Overview of Classical Force Fields and Parameter Optimization Strategy

Part I - Overview of CHARMM FF and Parameter Optimization

Part II - Introduction to Quantum Chemistry Calculations (SPARTAN)
Application to parameterization of thioester

Part III - Knowledge-based and Hybrid FF
Application to protein structure prediction and folding studies

Perth 2004

Classical Force Fields

- ✓ Physics-based, full atom FF like CHARMM, AMBER, OPLS - Mechanisms, function, protein/RNA/DNA interactions…
- Knowledge-based, coarse-grained model -

  Protein structure prediction, folding kinetics, docking,…
- Hybrid force fields like Go + full atom FF -

  Mutational effects on protein folding kinetics,….
- Hybrid QM/MM approaches

  Enzyme reactions, bond breaking and making, excited states,…

Each problem has a different goal and time scale!

General Considerations

- Description of molecules?
- Optimization of force field parameters?
- Training set of compounds/data?
- Test set of compounds/data?
- Limitations – questions you should not ask of your force field

Common Full Atom Force Fields

Class I

CHARMM, CHARMm (Accelyrs),AMBER, OPLS
ECEPP, GROMOS, SYBYL(Tripos)

Class II

MMFF94, UFF, …

Class III

QM/MM (CHARMM, AMBER,…)
Polarizable FF - Freisner/Berne(Schroedinger), AMOEBA (Tinker)

*Websites contain roadmaps of force field parameterization strategy. And they are different!!! So parameters from one cannot usually be used in another force field.

Overview and parameter optimization of CHARMM Force Field

Based on protocol established by

Alexander D. MacKerell, Jr , U. Maryland

See references: www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm

Especially Sanibel Conference 2003, JCC v21, 86,105 (2000)

Class I Potential Energy function

$$E_{Total} = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2$$

$$+ \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \delta)]$$

$$+ \sum_{impropers} k_\omega (\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u (r_{1,3} - r_{1,3,0})^2$$

Non-bonded Interaction Terms

$$+ \sum_{electrostatics} \left(\frac{q_i q_j}{\varepsilon r_{ij}}\right) + \sum_{VDW} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

From MacKerell

Class II force fields (e.g. MMFF) – Transferability, organic comb.

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

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

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

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

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

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

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

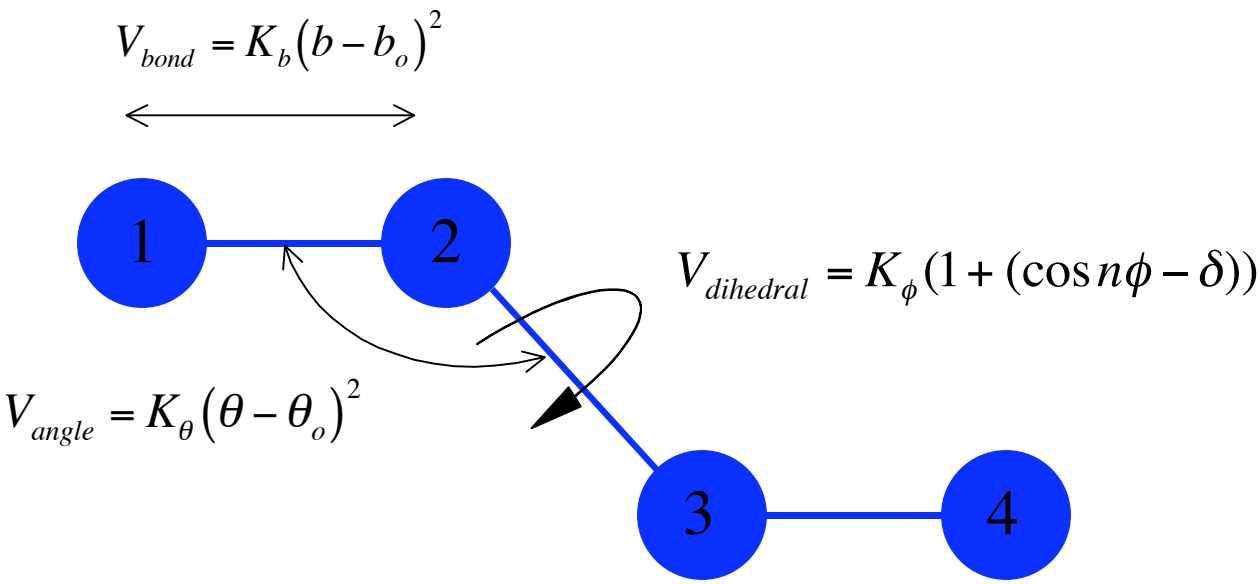
$$+\sum_{bonds'}\sum_{dihedrals}(b'-b_o')\left[K_{\phi,b'1}\cos\phi+K_{\phi,b'2}\cos 2\phi+K_{\phi,b'3}\cos 3\phi\right]$$

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

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

From MacKerell

Interactions between bonded atoms



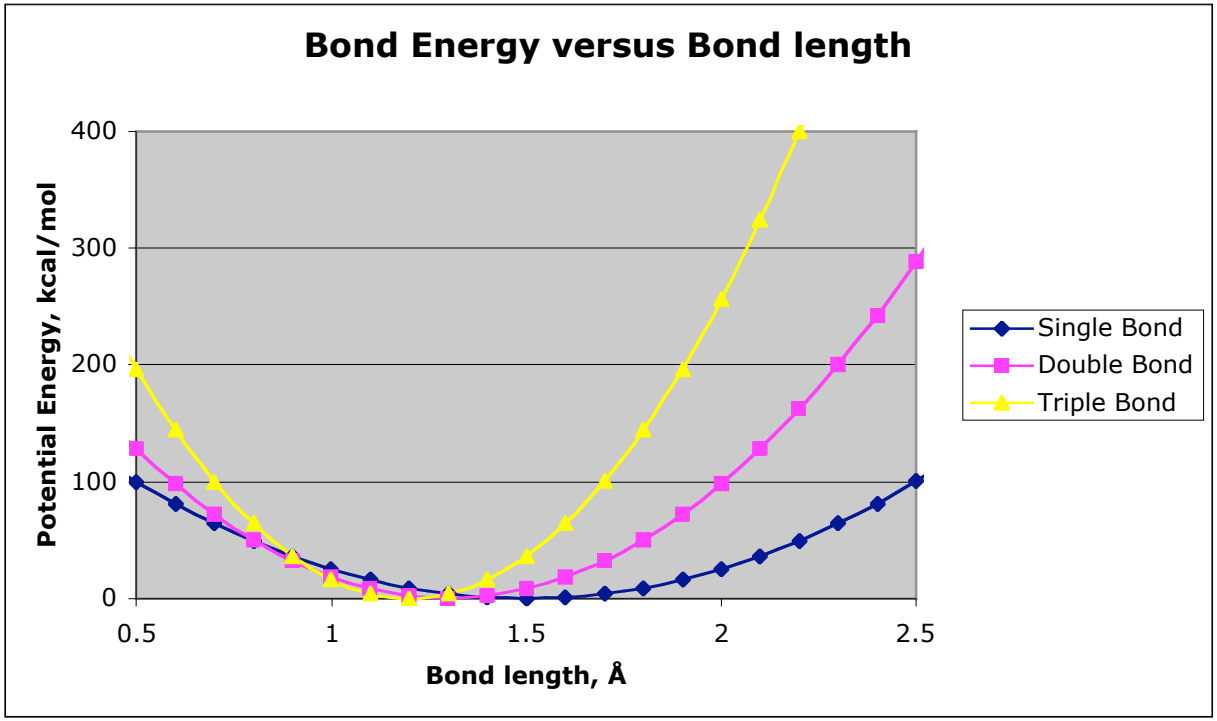
Intermolecular 1,4 interactions: 1 or scaled

