2R53 SG2R50 ang=117.20
```

```
CHARMM>    !C4-S5-C1    86.4369
CHARMM>    quick 5 7 1
QUICKA: The angle is:    86.314

CHARMM>    !S5-C1-N2   114.8867
CHARMM>    quick 7 1 3
QUICKA: The angle is:   114.688

CHARMM>    !C1-N2-N3   111.8949
CHARMM>    quick 1 3 4
QUICKA: The angle is:   112.155

CHARMM>    !N2-C1-H1   122.6403
CHARMM>    quick 3 1 2
QUICKA: The angle is:   125.656

CHARMM>    !S5-C1-H1   122.473
CHARMM>    quick 7 1 2
QUICKA: The angle is:   119.657
```

```
CHARMM>    !C4-S5-C1    86.4369
CHARMM>    quick 5 7 1
QUICKA: The angle is:    85.801

CHARMM>    !S5-C1-N2   114.8867
CHARMM>    quick 7 1 3
QUICKA: The angle is:   115.421

CHARMM>    !C1-N2-N3   111.8949
CHARMM>    quick 1 3 4
QUICKA: The angle is:   111.678

CHARMM>    !N2-C1-H1   122.6403
CHARMM>    quick 3 1 2
QUICKA: The angle is:   125.335

CHARMM>    !S5-C1-H1   122.473
CHARMM>    quick 7 1 2
QUICKA: The angle is:   119.244
```

See generate.inp

TDAZ force constants (1)

Before

```
ANGLES
CG2R53 NG2R50 NG2R50  110.00    104.50 ! TDAZ
CG2R53 SG2R50 CG2R53  110.00     96.60 ! TDAZ

DIHEDRALS
SG2R50 CG2R53 NG2R50 NG2R50     6.0000  2   180.00
HGR52  CG2R53 NG2R50 NG2R50     5.5000  2   180.00
NG2R50 CG2R53 SG2R50 CG2R53     8.5000  2   180.00
HGR52  CG2R53 SG2R50 CG2R53     5.5000  2   180.00
CG2R53 NG2R50 NG2R50 CG2R53    14.0000  2   180.00
```

```
 1   436.9  tRINGa  111.
 2   598.1  dRING    67.  ssC-S    19.
 3   625.9  tRING   105.
 4   734.8  saC-S    62.  dRINGa   36.
 5   814.3  waCH    111.
 6   833.3  ssC-S    52.  sN-N     24.
 7   859.7  wsCH    105.
 8   973.3  daCH     37.  ssC-N    25.  sN-N     22.
 9  1005.7  dRINGa   44.  saC-S    27.  saC-N    20.
10  1140.3  dsCH     55.  saC-N    41.
11  1320.4  sN-N     48.  ssC-N    20.  dRING    17.
12  1451.6  daCH     48.  ssC-N    36.  ssC-S    15.
13  1502.1  saC-N    37.  dsCH     36.  dRINGa   16.
14  2948.4  saCH     99.
15  2949.9  ssCH     99.
```

After

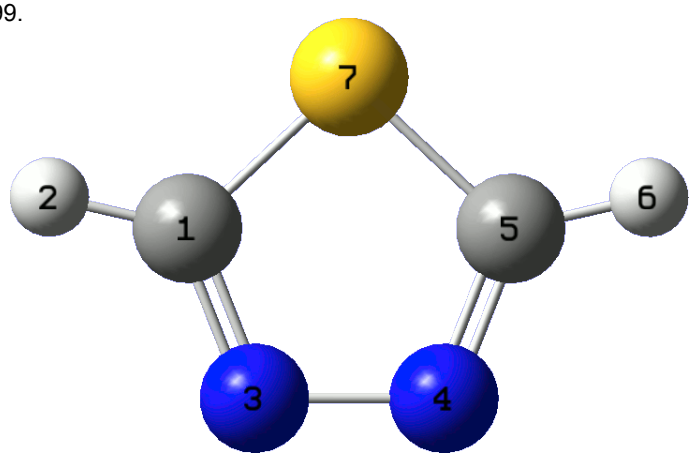
```
ANGLES
CG2R53 NG2R50 NG2R50   80.00    104.50 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     96.60 ! TDAZ

DIHEDRALS
SG2R50 CG2R53 NG2R50 NG2R50    10.9000  2   180.00
NG2R50 CG2R53 SG2R50 CG2R53     7.5000  2   180.00
CG2R53 NG2R50 NG2R50 CG2R53     4.4500  2   180.00
HGR52  CG2R53 NG2R50 NG2R50     4.3600  2   180.00
HGR52  CG2R53 SG2R50 CG2R53     4.3600  2   180.00
```

```
 1   460.1  tRINGa  110.
 2   567.5  dRING    72.  ssC-S    16.
 3   589.0  tRING   105.
 4   688.8  saC-S    50.  dRINGa   49.
 5   730.2  waCH    110.
 6   765.4  wsCH    105.
 7   826.7  ssC-S    55.  sN-N     21.
 8   961.9  daCH     35.  ssC-N    25.  sN-N     24.
 9   980.9  dRINGa   40.  saC-S    38.
10  1130.3  saC-N    52.  dsCH     47.
11  1308.9  sN-N     48.  ssC-N    20.  daCH     15.
12  1445.3  daCH     48.  ssC-N    37.
13  1473.7  dsCH     41.  saC-N    36.
14  2948.0  saCH     99.
15  2948.9  ssCH     99.
```

See
quick_molvib.inp
tdaz_molvib_qm.inp

QM:

```
 1   467.1  tRINGa  110.
 2   581.5  tRING    96.
 3   603.9  dRING    64.  ssC-S    38.
 4   731.7  wsCH     96.
 5   737.9  saC-S    96.
 6   770.7  waCH    110.
 7   877.3  dRINGa   88.
 8   887.3  ssC-S    58.  dRING    35.
 9   963.1  sN-N     86.
10  1173.2  dsCH     84.
11  1202.7  daCH     47.  ssC-N    33.
12  1327.5  saC-N    91.
13  1335.5  ssC-N    63.  daCH     37.
14  3121.6  saCH    100.
15  3124.9  ssCH     99.
```



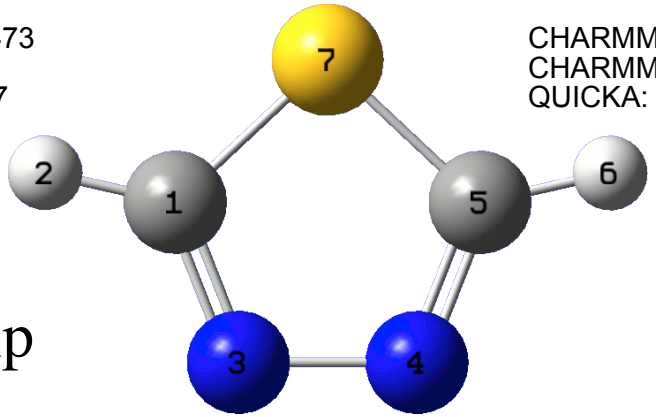
TDAZ reference angles (2)

```
ANGLES
CG2R53 NG2R50 NG2R50   80.00    104.50 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     96.60 ! TDAZ
```

```
CHARMM>    !C4-S5-C1    86.4369
CHARMM>    quick 5 7 1
QUICKA: The angle is:    84.737

CHARMM>    !S5-C1-N2    114.8867
CHARMM>    quick 7 1 3
QUICKA: The angle is:    115.897

CHARMM>    !C1-N2-N3    111.8949
CHARMM>    quick 1 3 4
QUICKA: The angle is:    111.735

CHARMM>    !N2-C1-H1    122.6403
CHARMM>    quick 3 1 2
QUICKA: The angle is:    125.126

CHARMM>    !S5-C1-H1    122.473
CHARMM>    quick 7 1 2
QUICKA: The angle is:    118.977
```

```
ANGLES
CG2R53 NG2R50 NG2R50   80.00    103.65 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     98.30 ! TDAZ
```

```
CHARMM>    !C4-S5-C1    86.4369
CHARMM>    quick 5 7 1
QUICKA: The angle is:    85.061

CHARMM>    !S5-C1-N2    114.8867
CHARMM>    quick 7 1 3
QUICKA: The angle is:    115.755

CHARMM>    !C1-N2-N3    111.8949
CHARMM>    quick 1 3 4
QUICKA: The angle is:    111.715

CHARMM>    !N2-C1-H1    122.6403
CHARMM>    quick 3 1 2
QUICKA: The angle is:    125.188

CHARMM>    !S5-C1-H1    122.473
CHARMM>    quick 7 1 2
QUICKA: The angle is:    119.057
```



See generate.inp

Reconsider geometry following adjustment of force constants

TDAZ force constants (2)

Before

```
ANGLES
CG2R53 NG2R50 NG2R50   80.00    104.50 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     96.60 ! TDAZ

DIHEDRALS
SG2R50 CG2R53 NG2R50 NG2R50    10.9000  2   180.00
NG2R50 CG2R53 SG2R50 CG2R53     7.5000  2   180.00
CG2R53 NG2R50 NG2R50 CG2R53     4.4500  2   180.00
HGR52  CG2R53 NG2R50 NG2R50     4.3600  2   180.00
HGR52  CG2R53 SG2R50 CG2R53     4.3600  2   180.00
```

```
 1   460.1  tRINGa  110.
 2   567.5  dRING    72.  ssC-S    16.
 3   589.0  tRING   105.
 4   688.8  saC-S    50.  dRINGa   49.
 5   730.2  waCH    110.
 6   765.4  wsCH    105.
 7   826.7  ssC-S    55.  sN-N     21.
 8   961.9  daCH     35.  ssC-N    25.  sN-N     24.
 9   980.9  dRINGa   40.  saC-S    38.
10  1130.3  saC-N    52.  dsCH     47.
11  1308.9  sN-N     48.  ssC-N    20.  daCH     15.
12  1445.3  daCH     48.  ssC-N    37.
13  1473.7  dsCH     41.  saC-N    36.
14  2948.0  saCH     99.
15  2948.9  ssCH     99.
```

After

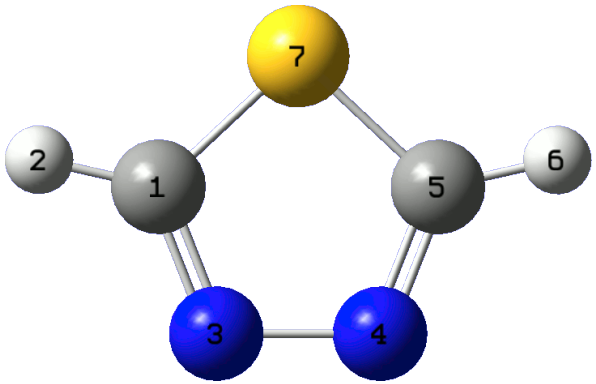
```
ANGLES
CG2R53 NG2R50 NG2R50   80.00    103.65 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     98.30 ! TDAZ
DIHEDRALS
SG2R50 CG2R53 NG2R50 NG2R50    10.9000  2   180.00
NG2R50 CG2R53 SG2R50 CG2R53     7.5000  2   180.00
CG2R53 NG2R50 NG2R50 CG2R53     4.4500  2   180.00
HGR52  CG2R53 NG2R50 NG2R50     4.3600  2   180.00
HGR52  CG2R53 SG2R50 CG2R53     4.3600  2   180.00
```

```
 1   461.3  tRINGa  110.
 2   566.6  dRING    72.  ssC-S    16.
 3   583.5  tRING   105.
 4   688.8  saC-S    50.  dRINGa   49.
 5   729.8  waCH    110.
 6   764.9  wsCH    105.
 7   827.0  ssC-S    56.  sN-N     21.
 8   962.3  daCH     35.  ssC-N    25.  sN-N     24.
 9   981.3  dRINGa   39.  saC-S    38.
10  1130.3  saC-N    52.  dsCH     47.
11  1308.2  sN-N     48.  ssC-N    20.
12  1445.2  daCH     49.  ssC-N    37.
13  1474.0  dsCH     41.  saC-N    36.
14  2948.0  saCH     99.
15  2948.9  ssCH     99.
```

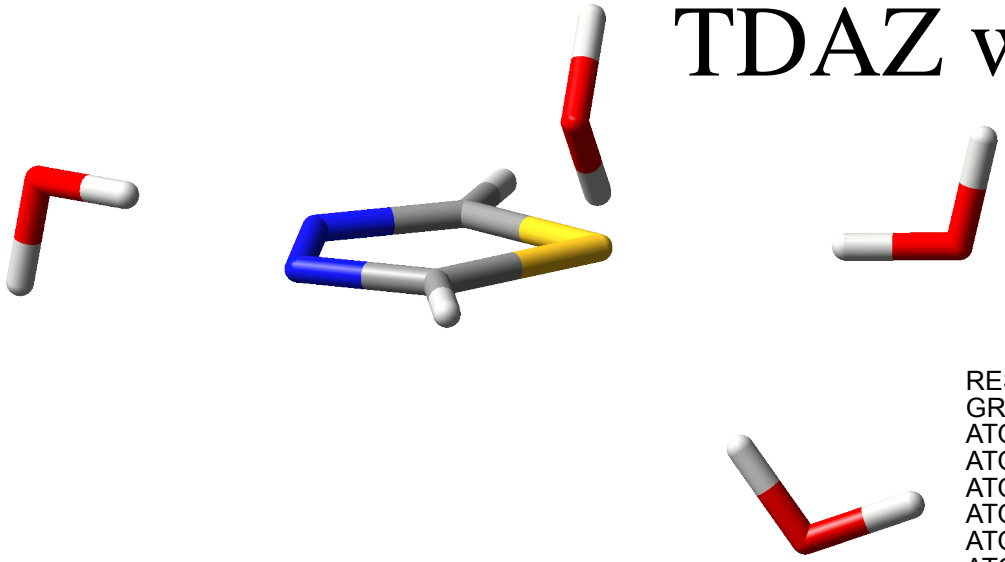
Resonsider FCs following adjustment of equilibrium terms: Self-consistent.

QM:

```
 1   467.1  tRINGa  110.
 2   581.5  tRING    96.
 3   603.9  dRING    64.  ssC-S    38.
 4   731.7  wsCH     96.
 5   737.9  saC-S    96.
 6   770.7  waCH    110.
 7   877.3  dRINGa   88.
 8   887.3  ssC-S    58.  dRING    35.
 9   963.1  sN-N     86.
10  1173.2  dsCH     84.
11  1202.7  daCH     47.  ssC-N    33.
12  1327.5  saC-N    91.
13  1335.5  ssC-N    63.  daCH     37.
14  3121.6  saCH    100.
15  3124.9  ssCH     99.
```



TDAZ water interactions (2)



See water_constr.inp
water.inp

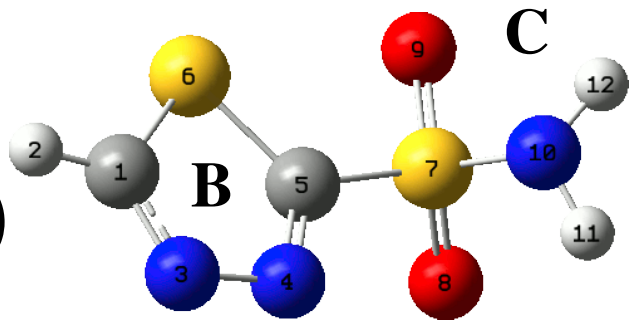
```
RESI TDAZ          0.000 ! 1,3,4-thiadiazole, jon
GROUP                    !          H1
ATOM C1    CG2R53  0.53  !        _____/
ATOM H1    HGR52   0.08  !    N2----C1
ATOM N2    NG2R50 -0.48  !     |     \
ATOM N3    NG2R50 -0.48  !     |      S5
ATOM C4    CG2R53  0.53  !     |_____/
ATOM H4    HGR52   0.08  !    N3----C4
ATOM S5    SG2R50 -0.26  !          \
                 !          H4
```

TDAZ WATER INTERACTION
######## USING MP2 GEOMETRY
QM DIPOLE (DEBYE)
MP2/6-31G*: X= -3.1938 Y= 1.2595 Z= 0.0011 TOT= 3.4332
HF/6-31G*: X= -3.4411 Y= 1.3570 Z= 0.0012 TOT= 3.6990
EMPIRICAL DIPOLE: -3.93696 1.55252 1.196776E-03 4.23202
1) C1H..OHH 90.0 DEGREES
A.I. -3.95 2.28
EMP. -3.89637 -3.99037 9.400732E-02 2.23
ENE DIFF: 5.363E-02 DIST DIFF: -5E-02
2) N2...HOH, 90. DEGREES
A.I. -5.33 2.19
EMP. -5.24768 -5.64785 0.400169 2
ENE DIFF: 8.232E-02 DIST DIFF: -0.19
3) S5...HOH, 90. DEGREES
A.I. -0.37 2.90
EMP. -0.644116 -0.284353 -0.359763 2.5
ENE DIFF: -0.274116 DIST DIFF: -0.4
4) S5...HOH LONE PAIR, HOH: 0. DEGREES
A.I. -0.44 2.84
EMP. -0.365761 7.597964E-02 -0.44174 2.51
ENE DIFF: 7.4239E-02 DIST DIFF: -0.33
AVEDIFF, RMSDIFF, AVERAGE ABSOLUTE ERROR
-1.598175E-02 0.149401 0.121076

TDAZ WATER INTERACTION
QM DIPOLE (DEBYE)
MP2/6-31G*: X= -3.1938 Y= 1.2595 Z= 0.0011 TOT= 3.4332
HF/6-31G*: X= -3.4411 Y= 1.3570 Z= 0.0012 TOT= 3.6990
EMPIRICAL DIPOLE: -4.0089 1.58089 1.381163E-03 4.30935
1) C1H..OHH 90.0 DEGREES
A.I. -3.95 2.28
EMP. -3.95937 -4.03811 7.874136E-02 2.23
ENE DIFF: -9.37E-03 DIST DIFF: -5E-02
2) N2...HOH, 90. DEGREES
A.I. -5.33 2.19
EMP. -5.36892 -5.76325 0.394332 2
ENE DIFF: -3.892E-02 DIST DIFF: -0.19
3) S5...HOH, 90. DEGREES
A.I. -0.37 2.90
EMP. -0.604375 -0.244857 -0.359518 2.5
ENE DIFF: -0.234375 DIST DIFF: -0.4
4) S5...HOH LONE PAIR, HOH: 0. DEGREES
A.I. -0.44 2.84
EMP. -0.317244 0.124194 -0.441438 2.51
ENE DIFF: 0.122756 DIST DIFF: -0.33
AVEDIFF, RMSDIFF, AVERAGE ABSOLUTE ERROR
-3.997725E-02 0.127682 0.101355

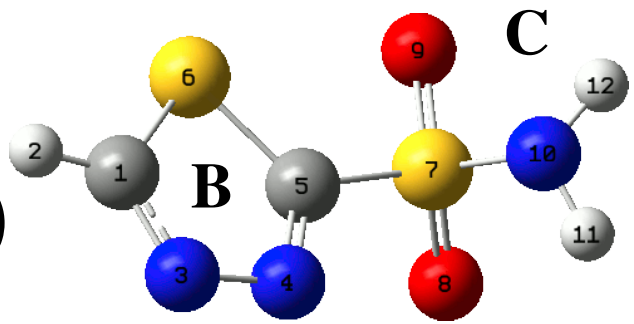
… nothing to do ☺

SULFAZ (1)




```
RESI SULFAZ        0.00 ! 5-sulfamoyl-1,3,4-thiadiazol-2-yl
GROUP ! charges put together manually from resi TDAZ and BSAM
ATOM C1     CG2R53  0.53
ATOM H1     HGR52   0.08 ! Following standard CHARMM rules, we should sum
ATOM N2     NG2R50 -0.48 ! both TDAZ H4 and BSAM CZ into C4. However, 2-thiadiazolyl
ATOM N3     NG2R50 -0.48 ! is significantly electron-deficient compared to phenyl
ATOM C4     CG2R53  0.82 ! We can't really take this into account quantitatively at
ATOM S5     SG2R50 -0.26 ! this stage, but we can sum a conservative +0.03 charge
ATOM S9     SG3O2   0.64 ! into S9 instead of C4.
ATOM O91    OG2P1  -0.42 ! To capture this effect quantitatively, water interactions
ATOM O92    OG2P1  -0.42 ! are necessary.
ATOM N10    NG321  -0.77
ATOM H101   HGP1    0.38
ATOM H102   HGP1    0.38

...

BONDS
CG2R53 SG3O2   190.00     1.7300 ! acazam , from CG2R61 SG3O2, penalty= 40

ANGLES
CG2R53 NG2R50 NG2R50   80.00    103.65 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     98.30 ! TDAZ

NG2R50 CG2R53 SG3O2    70.00    109.00 ! acazam , from NG2R53 CG2R53 SG311, penalty= 40
SG2R50 CG2R53 SG3O2    70.00    109.00 ! acazam , from NG2R53 CG2R53 SG311, penalty= 144
CG2R53 SG3O2  NG321    60.00     98.00 ! acazam , from CG2R61 SG3O2 NG321, penalty= 8.5
CG2R53 SG3O2  OG2P1    60.00    101.00 ! acazam , from CG2R61 SG3O2 OG2P1, penalty= 8.5
```

SULFAZ (2)

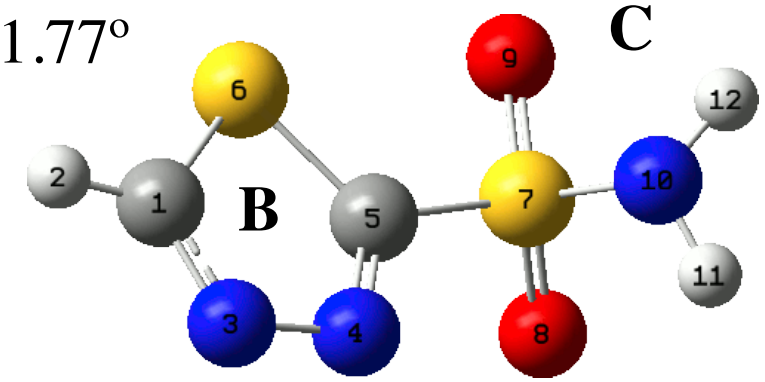



```
DIHEDRALS
SG2R50 CG2R53 NG2R50 NG2R50    10.9000  2   180.00 ! TDAZ
NG2R50 CG2R53 SG2R50 CG2R53     7.5000  2   180.00 ! TDAZ
CG2R53 NG2R50 NG2R50 CG2R53     4.4500  2   180.00 ! TDAZ
HGR52  CG2R53 NG2R50 NG2R50     4.3600  2   180.00 ! TDAZ
HGR52  CG2R53 SG2R50 CG2R53     4.3600  2   180.00 ! TDAZ

! SG3O2  CG2R53 NG2R50 NG2R50 is a compromise between
! SG2R50 CG2R53 NG2R50 NG2R50    10.9000  2   180.00
! and
! HGR52  CG2R53 NG2R50 NG2R50     4.3600  2   180.00
! in TDAZ. We know that heavy atoms substituents often need a higher force constant
! than hydrogens, but a higher force constant also unnaturally rigidifies the ring,
! and we don't have the freedom to compensate for that somewhere else.
! Similarly, G3O2  CG2R53 SG2R50 CG2R53 is a compromise between
! NG2R50 CG2R53 SG2R50 CG2R53     7.5000  2   180.00
! and
! HGR52  CG2R53 SG2R50 CG2R53     4.3600  2   180.00
! This is all rather coarse and should strictly spoken be refined by doing molvib
! and/or pes on TDAZ with a small (possibly methyl) substituent, but in the end,
! it will have only minor impact on the behavior of the system during MD simulations.
SG3O2  CG2R53 NG2R50 NG2R50     7.6300  2   180.00 ! ACAZAM & model compounds, average from TDAZ
SG3O2  CG2R53 SG2R50 CG2R53     5.9300  2   180.00 ! ACAZAM & model compounds, average from TDAZ

NG2R50 CG2R53 SG3O2  NG321      0.3500  2     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 NG321, penalty= 125
NG2R50 CG2R53 SG3O2  OG2P1      0.0000  6     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 OG2P1, penalty= 125
SG2R50 CG2R53 SG3O2  NG321      0.3500  2     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 NG321, penalty= 209
SG2R50 CG2R53 SG3O2  OG2P1      0.0000  6     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 OG2P1, penalty= 209
HGP1   NG321  SG3O2  CG2R53     1.5000  1   180.00 ! acazam , from HGP1 NG321 SG3O2 CG2R61, penalty= 8.5
HGP1   NG321  SG3O2  CG2R53     1.2000  2     0.00 ! acazam , from HGP1 NG321 SG3O2 CG2R61, penalty= 8.5
HGP1   NG321  SG3O2  CG2R53     0.1000  3     0.00 ! acazam , from HGP1 NG321 SG3O2 CG2R61, penalty= 8.5
```

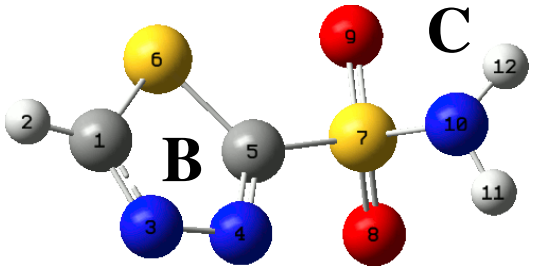
SULFAZ reference angles (1)

```
ANGLES
NG2R50 CG2R53 SG3O2    70.00    109.00 ! acazam , from NG2R53 CG2R53 SG311, penalty= 40
SG2R50 CG2R53 SG3O2    70.00    109.00 ! acazam , from NG2R53 CG2R53 SG311, penalty= 144


RESI SULFAZ center C4 bound to N3 S5 S9 : sum = 335.2 (360)
Parameters involved:
../../../toppar/par_all36_cgenff.prm line 1059: NG2R50 CG2R53 SG2R50 ang=117.20
../../../toppar/acazam_models_init.str line 145: SG2R50 CG2R53 SG3O2  ang=109.00
../../../toppar/acazam_models_init.str line 143: NG2R50 CG2R53 SG3O2  ang=109.00
RESI SULFAZ center N10 bound to S9 H101 H102 : sum = 340 (360)
Parameters involved:
../../../toppar/par_all36_cgenff.prm line 1924: SG3O2  NG321  HGP1   ang=115.00
../../../toppar/par_all36_cgenff.prm line 1925: HGP1   NG321  HGP1   ang=110.00
../../../toppar/par_all36_cgenff.prm line 1924: SG3O2  NG321  HGP1   ang=115.00


ANGLES
NG2R50 CG2R53 SG3O2    70.00    120.00 ! SULFAZ
SG2R50 CG2R53 SG3O2    70.00    122.80 ! SULFAZ
```

Angle (4,5,7) changed 121.09° → 121.77°
(QM target = 121.84°)

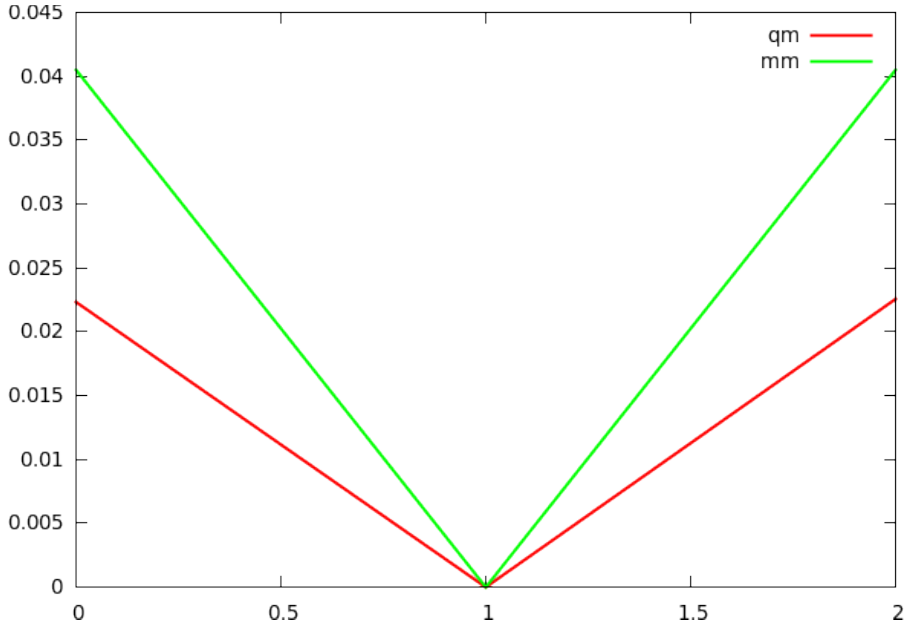


SULFAZ angle force constant: 3-point scan (1)

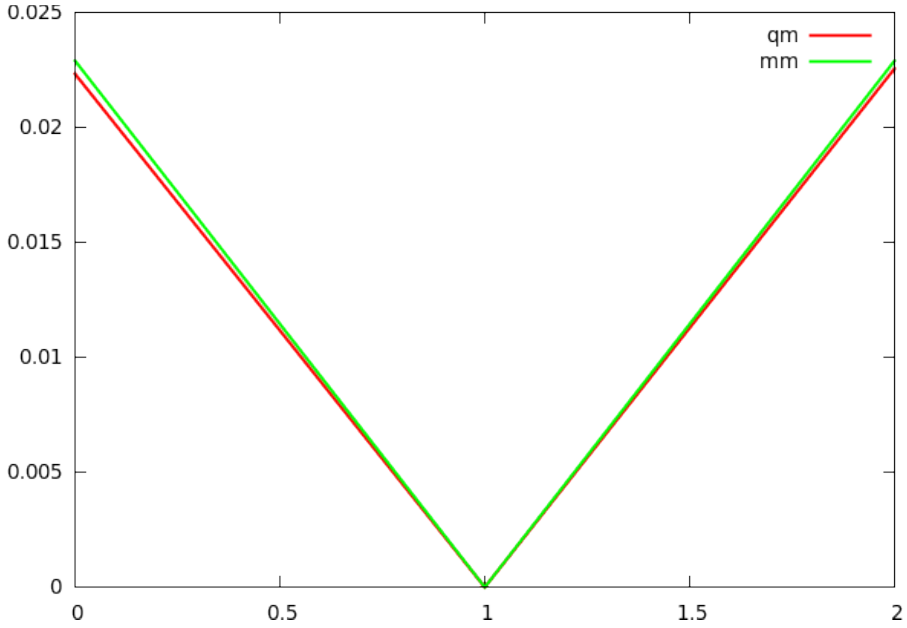


3-point scan: simple alternative to molvib if # of parameters is small. See angscan.inp