> 1,4 interactions: 1

From MacKerell

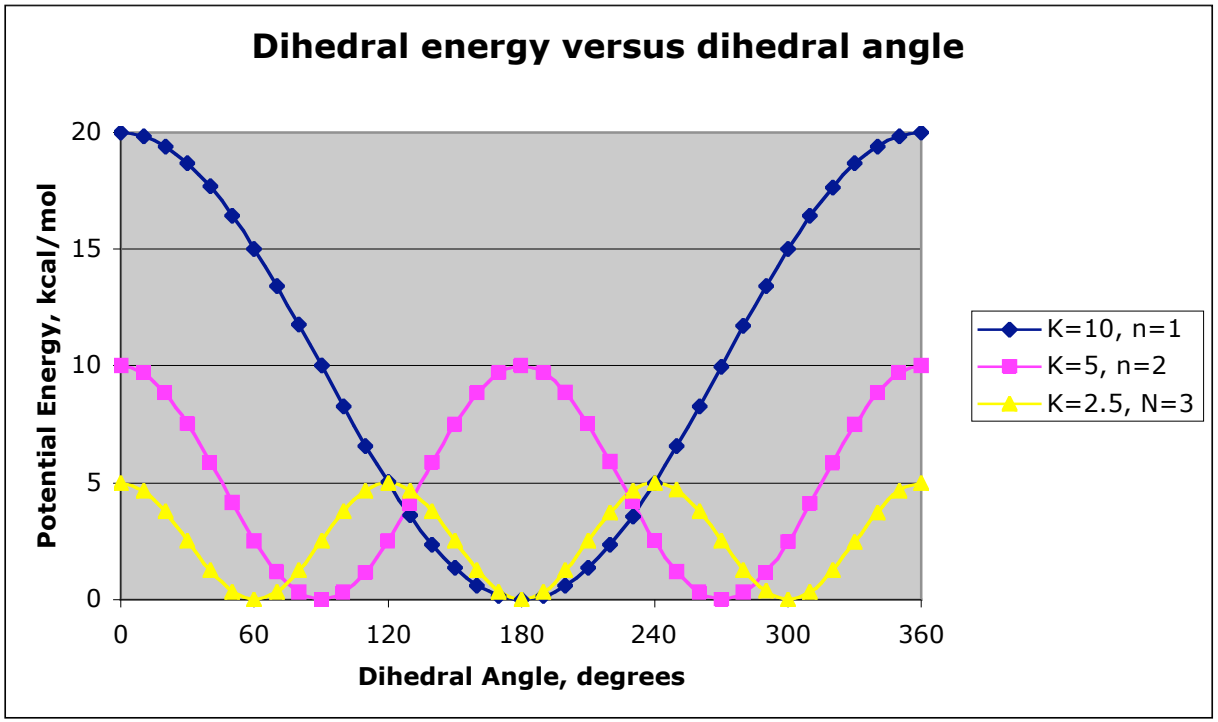
$$V_{bond} = K_b (b - b_o)^2$$

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



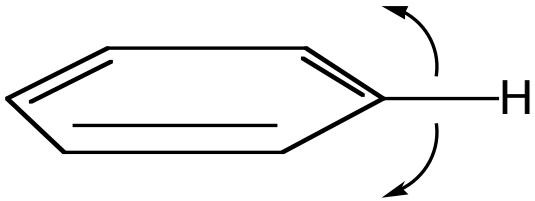
From MacKerell

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

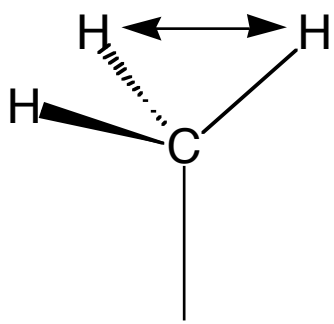


$$\delta = 0^{\circ}$$

From MacKerell



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



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

From MacKerell

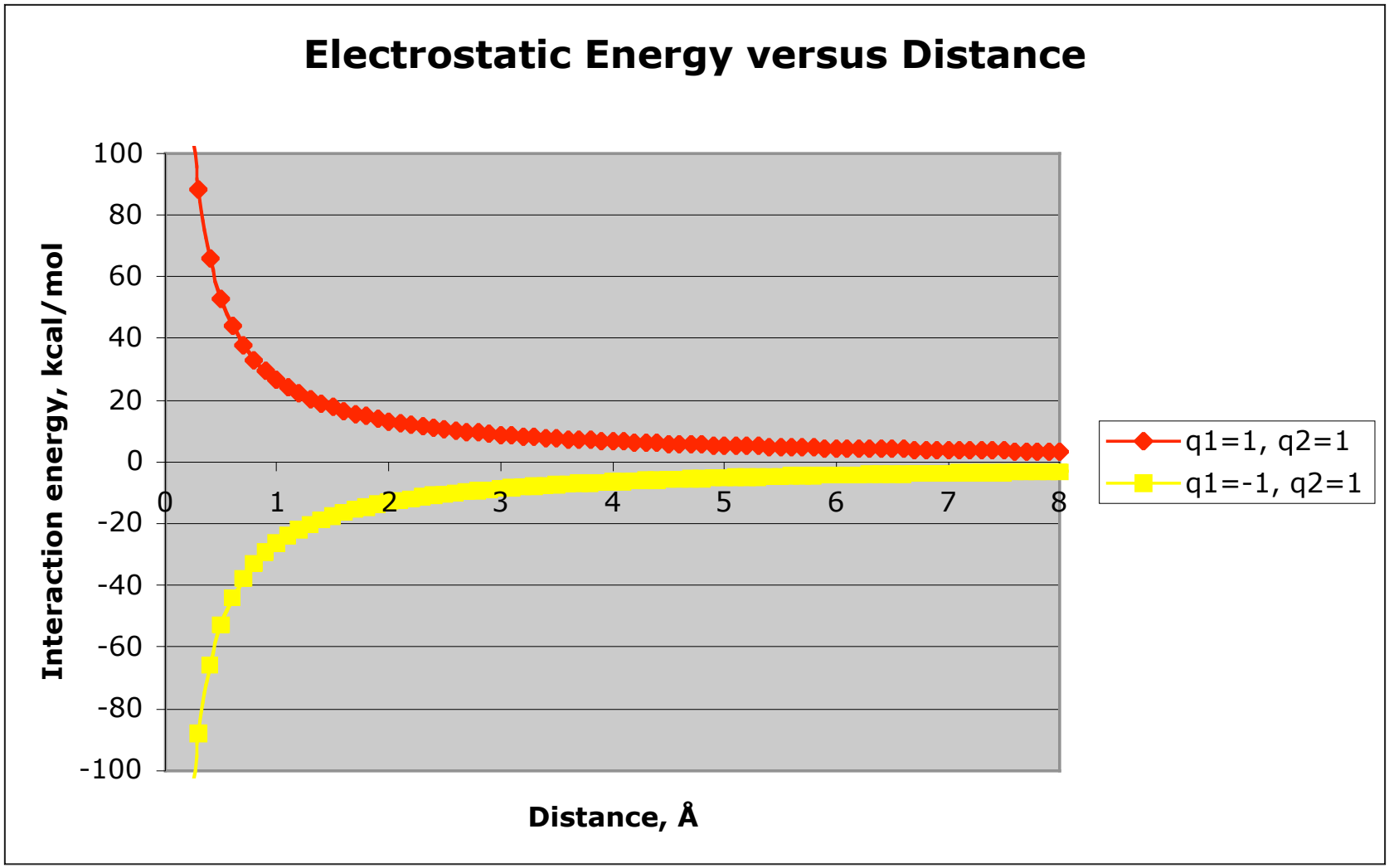
Intermolecular parameters

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

q_i: partial atomic charge
D: dielectric constant
ε: Lennard-Jones (LJ, vdW) well-depth
R_{min}: LJ radius ($R_{min}/2$ in CHARMM)
Combining rules (CHARMM, Amber)