```
ANGLES
NG2R50 CG2R53 SG3O2    70.00    120.00 ! SULFAZ
SG2R50 CG2R53 SG3O2    70.00    122.80 ! SULFAZ
```




```
ANGLES
NG2R50 CG2R53 SG3O2    30.00    120.00 ! SULFAZ
SG2R50 CG2R53 SG3O2    30.00    122.80 ! SULFAZ
```



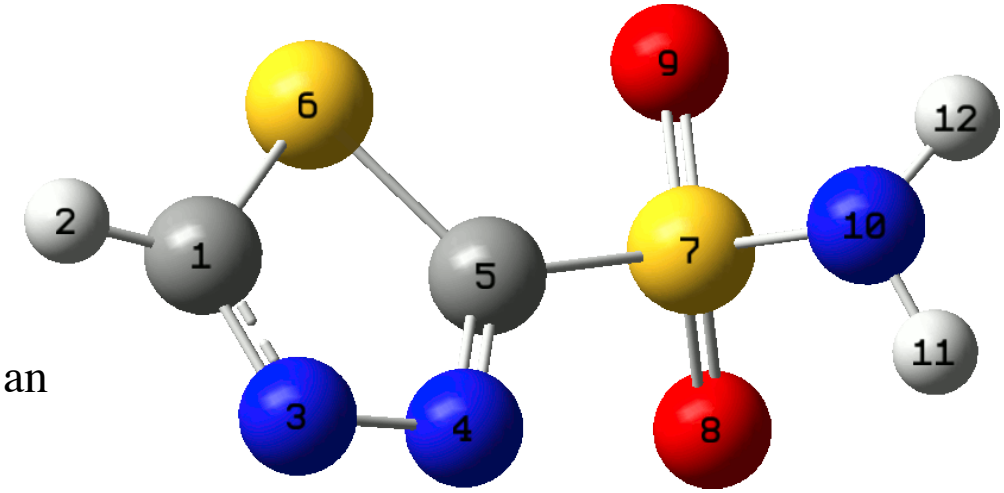
SULFAZ reference angles (2)

```
ANGLES
NG2R50 CG2R53 SG3O2    30.00    120.00 ! SULFAZ
SG2R50 CG2R53 SG3O2    30.00    122.80 ! SULFAZ
```

```
ANGLES
NG2R50 CG2R53 SG3O2    30.00    119.50 ! SULFAZ
SG2R50 CG2R53 SG3O2    30.00    123.30 ! SULFAZ
```

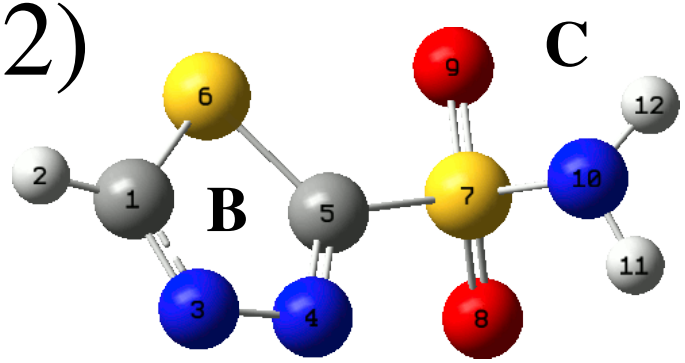
Angle (4,5,7) changed 122.21° → 121.86°
(QM target = 121.84°)

Sum of equilibrium angles for planar system should equal 540 for a 5-membered ring and 720 for a 6-membered ring. Sum of angles around an sp2 carbon should equal 360.

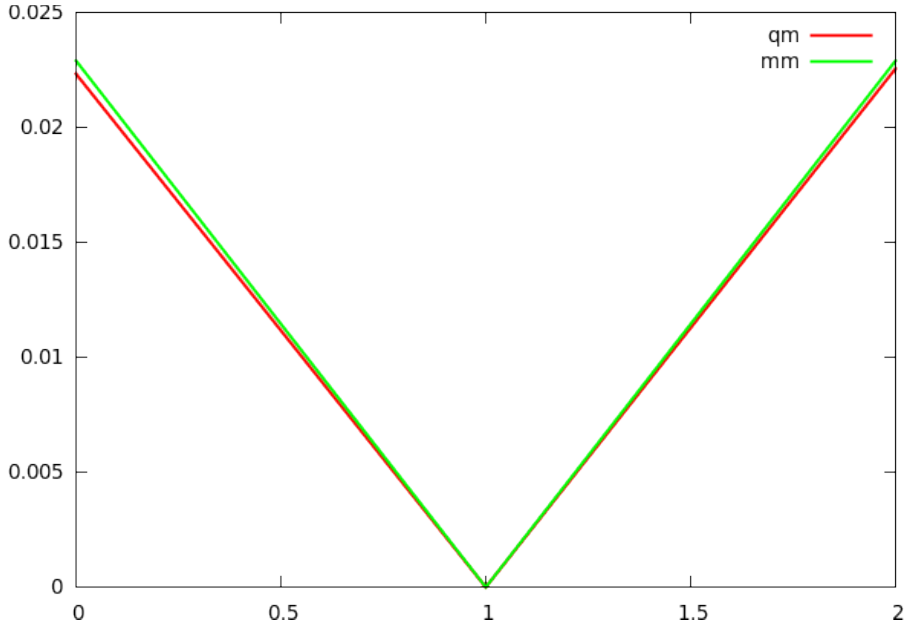


SULFAZ angle force constant: 3-point scan (2)

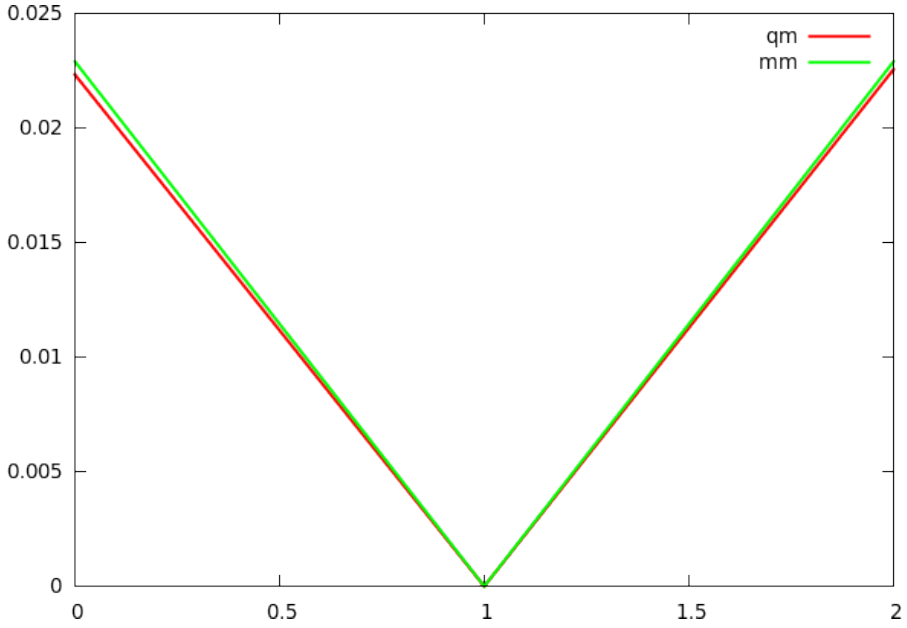
Self-consistency achieved




```
ANGLES
NG2R50 CG2R53 SG3O2    30.00    120.00 ! SULFAZ
SG2R50 CG2R53 SG3O2    30.00    122.80 ! SULFAZ
```

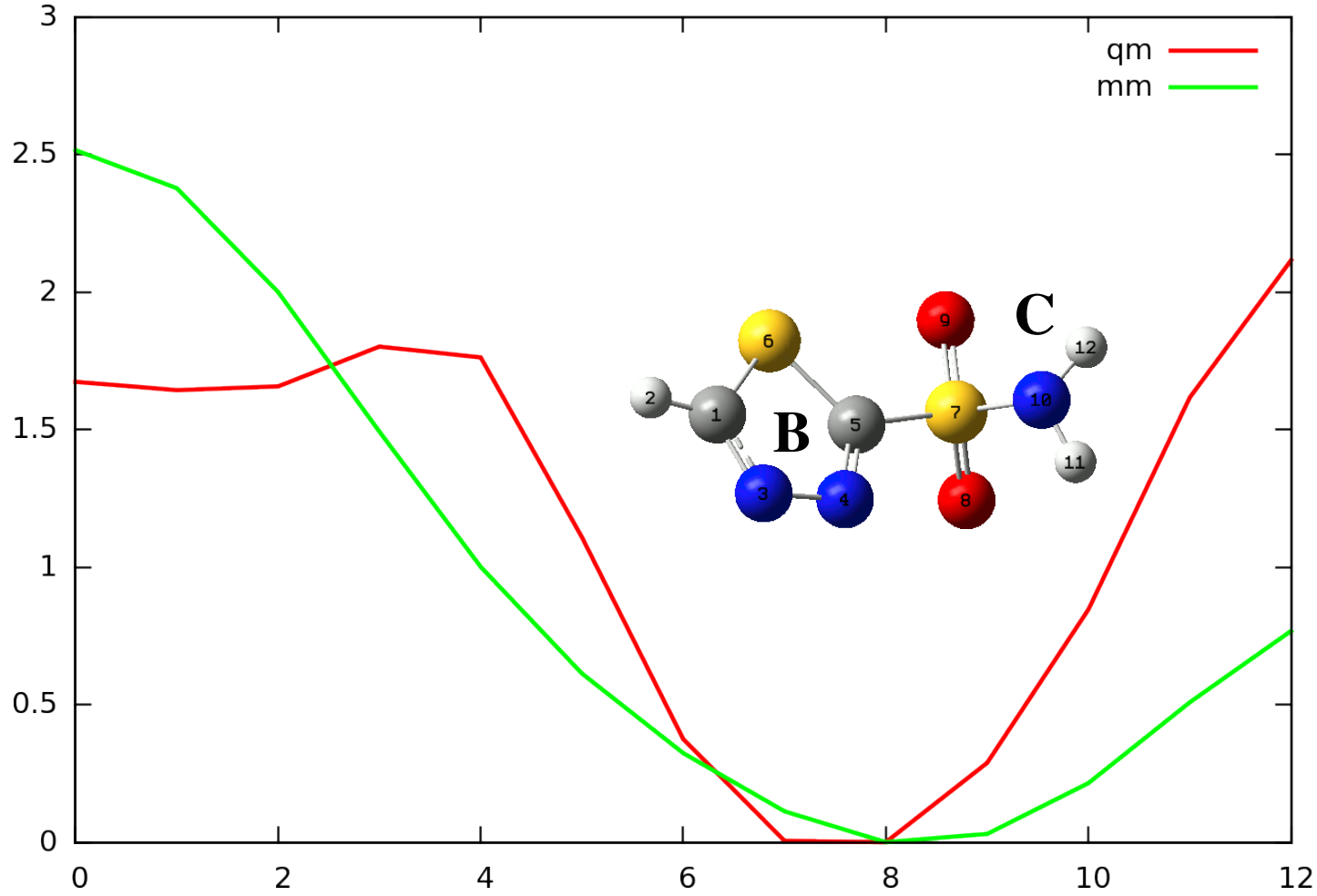



```
ANGLES
NG2R50 CG2R53 SG3O2    30.00    119.50 ! SULFAZ
SG2R50 CG2R53 SG3O2    30.00    123.30 ! SULFAZ
```



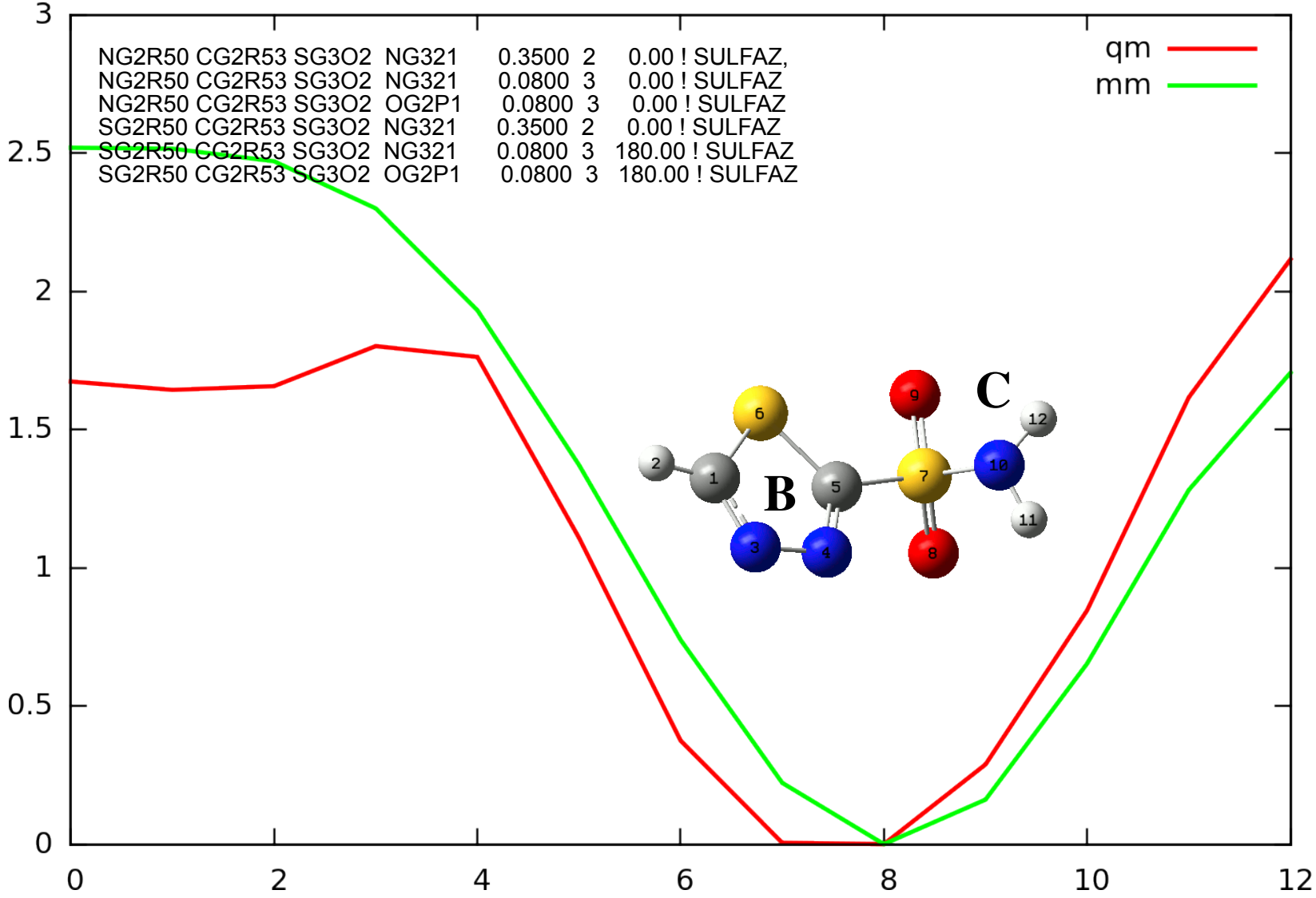
SULFAZ PES: before