$R_{min\ i,j} = R_{min\ i} + R_{min\ j}$
$\varepsilon_{i,j} = SQRT(\varepsilon_i * \varepsilon_j)$

From MacKerell

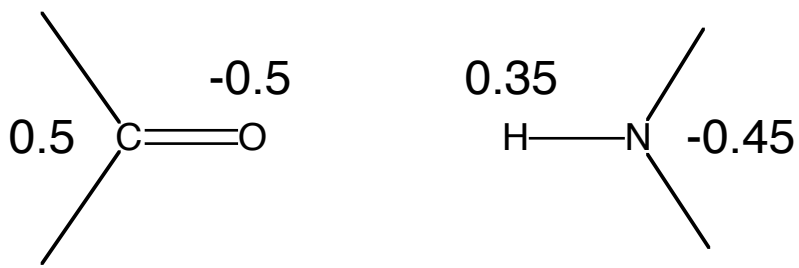


From MacKerell

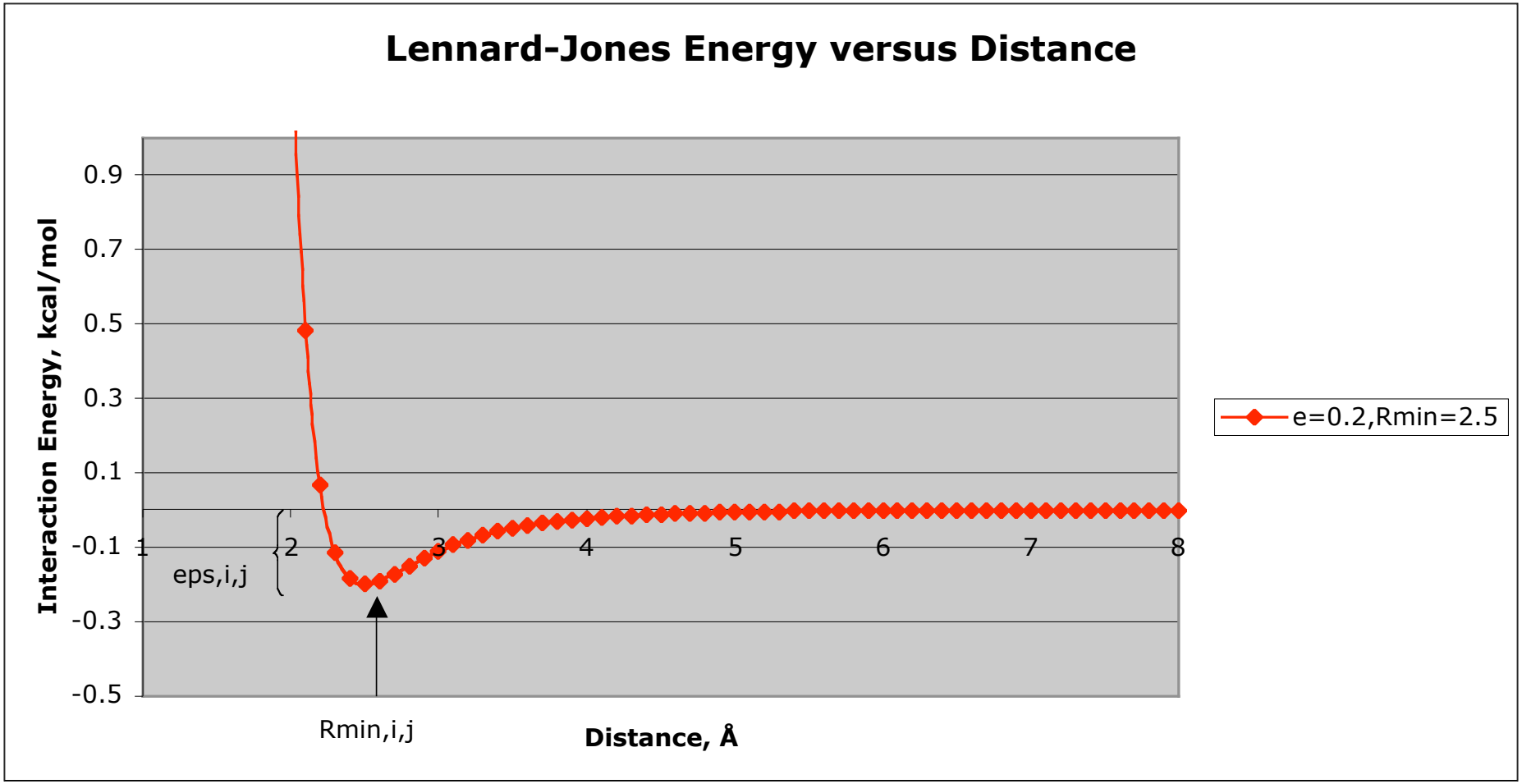
Charge Fitting Strategy

CHARMM- Mulliken* AMBER(ESP/RESP)

Partial atomic charges



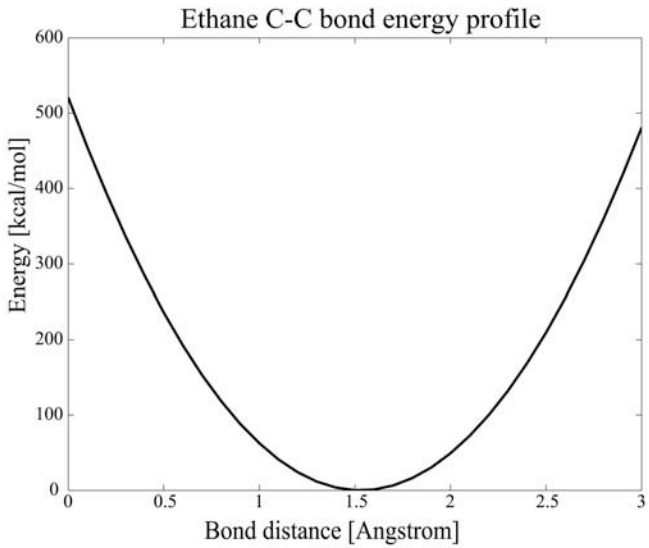
*Modifications based on interactions with TIP3 water