```
NG2R50 CG2R53 SG3O2  NG321      0.3500  2     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 NG321, penalty= 125
NG2R50 CG2R53 SG3O2  OG2P1      0.0000  6     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 OG2P1, penalty= 125
SG2R50 CG2R53 SG3O2  NG321      0.3500  2     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 NG321, penalty= 209
SG2R50 CG2R53 SG3O2  OG2P1      0.0000  6     0.00 ! acazam , from CG2R61 CG2R61 SG3O2 OG2P1, penalty= 209
```



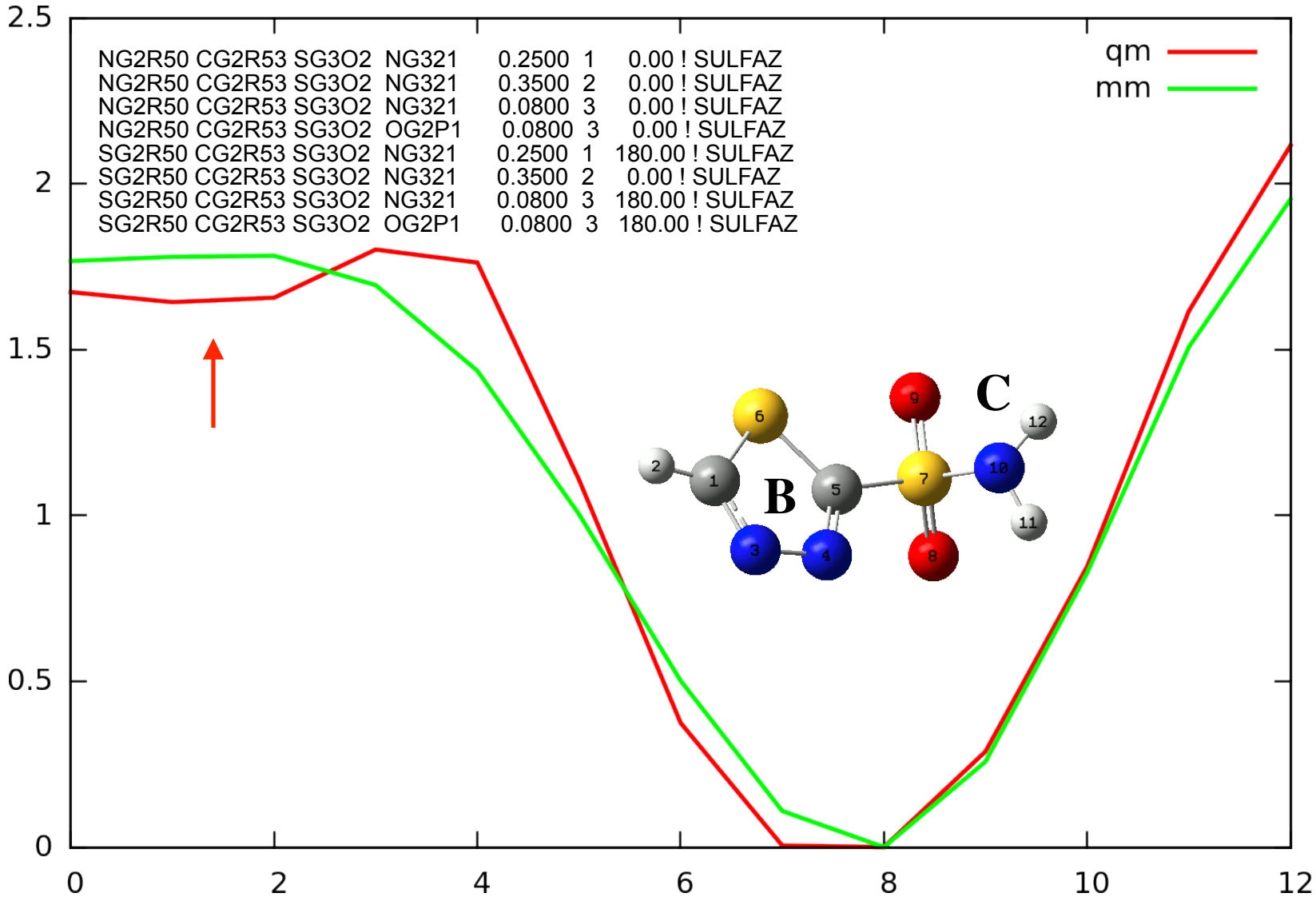
Most obvious deficiency: 3-fold See allscan.inp

SULFAZ PES: with 3-fold



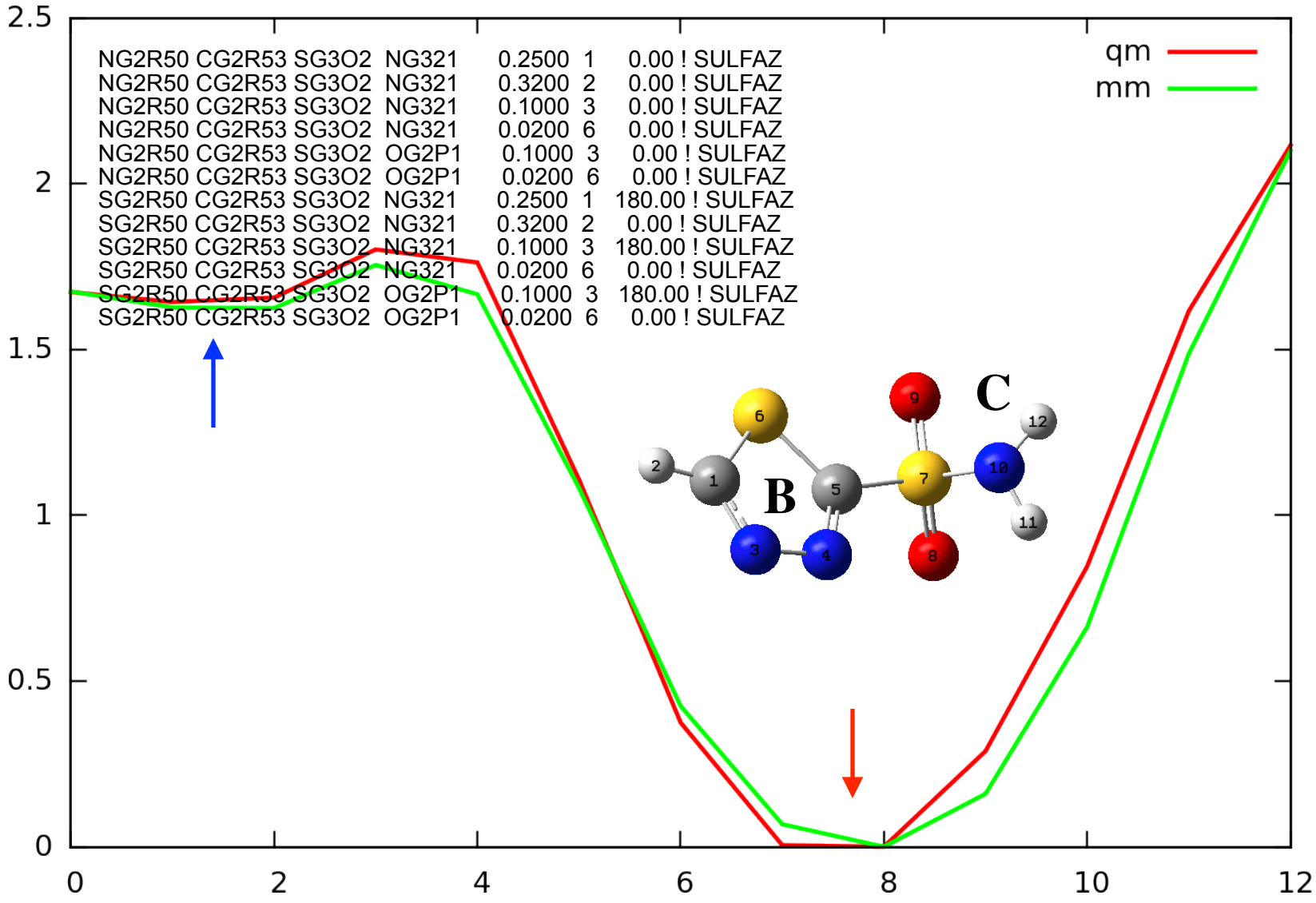
Most obvious deficiency: 1-fold

SULFAZ PES: with 1-fold



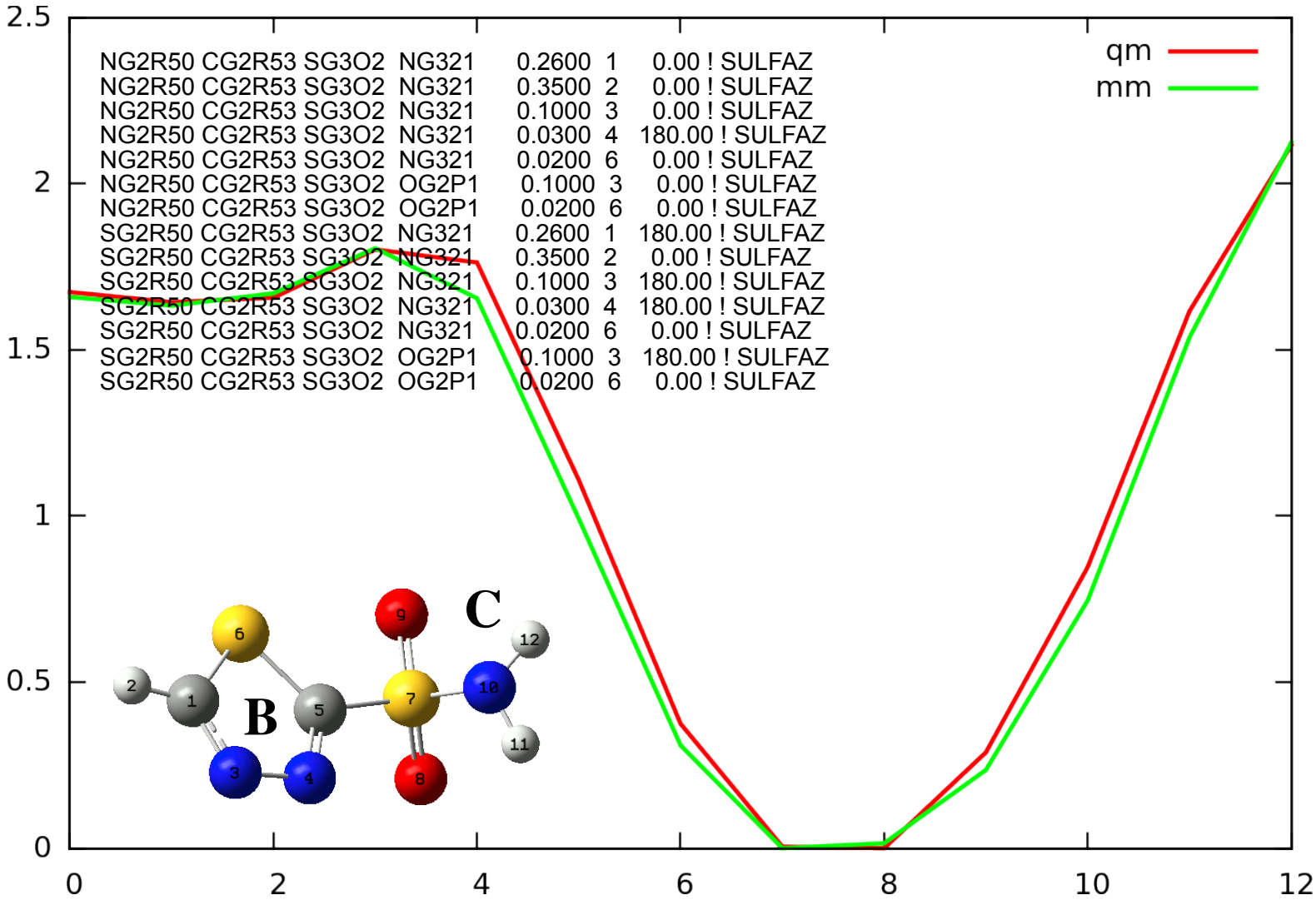
Most obvious deficiency: 6-fold

SULFAZ PES: with 6-fold



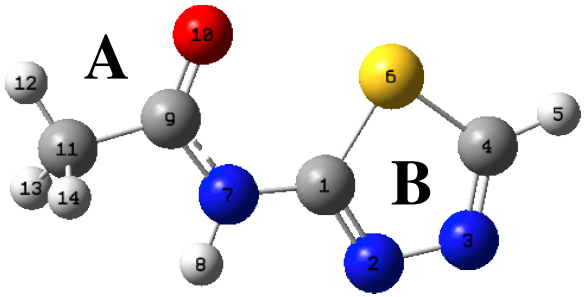
Introduction of 6-fold required rebalancing 3- & 1-fold

SULFAZ PES: with 4-fold



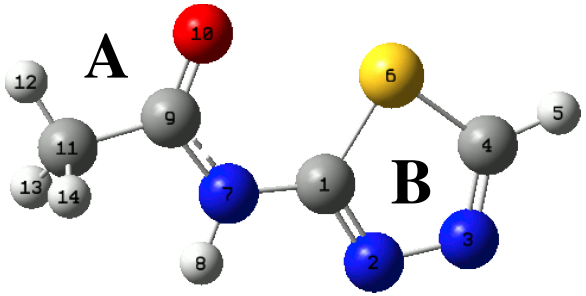
- Actually combination of MCSA fitting and manual tweaking
- Diminishing returns… accuracy of QM??

TDAZAM (1)




```
RESI TDAZAM         0.00 ! N-(1,3,4-thiadiazol-2-yl)acetamide
GROUP                    ! Charges put together manually from resi TDAZ and PACP
ATOM C1     CG2R53  0.71
ATOM N2     NG2R50 -0.48 ! Following standard CHARMM rules, we should sum
ATOM N3     NG2R50 -0.48 ! both TDAZ H1 and PACP C22 into C1. However, 2-thiadiazolyl
ATOM C4     CG2R53  0.53 ! is significantly electron-deficient compared to 4-phenolyl
ATOM H4     HGR52   0.08 ! We can't really take this into account quantitatively at
ATOM S5     SG2R50 -0.26 ! this stage, but we can sum a conservative +0.04 charge
ATOM N6     NG2S1  -0.43 ! into N6 instead of C1.
ATOM H6     HGP1    0.33 ! To capture this effect quantitatively, water interactions
ATOM C7     CG2O1   0.52 ! are necessary.
ATOM O7     OG2D1  -0.52
ATOM C8     CG331  -0.27
ATOM H81    HGA3    0.09
ATOM H82    HGA3    0.09
ATOM H83    HGA3    0.09

...