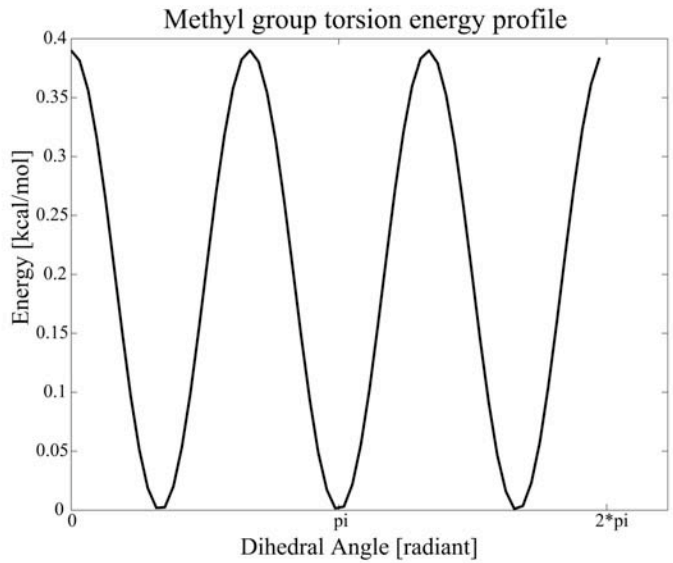


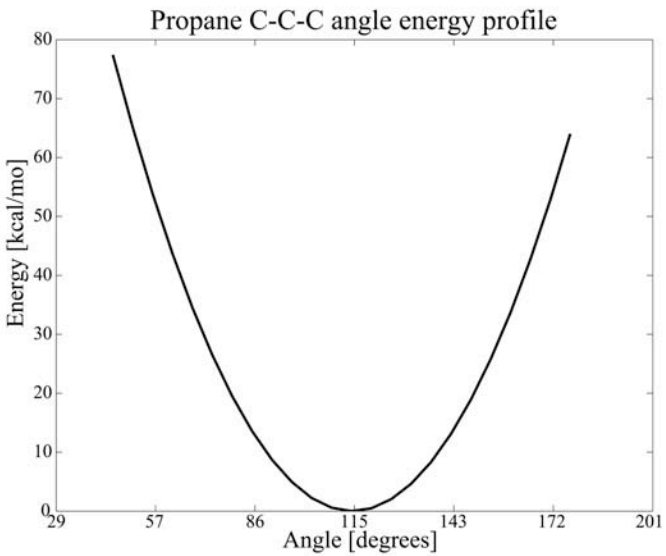
$$\varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

From MacKerell

Summary of Potential Terms

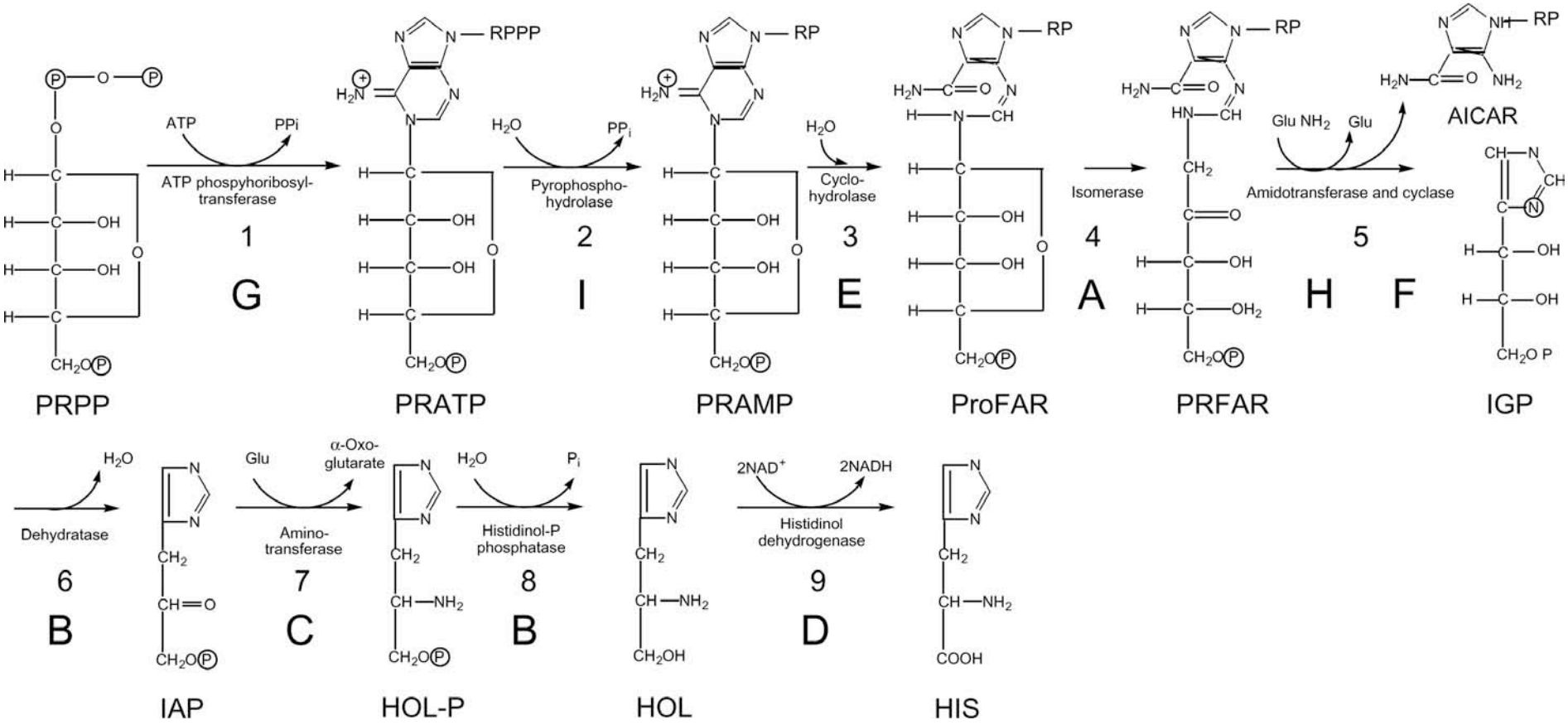






What about a new ligand not present in the CHARMM force field??????