BONDS
CG2R53 NG2S1   305.00     1.4140 ! acazam , from CG2R61 NG2S1, penalty= 40

ANGLES
CG2R53 NG2R50 NG2R50   80.00    103.65 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     98.30 ! TDAZ

NG2R50 CG2R53 NG2S1    65.00    127.80 ! acazam , from NG2R53 CG2R53 OG2D1, penalty= 57
NG2S1  CG2R53 SG2R50   30.00    118.00 ! acazam , from SG2R50 CG2R53 HGR52, penalty= 144
CG2O1  NG2S1  CG2R53   50.00    120.00 ! acazam , from CG2O1 NG2S1 CG2R61, penalty= 8.5
CG2R53 NG2S1  HGP1     34.00    117.00 ! acazam , from CG2R61 NG2S1 HGP1, penalty= 8.5
```

TDAZAM (2)



DIHEDRALS
SG2R50 CG2R53 NG2R50 NG2R50 10.9000 2 180.00 ! TDAZ
NG2R50 CG2R53 SG2R50 CG2R53 7.5000 2 180.00 ! TDAZ
CG2R53 NG2R50 NG2R50 CG2R53 4.4500 2 180.00 ! TDAZ
HGR52 CG2R53 NG2R50 NG2R50 4.3600 2 180.00 ! TDAZ
HGR52 CG2R53 SG2R50 CG2R53 4.3600 2 180.00 ! TDAZ

NG2S1 CG2R53 NG2R50 NG2R50 7.6300 2 180.00 ! ACAZAM & model compounds, average of endocyclic and exocyclic from TDAZ
NG2S1 CG2R53 SG2R50 CG2R53 5.9300 2 180.00 ! ACAZAM & model compounds, average of endocyclic and exocyclic from TDAZ
SG3O2 CG2R53 NG2R50 NG2R50 7.6300 2 180.00 ! ACAZAM & model compounds, average of endocyclic and exocyclic from TDAZ
SG3O2 CG2R53 SG2R50 CG2R53 5.9300 2 180.00 ! ACAZAM & model compounds, average of endocyclic and exocyclic from TDAZ

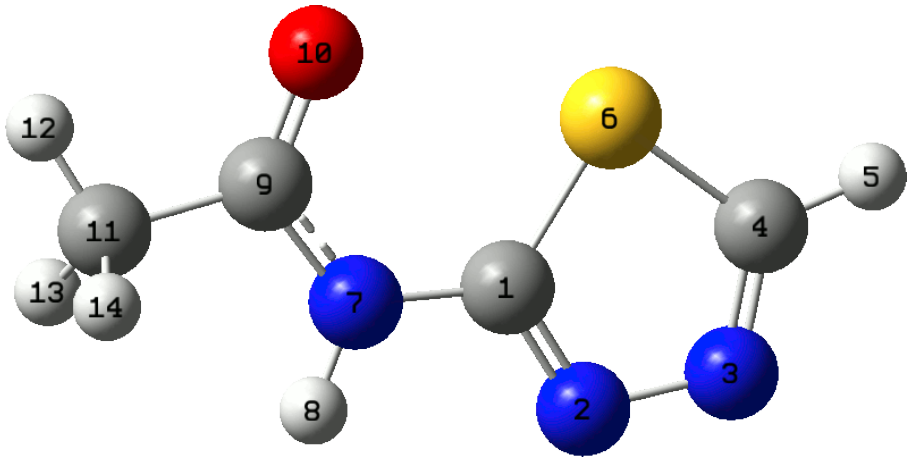
CG331 CG2O1 NG2S1 CG2R53 1.6000 1 0.00 ! acazam , from CG331 CG2O1 NG2S1 CG2R61, penalty= 8.5
CG331 CG2O1 NG2S1 CG2R53 2.5000 2 180.00 ! acazam , from CG331 CG2O1 NG2S1 CG2R61, penalty= 8.5
OG2D1 CG2O1 NG2S1 CG2R53 2.5000 2 180.00 ! acazam , from OG2D1 CG2O1 NG2S1 CG2R61, penalty= 8.5
NG2R50 CG2R53 NG2S1 CG2O1 1.5000 1 180.00 ! acazam , from NG2R60 CG2R64 NG2S1 CG2O1, penalty= 101
NG2R50 CG2R53 NG2S1 CG2O1 2.6000 2 180.00 ! acazam , from NG2R60 CG2R64 NG2S1 CG2O1, penalty= 101
NG2R50 CG2R53 NG2S1 CG2O1 0.1800 3 180.00 ! acazam , from NG2R60 CG2R64 NG2S1 CG2O1, penalty= 101
NG2R50 CG2R53 NG2S1 HGP1 0.5000 2 180.00 ! acazam , from NG2R60 CG2R64 NG2S1 HGP1, penalty= 101
SG2R50 CG2R53 NG2S1 CG2O1 1.2000 2 180.00 ! acazam , from CG2R61 CG2R61 NG2S1 CG2O1, penalty= 209
SG2R50 CG2R53 NG2S1 HGP1 0.5000 2 180.00 ! acazam , from CG2R61 CG2R61 NG2S1 HGP1, penalty= 209

IMPROPERS
CG2R53 NG2R50 NG2S1 SG2R50 43.0000 0 0.00 ! acazam , from CG2R53 NG2R53 OG2D1 SG311, penalty= 114

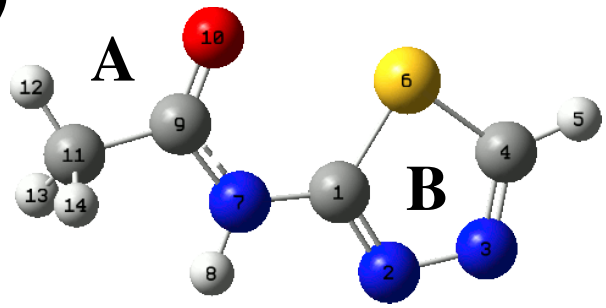
TDAZAM reference angles (1)

```
ANGLES
NG2R50 CG2R53 NG2S1    65.00    127.80 ! acazam , from NG2R53 CG2R53 OG2D1, penalty= 57
NG2S1  CG2R53 SG2R50   30.00    118.00 ! acazam , from SG2R50 CG2R53 HGR52, penalty= 144


# RESI TDAZAM center C1 bound to S5 N2 N6 : sum = 363 (360)
# Parameters involved:
# ../../../toppar/par_all36_cgenff.prm line 1059: NG2R50 CG2R53 SG2R50 ang=117.20
# ../../../toppar/acazam_sulfaz_4fold.str line 145: NG2R50 CG2R53 NG2S1  ang=127.80
# ../../../toppar/acazam_sulfaz_4fold.str line 146: NG2S1  CG2R53 SG2R50 ang=118.00


ANGLES
NG2R50 CG2R53 NG2S1    65.00    129.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   30.00    113.70 ! TDAZAM
```

Angle (2,1,7) changed 117.96° → 119.74°
(QM target = 119.75°)

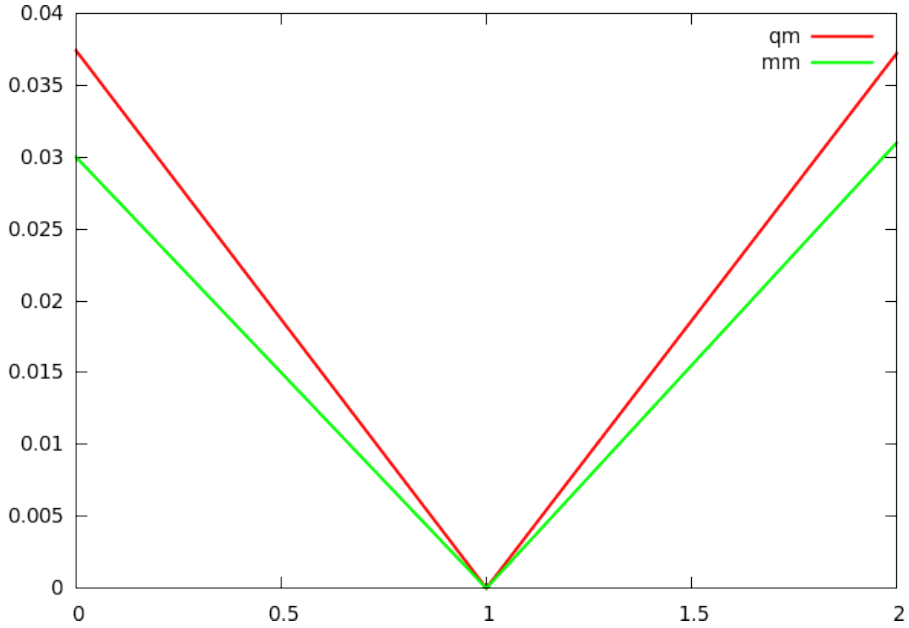


TDAZAM angle force constant: 3-point scan (1)

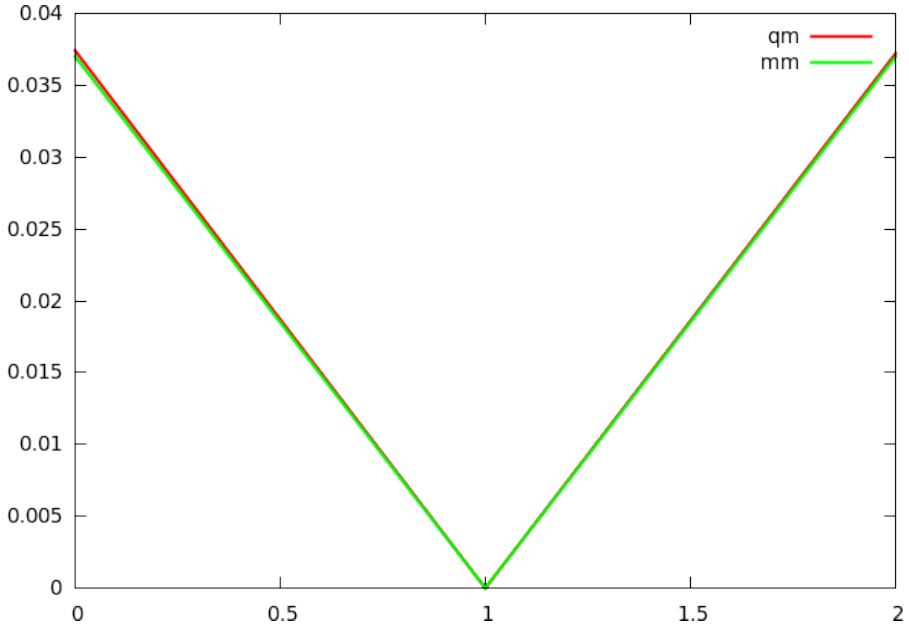


See fit_ang/angscan.inp

```
ANGLES
NG2R50 CG2R53 NG2S1    65.00    129.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   30.00    113.70 ! TDAZAM
```




```
ANGLES
NG2R50 CG2R53 NG2S1    68.00    129.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   68.00    113.70 ! TDAZAM
```

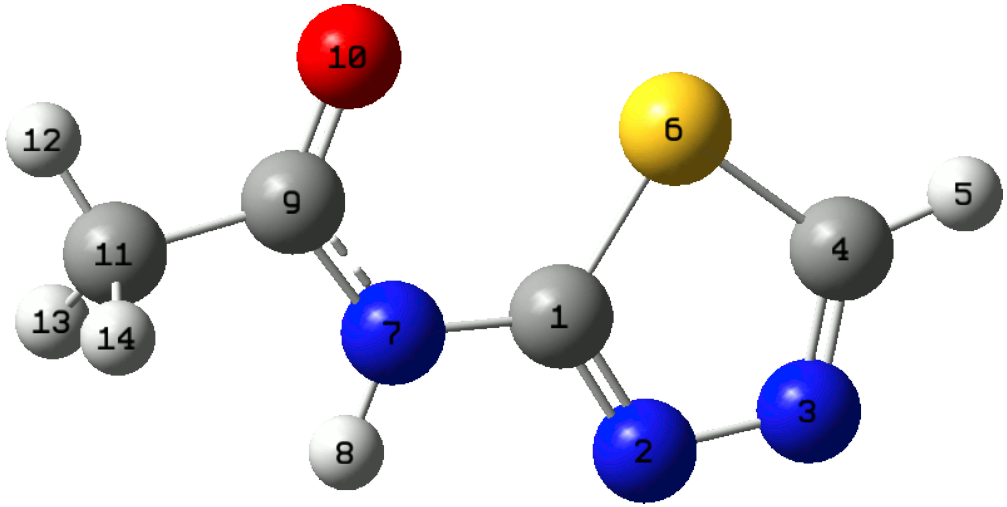


TDAZAM reference angles (2)

```
ANGLES
NG2R50 CG2R53 NG2S1    68.00    129.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   68.00    113.70 ! TDAZAM
```

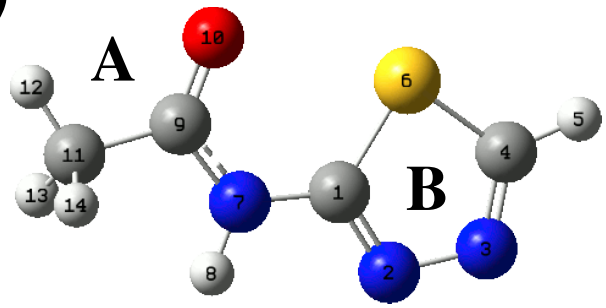
```
ANGLES
NG2R50 CG2R53 NG2S1    69.00    126.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   69.00    116.70 ! TDAZAM
```

Angle (2,1,7) changed 122.36º → 119.78º
(QM target = 119.75º)

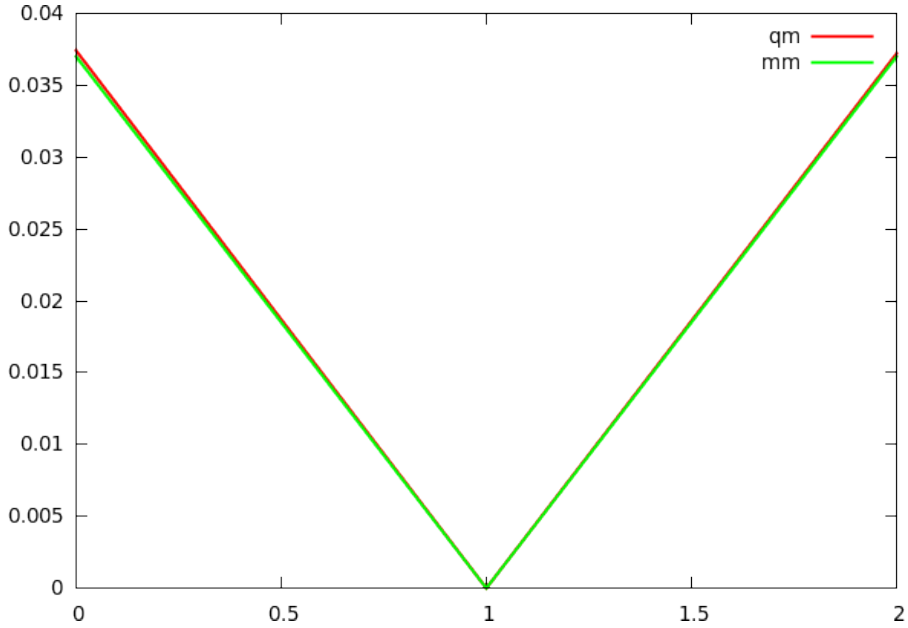


TDAZAM angle force constant: 3-point scan (2)

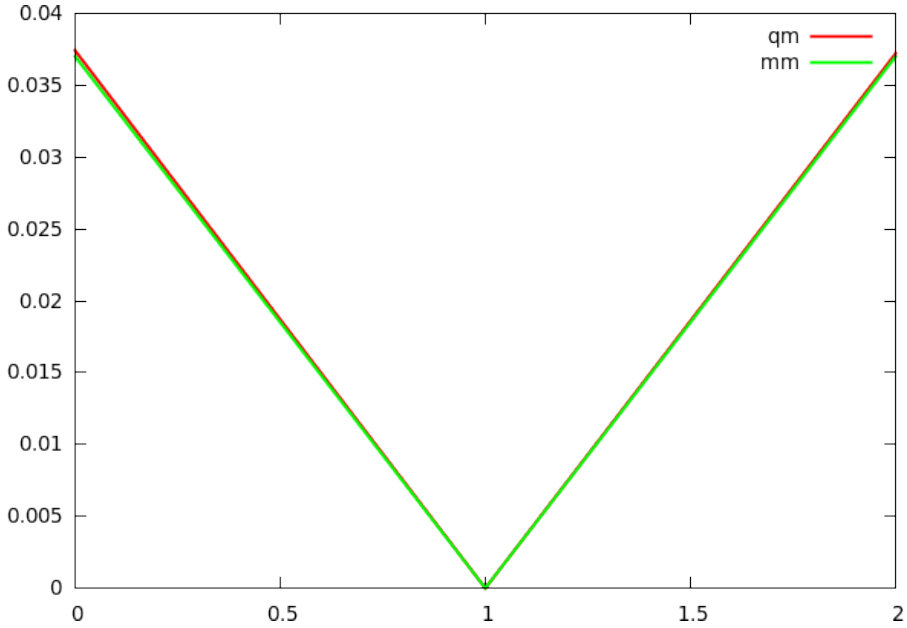
Self-consistency achieved




```
ANGLES
NG2R50 CG2R53 NG2S1    68.00    129.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   68.00    113.70 ! TDAZAM
```

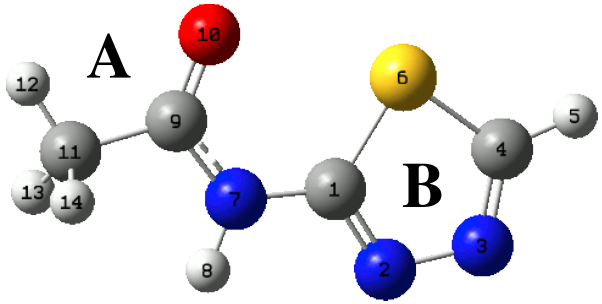



```
ANGLES
NG2R50 CG2R53 NG2S1    69.00    126.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   69.00    116.70 ! TDAZAM
```

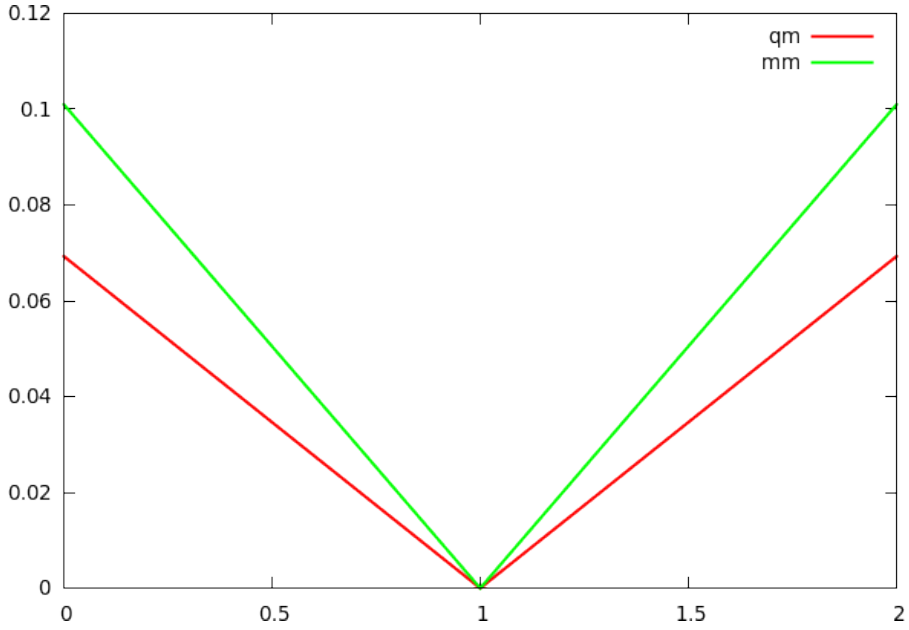


TDAZAM improper: 3-point scan

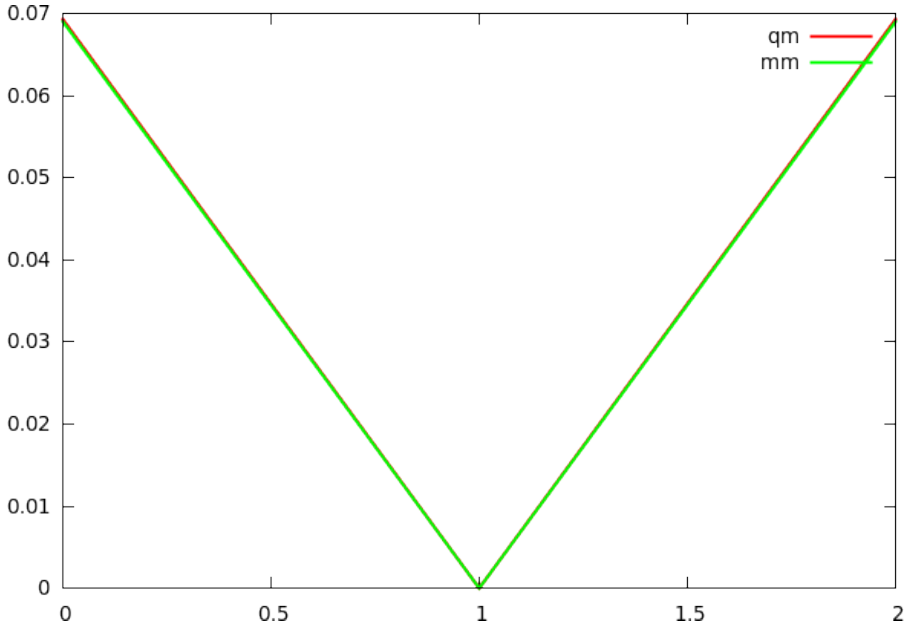
See fit_imp/allscan.inp



IMPROPERS
CG2R53 NG2R50 NG2S1 SG2R50 43.0000 0 0.00
! acazam , from CG2R53 NG2R53 OG2D1 SG311, penalty= 114

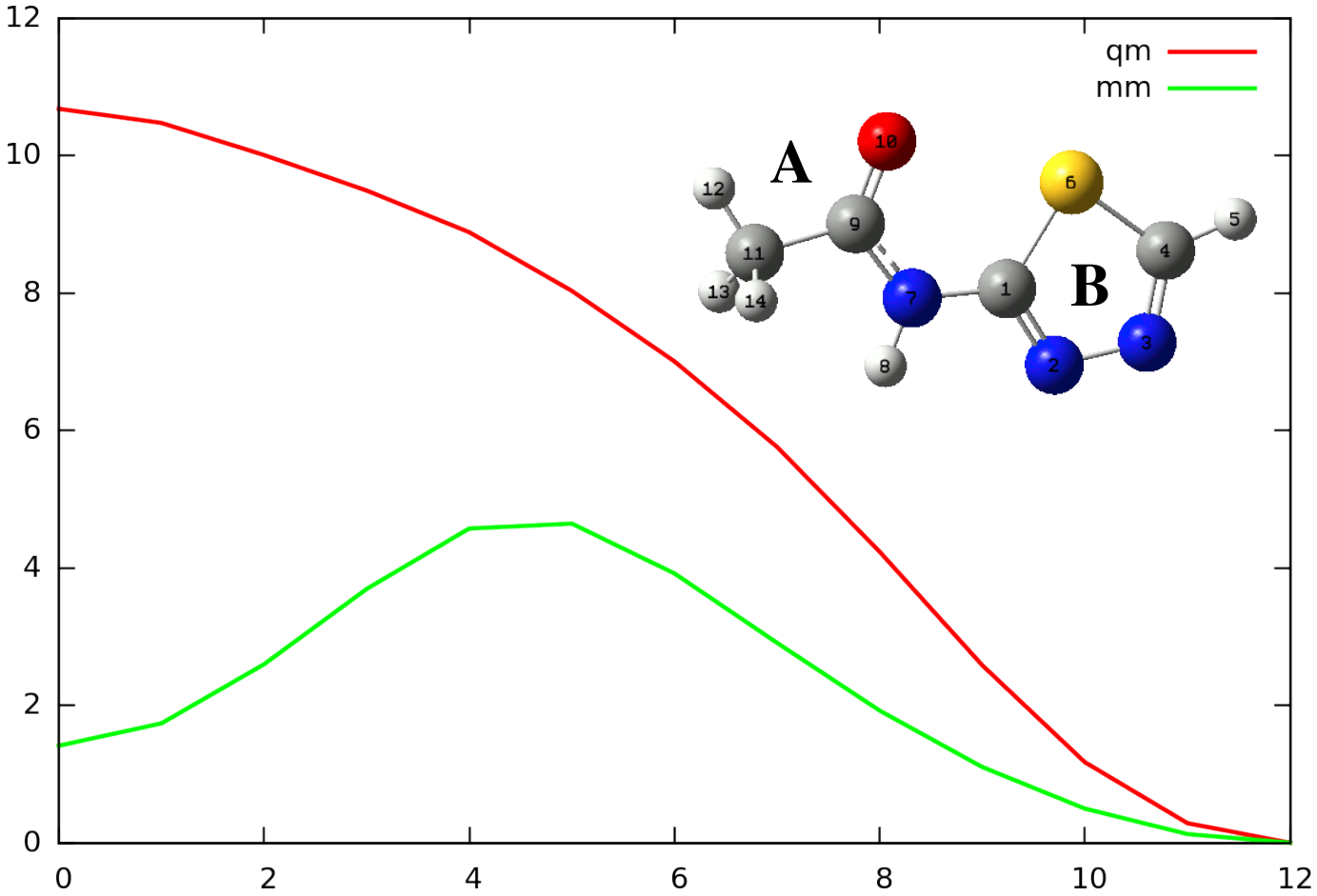


IMPROPERS
CG2R53 NG2R50 NG2S1 SG2R50 23.0000 0 0.00
! TDAZAM



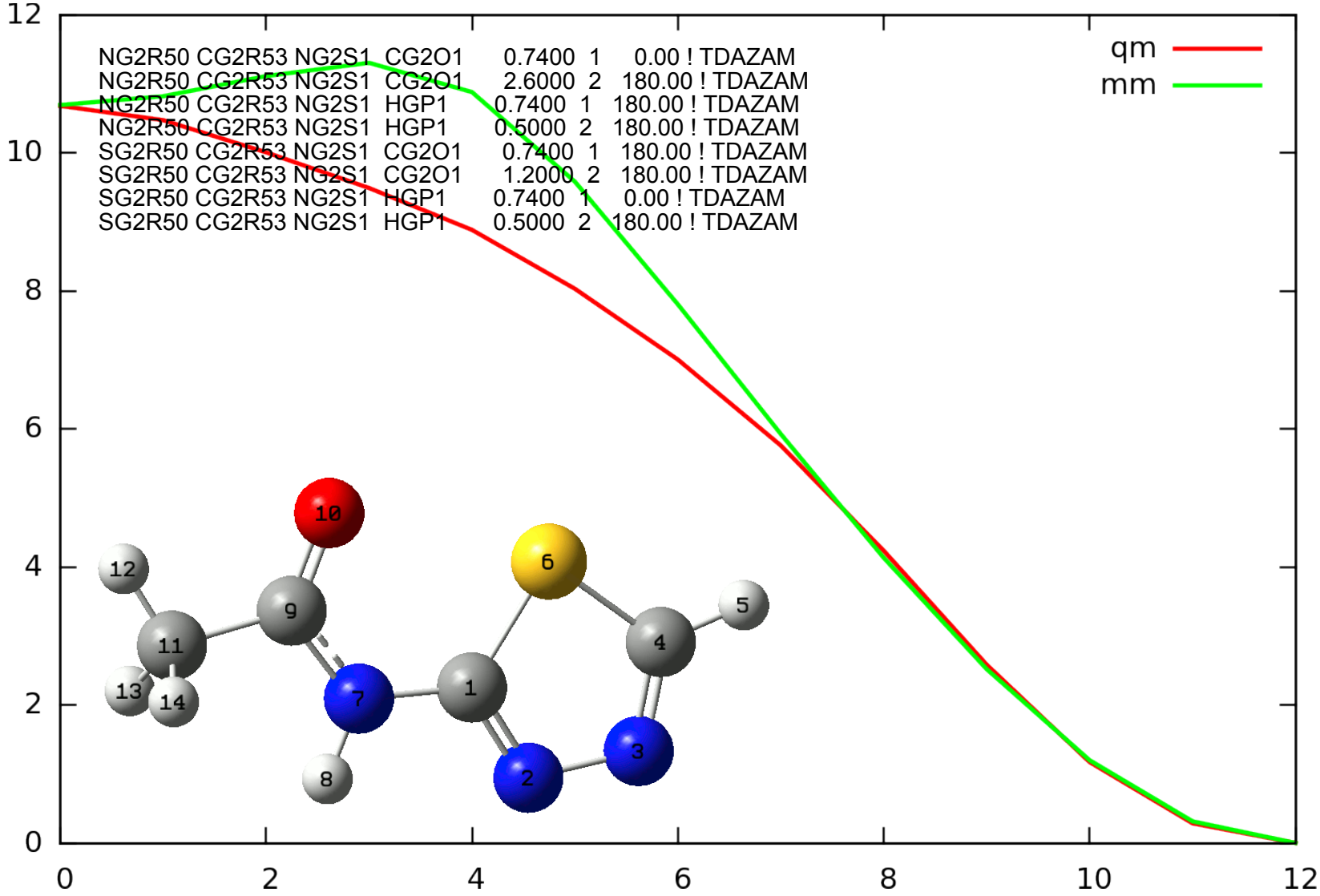
TDAZAM PES: before