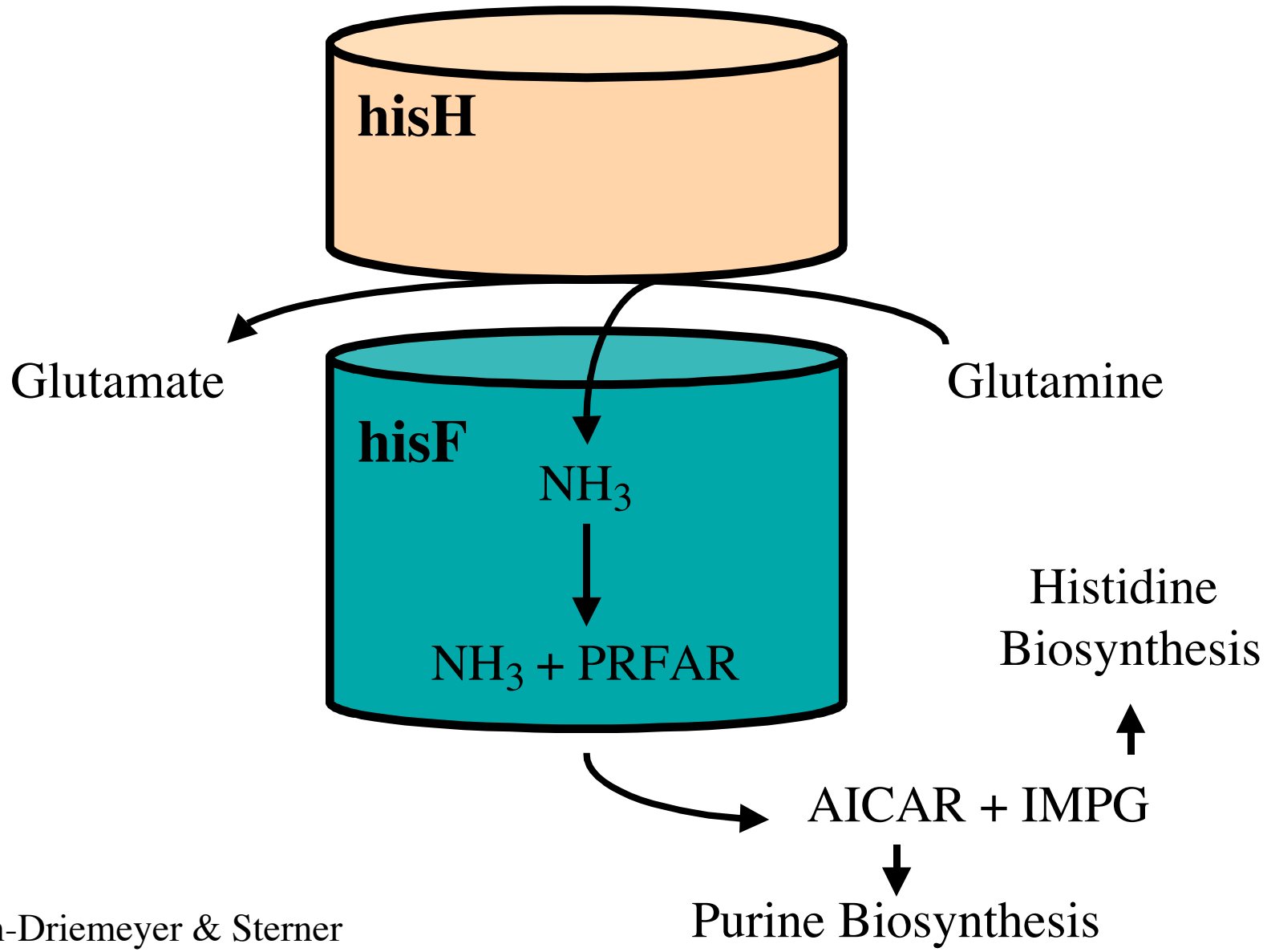
Histidine Anabolic Pathway



- **Imidazole Glycerol Phosphate Synthase**: 5th step in Histidine Biosynthesis.
- Branch point between Purine (nucleotide) and Histidine synthesis.

Imidazole Glycerol Phosphate Synthase:

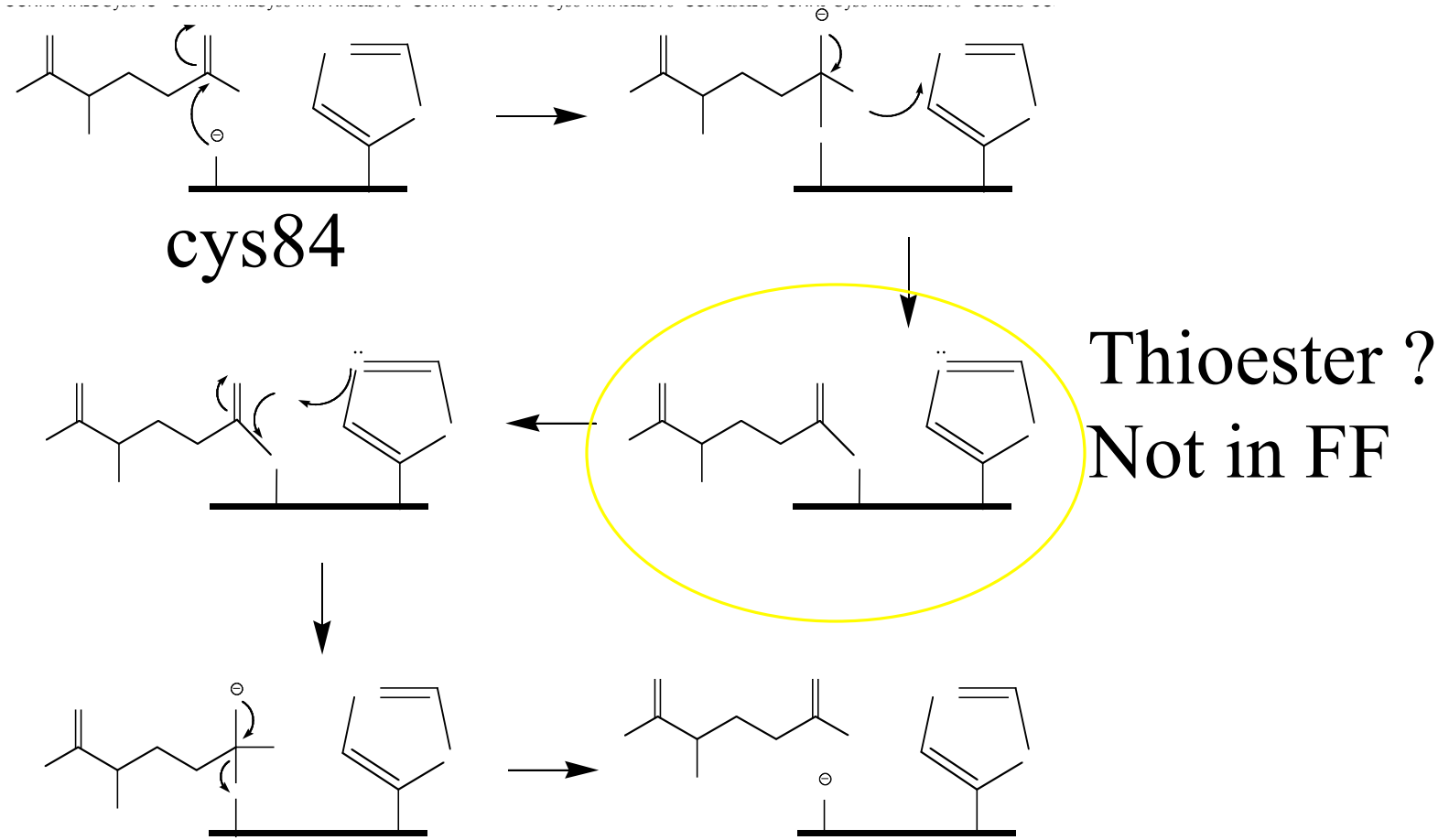
Proposed Mechanism*



*Beismann-Driemeyer & Sterner

HisH Mechanism

- HisH glutamine amidotransferase
- Conserved catalytic triad: CYS84, HIS178, GLU180



Parameter Optimization Strategies

Minimal optimization

By analogy (i.e. direct transfer of known parameters)
Quick, starting point

Maximal optimization

Time-consuming
Requires appropriate target data

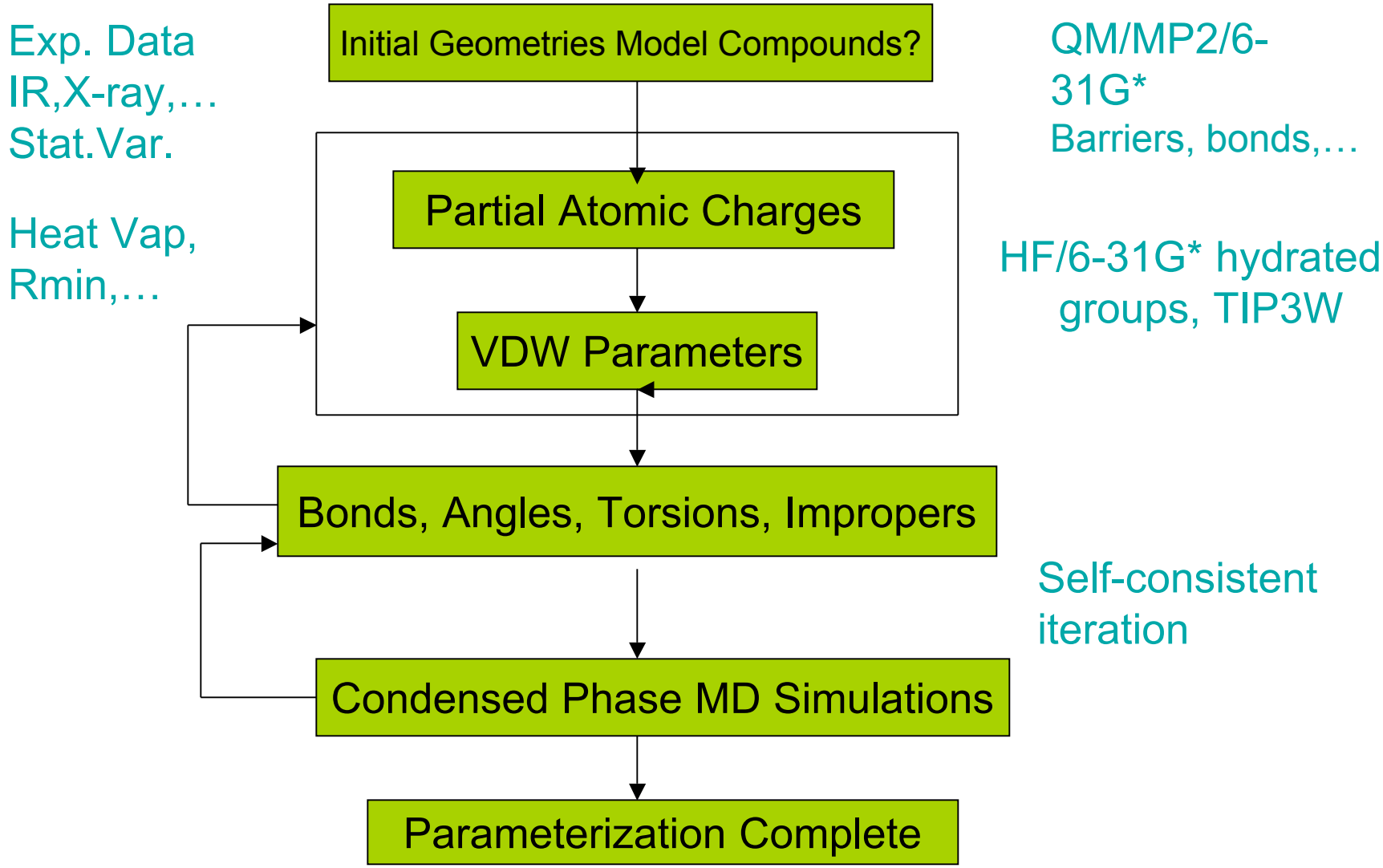
Choice based on goal of the calculations

Minimal
- database screening
- NMR/X-ray structure determination

Maximal
- free energy calculations, mechanistic studies,
- subtle environmental effects

Manual or Automated Fitting Procedures ?

Roadmap Charmm27 Optimization*



*based on MacKerell, JCC v21, 86,105 (2000)

Getting Started

- Identify previously parameterized compounds
- Access topology information – assign atom types, connectivity, and charges – annotate changes

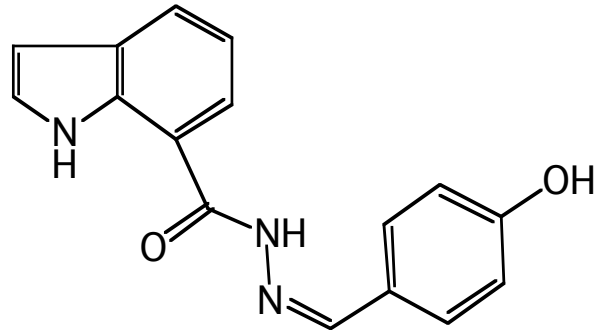
CHARMM topology (parameter files)

top_all22_model.inp (par_all22_prot.inp)
top_all22_prot.inp (par_all22_prot.inp)
top_all22_sugar.inp (par_all22_sugar.inp)
top_all27_lipid.rtf (par_all27_lipid.prm)
top_all27_na.rtf (par_all27_na.prm)
top_all27_na_lipid.rtf (par_all27_na_lipid.prm)
top_all27_prot_lipid.rtf (par_all27_prot_lipid.prm)
top_all27_prot_na.rtf (par_all27_prot_na.prm)
toph19.inp (param19.inp)

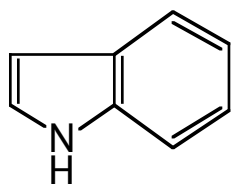
NA and lipid force fields have new LJ parameters for the alkanes, representing increased optimization of the protein alkane parameters. Tests have shown that these are compatible (e.g. in protein-nucleic acid simulations). For new systems is suggested that the new LJ parameters be used. Note that only the LJ parameters were changed; the internal parameters are identical

www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm

Break Desired Compound into 3 Smaller Ones

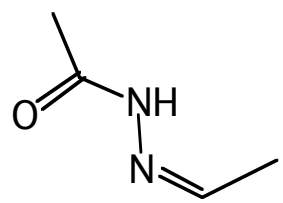


A



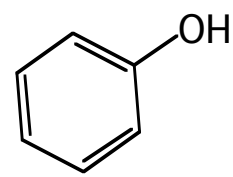
Indole

B



Hydrazine

C

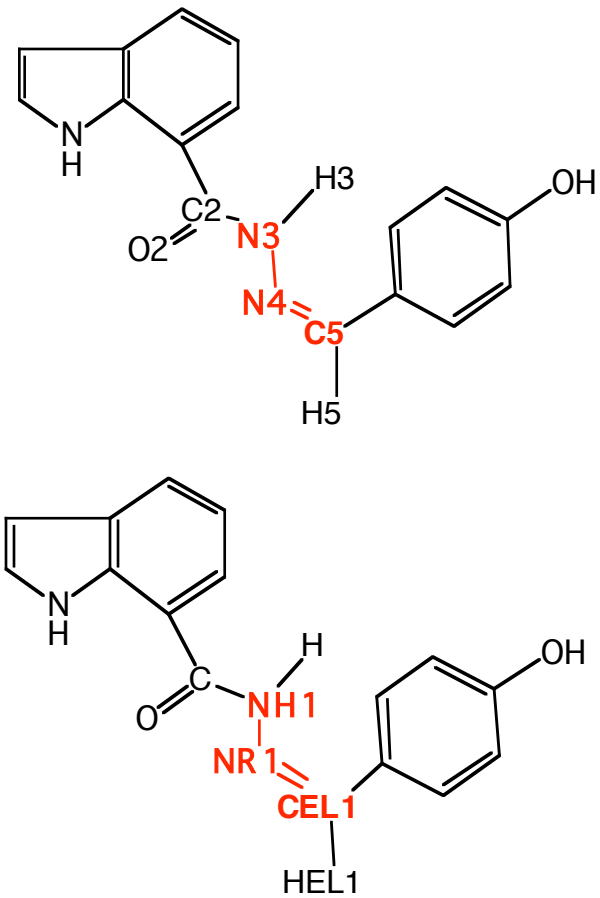


Phenol

When creating a covalent link between model compounds move the charge on the deleted H into the carbon to maintain integer charge (i.e. methyl (q_C=-0.27, q_H=0.09) to methylene (q_C=-0.18, q_H=0.09)

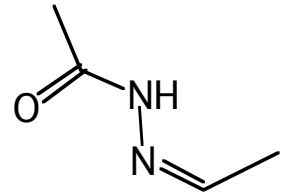
From MacKerell

Comparison of atom names (upper) and atom types (lower)



From MacKerell

Creation of topology for central model compound




```
Resi  Mod1 ! Model compound 1
Group        !specifies integer charge group of atoms (not
essential)
ATOM C1    CT3   -0.27
ATOM H11  HA3    0.09
ATOM H12  HA3    0.09
ATOM H13  HA3    0.09
GROUP
ATOM C2   C      0.51
ATOM O2   O     -0.51
GROUP
ATOM N3   NH1   -0.47
ATOM H3   H      0.31
ATOM N4   NR1    0.16   !new atom
ATOM C5   CEL1  -0.15
ATOM H51 HEL1   0.15
ATOM C6    CT3   -0.27
ATOM H61  HA     0.09
ATOM H62  HA     0.09
ATOM H63  HA     0.09
BOND C1 H11 C1 H12 C1 H13 C1 C2 C2 O2 C2 N3 N3 H3
BOND N3 N4 C5 H51 C5 C6 C6 H61 C6 H62 C6 H63
DOUBLE N4 C5 (DOUBLE only required for MMFF)
```

Start with alanine dipeptide.
Note use of new aliphatic LJ parameters and, importantly, atom types.