```
NG2R50 CG2R53 NG2S1  CG2O1     1.5000  1   180.00 ! acazam , from NG2R60 CG2R64 NG2S1 CG2O1, penalty= 101
NG2R50 CG2R53 NG2S1  CG2O1     2.6000  2   180.00 ! acazam , from NG2R60 CG2R64 NG2S1 CG2O1, penalty= 101
NG2R50 CG2R53 NG2S1  CG2O1     0.1800  3   180.00 ! acazam , from NG2R60 CG2R64 NG2S1 CG2O1, penalty= 101
NG2R50 CG2R53 NG2S1  HGP1      0.5000  2   180.00 ! acazam , from NG2R60 CG2R64 NG2S1 HGP1, penalty= 101
SG2R50 CG2R53 NG2S1  CG2O1     1.2000  2   180.00 ! acazam , from CG2R61 CG2R61 NG2S1 CG2O1, penalty= 209
SG2R50 CG2R53 NG2S1  HGP1      0.5000  2   180.00 ! acazam , from CG2R61 CG2R61 NG2S1 HGP1, penalty= 209
```



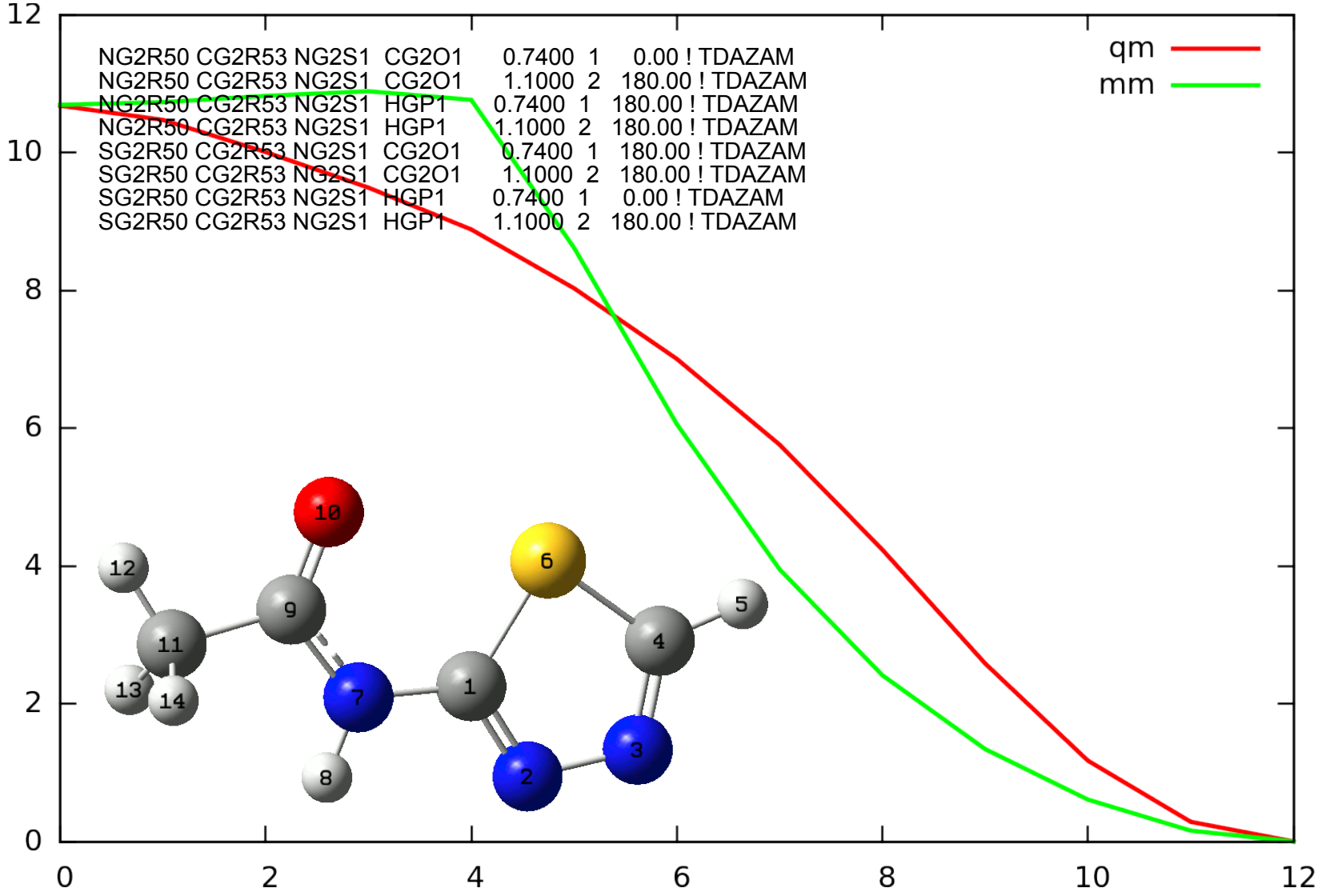
Most obvious deficiency: 1-fold, See fit_dih/allscan.inp

TDAZAM PES: improved 1-fold



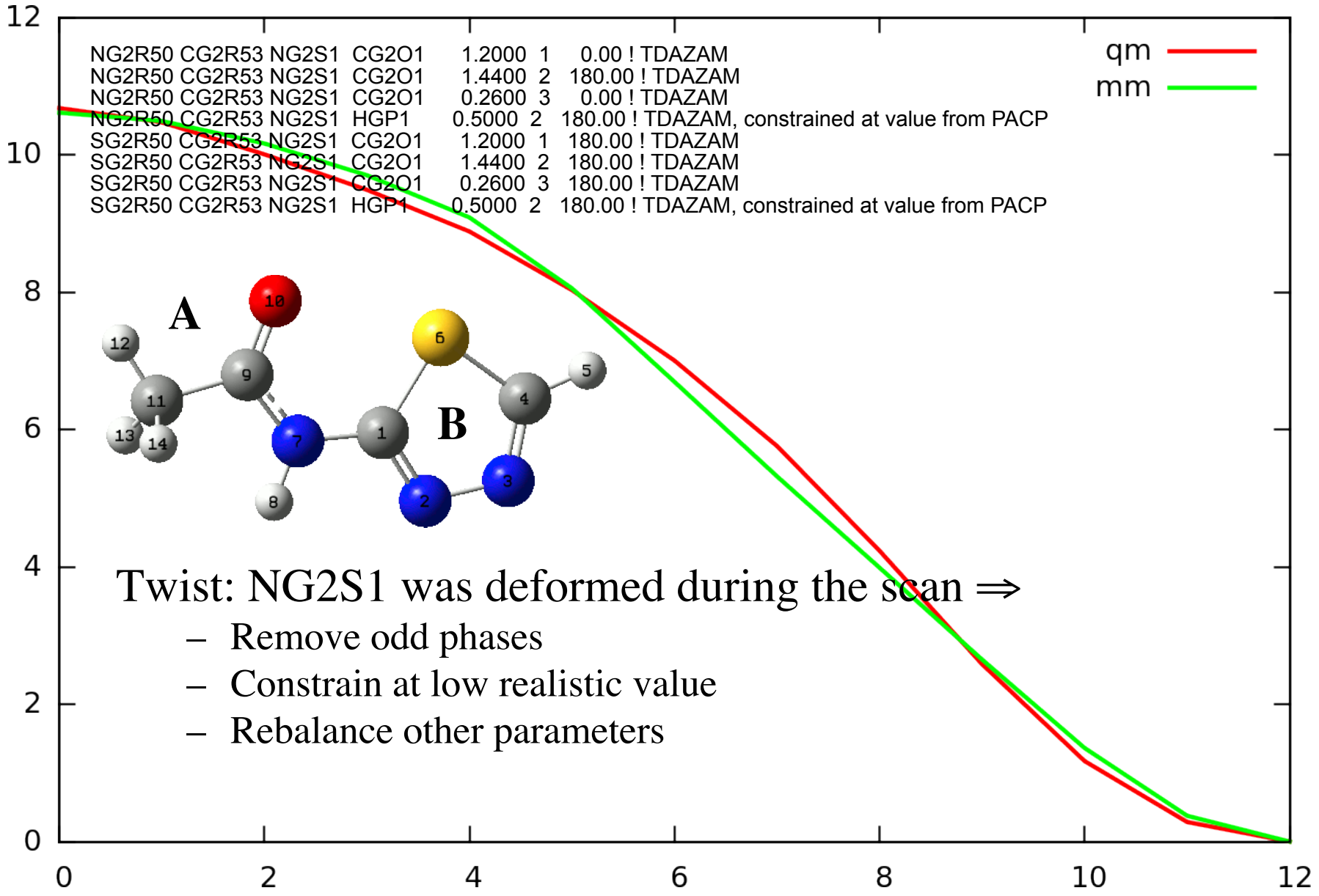
Most obvious deficiency: 2-fold, See fit_dih/allscan.inp

TDAZAM PES: improved 2-fold



Most obvious deficiency: 3-fold

TDAZAM PES: with 3-fold

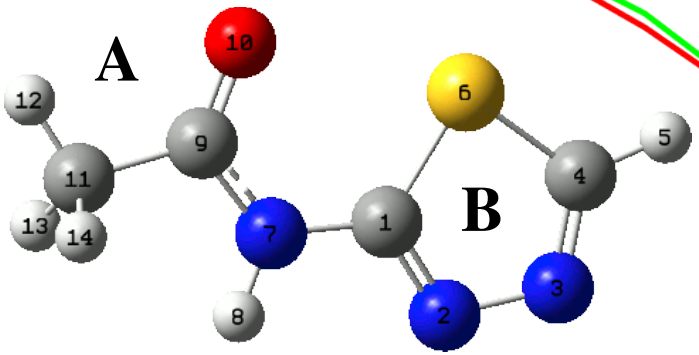


Twist: NG2S1 was deformed during the scan ⇒

- Remove odd phases
- Constrain at low realistic value
- Rebalance other parameters

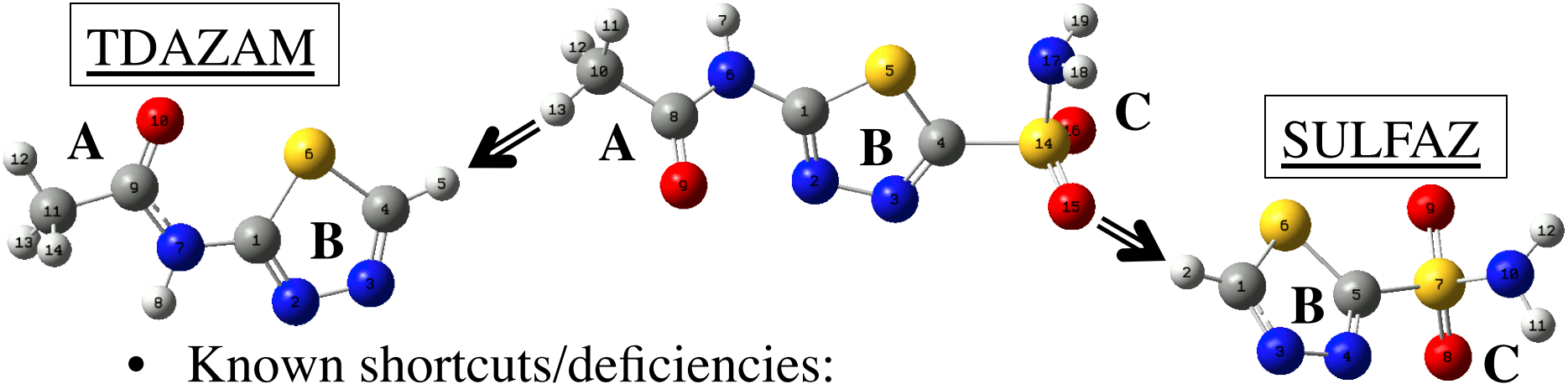
TDAZAM PES: with 4-fold