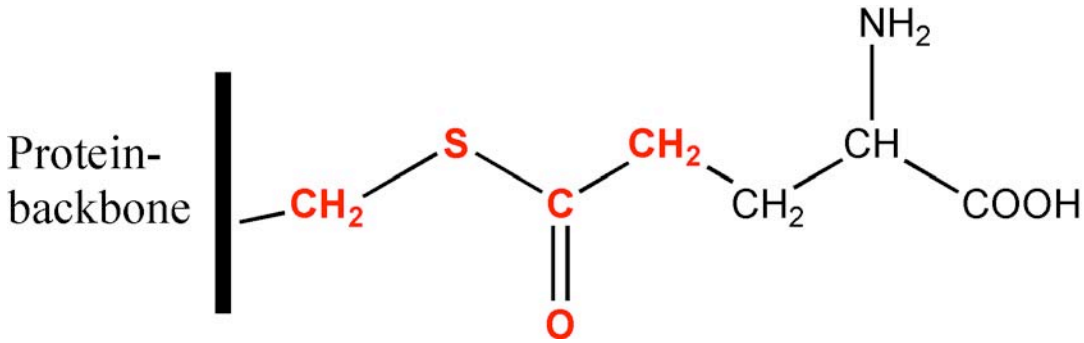
NR1 from histidine unprotonated ring nitrogen. Charge (very bad) initially set to yield unit charge for the group.

CEL1/HEL1 from propene (lipid model compound). See top_all27_prot_lipid.rtf

Note use of large group to allow flexibility in charge optimization.

From MacKerell

```
RESI CYG 0.00
GROUP
ATOM N     NH1     -0.47  !      |
ATOM HN    H        0.31  !  HN-N
ATOM CA    CT1      0.07  !      |   HB1
ATOM HA    HB       0.09  !      |   |
GROUP                     !  HA-CA--CB--SG
ATOM CB    CT2     -0.11  !      |   |   |
ATOM HB1   HA       0.09  !      |   HB2 |
ATOM HB2   HA       0.09  !   O=C       |
ATOM SG    S       -0.07  !      |        \
!ATOM HG1  HS       0.16  !                \
GROUP                     !                 \
ATOM CDG   CC       0.55  !                  \
ATOM OE1   O       -0.55  !                   \
GROUP                      !     HN2G          \
ATOM CGG   CT2     -0.18   !      |             \
ATOM HG1G  HA       0.09   ! HN1G-NG   HB1G HG1G\
ATOM HG2G  HA       0.09   !      |     |      |   \
GROUP                      !  HAG-CAG--CBG--CGG--CDG=OE1
ATOM CBG   CT2     -0.18   !      |     |      |
ATOM HB1G  HA       0.09   !      |    HB2G HG2G
ATOM HB2G  HA       0.09   !  O1G=CG
GROUP                      !      |
ATOM CG    CD       0.75   !      O2G-HO2G
ATOM O1G   OB      -0.55
ATOM O2G   OH1     -0.61
ATOM HO2G  H        0.44
ATOM CAG   CT1     -0.12
ATOM HAG   HB       0.09
ATOM NG    NH3      -0.62
ATOM HN1G  HC        0.31
ATOM HN2G HC       0.31
GROUP
ATOM C     C        0.51
ATOM O     O       -0.51
```



HG1 deleted from CYS and the charge was moved to SG (-0.23 +0.16=0.07) so that the SG charge becomes 0.07 in final compound and the group remains neutral

Changes annotated!

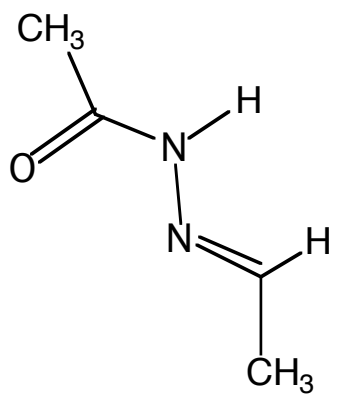
Partial Atomic Charge Determination

Additive Models: account for lack of explicit inclusion of polarizability via "overcharging" of atoms.

- RESP: HF/6-31G overestimates dipole moments (AMBER)
- Interaction based optimization (CHARMM, OPLS)
  - local polarization included
  - scale target interaction energies (CHARMM)
    - 1.16 for polar neutral compounds
    - 1.0 for charged compounds

For a particular force field do NOT change the QM level of theory. This is necessary to maintain consistency with the remainder of the force field.

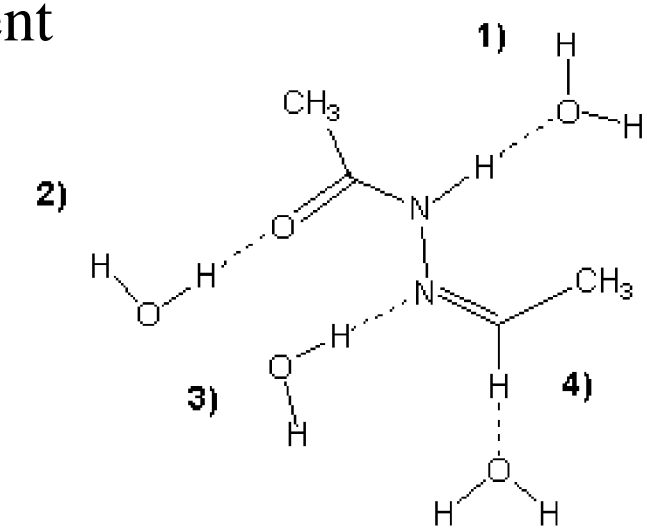
From MacKerell



Starting charges??
- peptide bond
- methyl
- imidazole (N-N=C)?
- Mulliken population analysis

Final charges (methyl, vary q_C to maintain integer charge, always q_H = 0.09)

- interactions with water (HF/6-31G*, monohydrates!)
- dipole moment

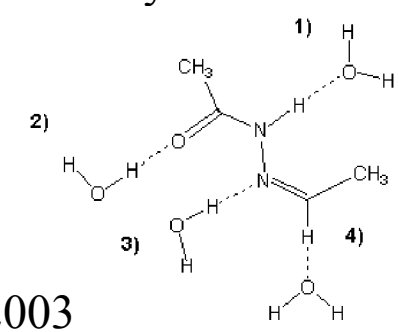


From MacKerell

Model compound 1-water interaction energies/geometries

	Interaction Energies (kcal/mole)			Interaction Distances (Å)		
	Ab initio	Analogy	Optimized	*Ab initio*	Analogy	Optimized
1) O2...HOH	-6.12	-6.56	-6.04	2.06	1.76	1.78
2) N3-H..OHH	-7.27	-7.19	-7.19	2.12	1.91	1.89
3) N4...HOH	-5.22	-1.16	-5.30	2.33	2.30	2.06
4) C5-H..OHH	-3.86	-3.04	-3.69	2.46	2.51	2.44
Energetic statistical analysis						
Ave. Difference		1.13	0.06			
RMS Difference		1.75	0.09			
Dipole Moments (debeye)						
	5.69	4.89	6.00			

Ab initio interaction energies scaled by 1.16.