```
NG2R50 CG2R53 NG2S1  CG2O1     1.1700  1    0.00 ! TDAZAM
NG2R50 CG2R53 NG2S1  CG2O1     1.5100  2  180.00 ! TDAZAM
NG2R50 CG2R53 NG2S1  CG2O1     0.3300  3    0.00 ! TDAZAM
NG2R50 CG2R53 NG2S1  CG2O1     0.0800  4    0.00 ! TDAZAM
NG2R50 CG2R53 NG2S1  HGP1      0.5000  2  180.00 ! TDAZAM, constrained at value from PACP
SG2R50 CG2R53 NG2S1  CG2O1     1.1700  1  180.00 ! TDAZAM
SG2R50 CG2R53 NG2S1  CG2O1     1.5100  2  180.00 ! TDAZAM
SG2R50 CG2R53 NG2S1  CG2O1     0.3300  3  180.00 ! TDAZAM
SG2R50 CG2R53 NG2S1  CG2O1     0.0800  4    0.00 ! TDAZAM
SG2R50 CG2R53 NG2S1  HGP1      0.5000  2  180.00 ! TDAZAM, constrained at value from PACP
```



- MCSA fitting re-introduced N deformation and therefore failed to find good parameters
- Above parameters: complicated combination of MCSA fitting and manual tweaking in order to avoid this
- Worth the effort???

Full acetazolamide model (3)

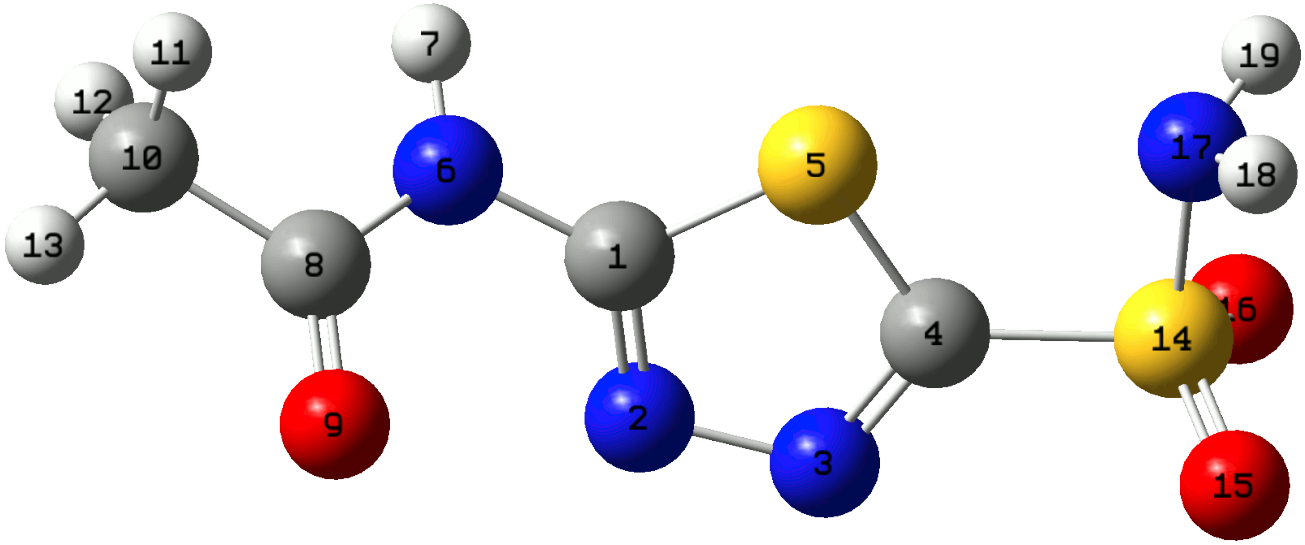


- Known shortcuts/deficiencies:
  - Combining charges for linkages ambiguous (water interactions on TDAZAM & SULFAZ?)
  - Ring substituent dihedrals (MOLVIB or PES on TDAZAM & SULFAZ ???)
  - Linkage bond parameters (crystal surveys and/or 3-point scans!)
  - In principle requires refitting everything (espcially diedrals) to maintain self-consistency

- Final validation against relevant experimental data! (the proof is in the pudding)

Full acetazolamide model(1)

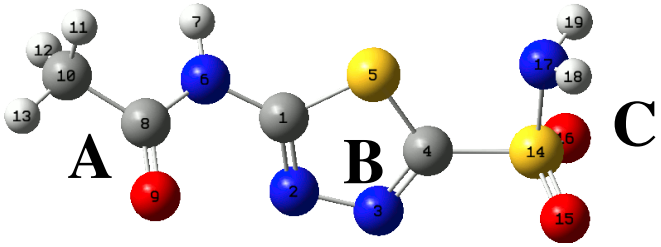
```
RESI ACAZAM        0.00 ! acetazolamide, N-(5-sulfamoyl-1,3,4-thiadiazol-2-yl)acetamide
GROUP
ATOM C1     CG2R53  0.71
ATOM N2     NG2R50 -0.48
ATOM N3     NG2R50 -0.48
ATOM C4     CG2R53  0.82
ATOM S5     SG2R50 -0.26
ATOM N6     NG2S1  -0.43
ATOM H6     HGP1    0.33
ATOM C7     CG2O1   0.52
ATOM O7     OG2D1  -0.52
ATOM C8     CG331  -0.27
ATOM H81    HGA3    0.09
ATOM H82    HGA3    0.09
ATOM H83    HGA3    0.09
ATOM S9     SG3O2   0.64
ATOM O91    OG2P1  -0.42
ATOM O92    OG2P1  -0.42
ATOM N10    NG321  -0.77
ATOM H101   HGP1    0.38
ATOM H102   HGP1    0.38

BOND ...
IMPR C1     N2     N6     S5
IMPR C7     C8     N6     O7

...

BONDS
CG2R53 NG2S1   305.00     1.4140 ! acazam , from CG2R61 NG2S1, penalty= 40
CG2R53 SG3O2   190.00     1.7300 ! acazam , from CG2R61 SG3O2, penalty= 40

ANGLES
CG2R53 NG2R50 NG2R50   80.00    103.65 ! TDAZ
CG2R53 SG2R50 CG2R53   73.00     98.30 ! TDAZ
NG2R50 CG2R53 SG3O2    30.00    119.50 ! SULFAZ
SG2R50 CG2R53 SG3O2    30.00    123.30 ! SULFAZ
NG2R50 CG2R53 NG2S1    69.00    126.10 ! TDAZAM
NG2S1  CG2R53 SG2R50   69.00    116.70 ! TDAZAM
CG2O1  NG2S1  CG2R53   50.00    120.00 ! acazam , from CG2O1 NG2S1 CG2R61, penalty= 8.5
CG2R53 NG2S1  HGP1     34.00    117.00 ! acazam , from CG2R61 NG2S1 HGP1, penalty= 8.5
CG2R53 SG3O2  NG321    60.00     98.00 ! acazam , from CG2R61 SG3O2 NG321, penalty= 8.5
CG2R53 SG3O2  OG2P1    60.00    101.00 ! acazam , from CG2R61 SG3O2 OG2P1, penalty= 8.5
```



Full acetazolamide model (2)

DIHEDRALS
SG2R50 CG2R53 NG2R50 NG2R50 10.9000 2 180.00 ! TDAZ
NG2R50 CG2R53 SG2R50 CG2R53 7.5000 2 180.00 ! TDAZ
CG2R53 NG2R50 NG2R50 CG2R53 4.4500 2 180.00 ! TDAZ
NG2S1 CG2R53 NG2R50 NG2R50 7.6300 2 180.00 ! ACAZAM & model compounds, average from TDAZ
NG2S1 CG2R53 SG2R50 CG2R53 5.9300 2 180.00 ! ACAZAM & model compounds, average from TDAZ
SG3O2 CG2R53 NG2R50 NG2R50 7.6300 2 180.00 ! ACAZAM & model compounds, average from TDAZ
SG3O2 CG2R53 SG2R50 CG2R53 5.9300 2 180.00 ! ACAZAM & model compounds, average from TDAZ
NG2R50 CG2R53 SG3O2 NG321 0.2600 1 0.00 ! SULFAZ
NG2R50 CG2R53 SG3O2 NG321 0.3500 2 0.00 ! SULFAZ
NG2R50 CG2R53 SG3O2 NG321 0.1000 3 0.00 ! SULFAZ
NG2R50 CG2R53 SG3O2 NG321 0.0300 4 180.00 ! SULFAZ
NG2R50 CG2R53 SG3O2 NG321 0.0200 6 0.00 ! SULFAZ
NG2R50 CG2R53 SG3O2 OG2P1 0.1000 3 0.00 ! SULFAZ
NG2R50 CG2R53 SG3O2 OG2P1 0.0200 6 0.00 ! SULFAZ
SG2R50 CG2R53 SG3O2 NG321 0.2600 1 180.00 ! SULFAZ
SG2R50 CG2R53 SG3O2 NG321 0.3500 2 0.00 ! SULFAZ
SG2R50 CG2R53 SG3O2 NG321 0.1000 3 180.00 ! SULFAZ
SG2R50 CG2R53 SG3O2 NG321 0.0300 4 180.00 ! SULFAZ
SG2R50 CG2R53 SG3O2 NG321 0.0200 6 0.00 ! SULFAZ
SG2R50 CG2R53 SG3O2 OG2P1 0.1000 3 180.00 ! SULFAZ
SG2R50 CG2R53 SG3O2 OG2P1 0.0200 6 0.00 ! SULFAZ
NG2R50 CG2R53 NG2S1 CG2O1 1.1700 1 0.00 ! TDAZAM
NG2R50 CG2R53 NG2S1 CG2O1 1.5100 2 180.00 ! TDAZAM
NG2R50 CG2R53 NG2S1 CG2O1 0.3300 3 0.00 ! TDAZAM
NG2R50 CG2R53 NG2S1 CG2O1 0.0800 4 0.00 ! TDAZAM
NG2R50 CG2R53 NG2S1 HGP1 0.5000 2 180.00 ! TDAZAM, constrained at value from PACP
SG2R50 CG2R53 NG2S1 CG2O1 1.1700 1 180.00 ! TDAZAM
SG2R50 CG2R53 NG2S1 CG2O1 1.5100 2 180.00 ! TDAZAM
SG2R50 CG2R53 NG2S1 CG2O1 0.3300 3 180.00 ! TDAZAM
SG2R50 CG2R53 NG2S1 CG2O1 0.0800 4 0.00 ! TDAZAM
SG2R50 CG2R53 NG2S1 HGP1 0.5000 2 180.00 ! TDAZAM, constrained at value from PACP
CG331 CG2O1 NG2S1 CG2R53 1.6000 1 0.00 ! acazam , from CG331 CG2O1 NG2S1 CG2R61, penalty= 8.5
CG331 CG2O1 NG2S1 CG2R53 2.5000 2 180.00 ! acazam , from CG331 CG2O1 NG2S1 CG2R61, penalty= 8.5
OG2D1 CG2O1 NG2S1 CG2R53 2.5000 2 180.00 ! acazam , from OG2D1 CG2O1 NG2S1 CG2R61, penalty= 8.5
HGP1 NG321 SG3O2 CG2R53 1.5000 1 180.00 ! acazam , from HGP1 NG321 SG3O2 CG2R61, penalty= 8.5
HGP1 NG321 SG3O2 CG2R53 1.2000 2 0.00 ! acazam , from HGP1 NG321 SG3O2 CG2R61, penalty= 8.5
HGP1 NG321 SG3O2 CG2R53 0.1000 3 0.00 ! acazam , from HGP1 NG321 SG3O2 CG2R61, penalty= 8.5

IMPROPERS
CG2R53 NG2R50 NG2S1 SG2R50 23.0000 0 0.00 ! TDAZAM



LJ (vdW) parameters

- Direct transfer from available parameters is generally adequate
- Test via
  - Heat of vaporization
  - Density (Molecular Volume)
  - Partial molar volume
  - Crystal simulations

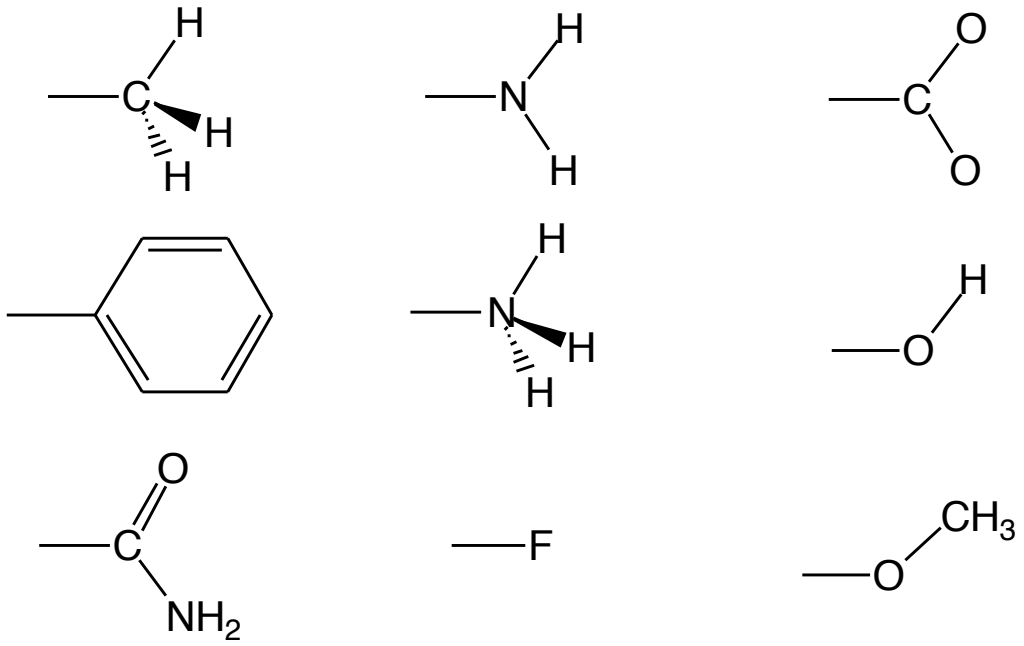
Experimental target data for bonded parameters

- Geometries (equilibrium bond, angle, dihedral and improper terms)
  - microwave, electron diffraction
  - small molecule x-ray crystallography (CSD)
  - crystal surveys of geometries
- Vibrational spectra (force constants)
  - infrared, Raman
- Conformational energies (force constants)
  - microwave

- For most drug molecules the amount of experimental data is minimal, requiring the use of QM data. However, for geometries it is often possible to do surveys of the Cambridge Structural Database for a type of linkage to obtain target geometries.

Lead Optimization

Addition of simple functional groups is generally straightforward once the full compound parameters have been optimized.



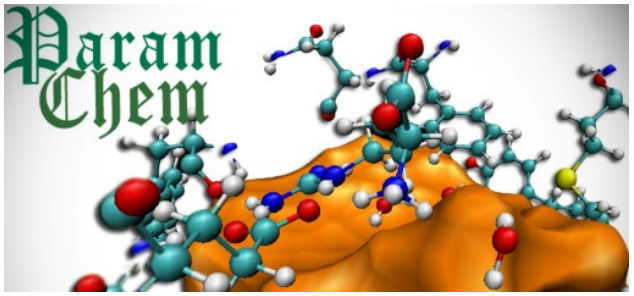
1) Delete appropriate hydrogens (i.e. at site of covalent bond)
2) Shift charge of deleted hydrogen into carbon being functionalized.
3) Add functional group
4) Offset charge on functionalized carbon to account for functional group charge requirements
    1) Aliphatics: just neutralize added functional group, q_H=0.09
    2) Phenol OH: q_C=0.11, q_O=-0.54, q_H=0.43
    3) Aliphatic OH: q_C=-0.04, q_O=-0.66, q_H=0.43
    4) Amino: q_C=0.16, q_{CH}=0.05, q_N=-0.30, q_H=0.33
    5) Carboxylate: q_C=-0.37, q_{CO}=-0.62, q_O=-0.76
5) Internal parameters should be present. Add by analogy if needed.
6) Optimize necessary parameters.

Perform above via the CHARMM PATCH (PRES) command

Summary

1) Junk in, junk out: Parameter optimization effort based on application requirements.
2) Follow standard protocol for the force field of interest (higher level QM is not necessarily better).
3) Careful parameter optimization of lead molecules.
4) Simple substitutions often require minimal or no optimization.

ParamChem



https://www.paramchem.org/index.htm

ParamChem Force Field Engine: Web based utility to automatically generate topology and additional parameters to model drug like molecules in the context of CGenFF (and other FFs in the future).

1)Automatic atom typing, partial charge assignment and parameter guess (with scores!) based on mol2 input format.

2)“Automatic” parameter validation via reproduction of QM potential energy surfaces etc.

3)“Automatic” parameter optimization via reproduction of QM potential energy surfaces etc.

Acknowledgements

MacKerell lab members
(Kenno Vanommeslaeghe)
ParamChem Team
Science Gateways group, IU
NIH, NSF, DoD HPC, NPACI, PSC Terascale Computing