MacKerell: Sanibel Conference 2003

Summary of Parameterization

1. LJ (VDW) parameters – normally direct transfer from available parameters is adequate, but should be tested by comparison to heats of vaporization, density, partial molar volumes, crystal simulations,.... (MacKerell JCC 2002). Other solvents?

2. Bond, angle, dihedral, UB and improper force constants

Vibrational spectra- Frequencies
Conformational Energetics -
Relative energies
Potential energy surfaces

Vibrations are generally used to optimize the bond, angle, UB and improper FCs while conformational energies are used for the dihedral FCs. However, vibrations will also be used for a number of the dihedral FCs, especially those involving hydrogens and in rings.(MacKerell 2003)

Vibrational Spectra of Model Compound 1 from MP2/6-31G* QM calculations

#	Freq	Assign	%	Assign	%	Assign	%	#	Freq	Assign	%	Assign	%
1	62	tC2N	64	tN3N	46			21	1446	rNH	35		
2	133	tC1H3	50	tN3N	18	tC2N	17	22	1447	rC5H	47	sC-N	18
3	148	tC1H3	46	tC6H3	25			23	1527	dCH3	77		
4	154	dC2NN	44	dN3NC	28	dN4CC	16	24	1532	dCH3	88		
5	205	tC6H3	59	tN4C	22	tN3N	21	25	1599	dCH3a'	50	dCH3a	17
6	333	tN4C	73	tC2N	22			26	1610	dCH3a	71	dCH3a'	24
7	361	dC1CN	45	dN4CC	21	dN3NC	16	27	1612	dCH3a'	30		
8	446	rC=O	32	dN4CC	20			28	1613	dCH3a	70	dCH3a'	23
9	568	wNH	77					29	1622	dCH3a'	57	dCH3a	19
10	586	dC1CN	21	dC2NN	20	rC=O	18	30	1782	sN=C	71		
11	618	wC=O	83	wNH	28	tC2N	-26	31	1901	sC=O	78		
12	649	rC=O	27	dN4CC	19			32	3250	sCH3	76	sC5-H	21
13	922	sC1-C	62					33	3258	sC5-H	78	sCH3	21
14	940	wC5H	80					34	3280	sCH3	99		
15	1031	rCH3'	33	sC5-C	31			35	3330	sCH3a	75	sCH3a'	25
16	1114	rCH3	66					36	3372	sCH3a'	100		
17	1139	rCH3'	76	wC=O	20			37	3377	sCH3a'	73	sCH3a	24
18	1157	rCH3	61	wC5H	21			38	3403	sCH3a	99		
19	1234	sC5-C	33	sN-N	32			39	3688	sN-H	100		
20	1269	sN-N	36	rCH3'	18								

Frequencies in cm^{-1}. Assignments and % are the modes and there respective percents contributing to each vibration.

From MacKerell

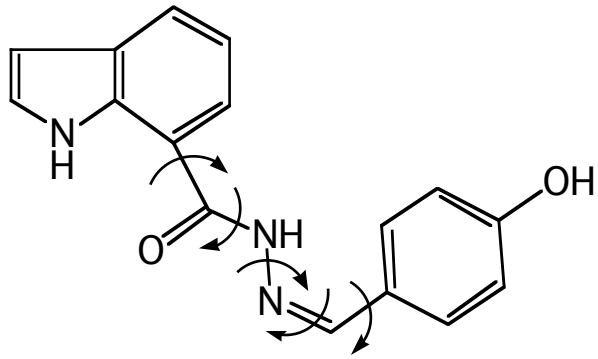
Comparison of the scaled ab initio, by analogy and optimized vibrations for selected modes

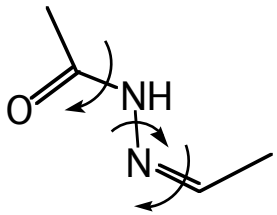
g98				Analogy				Optimized		
#	Freq	Assi	%	#	Freq	Assi	%	#	Freq	Assi
sN=N										
30	1782	sN=C	71	21	1228	sN=C	37	31	1802	sN=C
						rC5H	36			sN-N
				30	1646	sN=C	28			
						sC5-C	24			
						rC5H	18			
sN-N										
19	1234	sC5-C	33	20	1113	sN-N	53	20	1200	rNH
		sN-N	32			rNH	26			sN-N
20	1269	sN-N	36							rC5H
		rCH3'	18					23	1395	dCH3
										sN-N
								31	1802	sN=C
										sN-N
dC2NN										
4	154	dC2NN	44	5	207	dC2NN	36	4	158	dC2NN
		dN3NC	28			tN4C	31			dN3NC
		dN4CC	16							dN4CC
10	586	dC1CN	21	12	607	dC1CN	26	11	574	dC1CN
		dC2NN	20			dC2NN	25			dC2NN
		rC=O	18							dN4CC

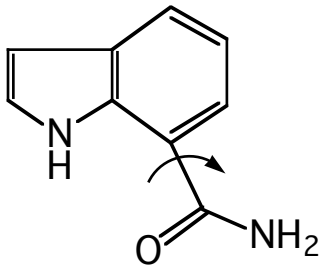
NH1-NR1 from 400/1.38 to 550/1.36
NR1=CEL1 from 500/1.342 to 680/1.290:
C-NH1-NR1 from 50.0/120.0 to 50.0/115.0,

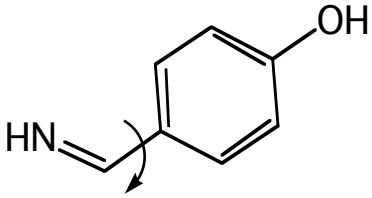
From MacKerell

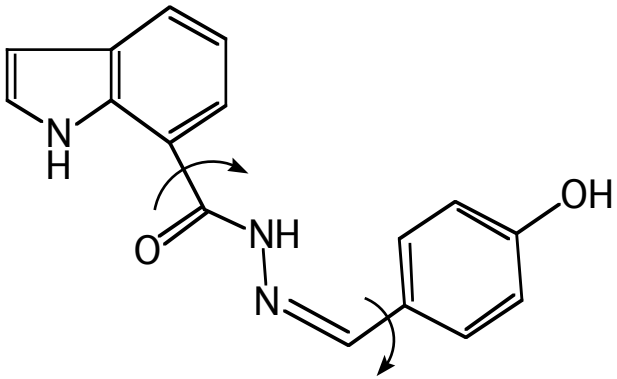
Dihedral optimization based on QM potential energy surfaces (HF/6-31G* or MP2/6-31G*).





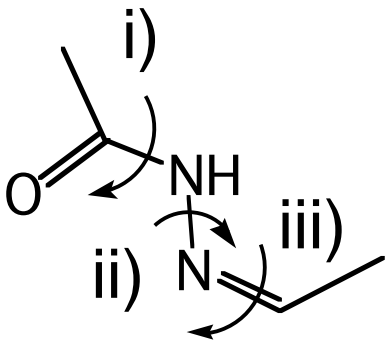






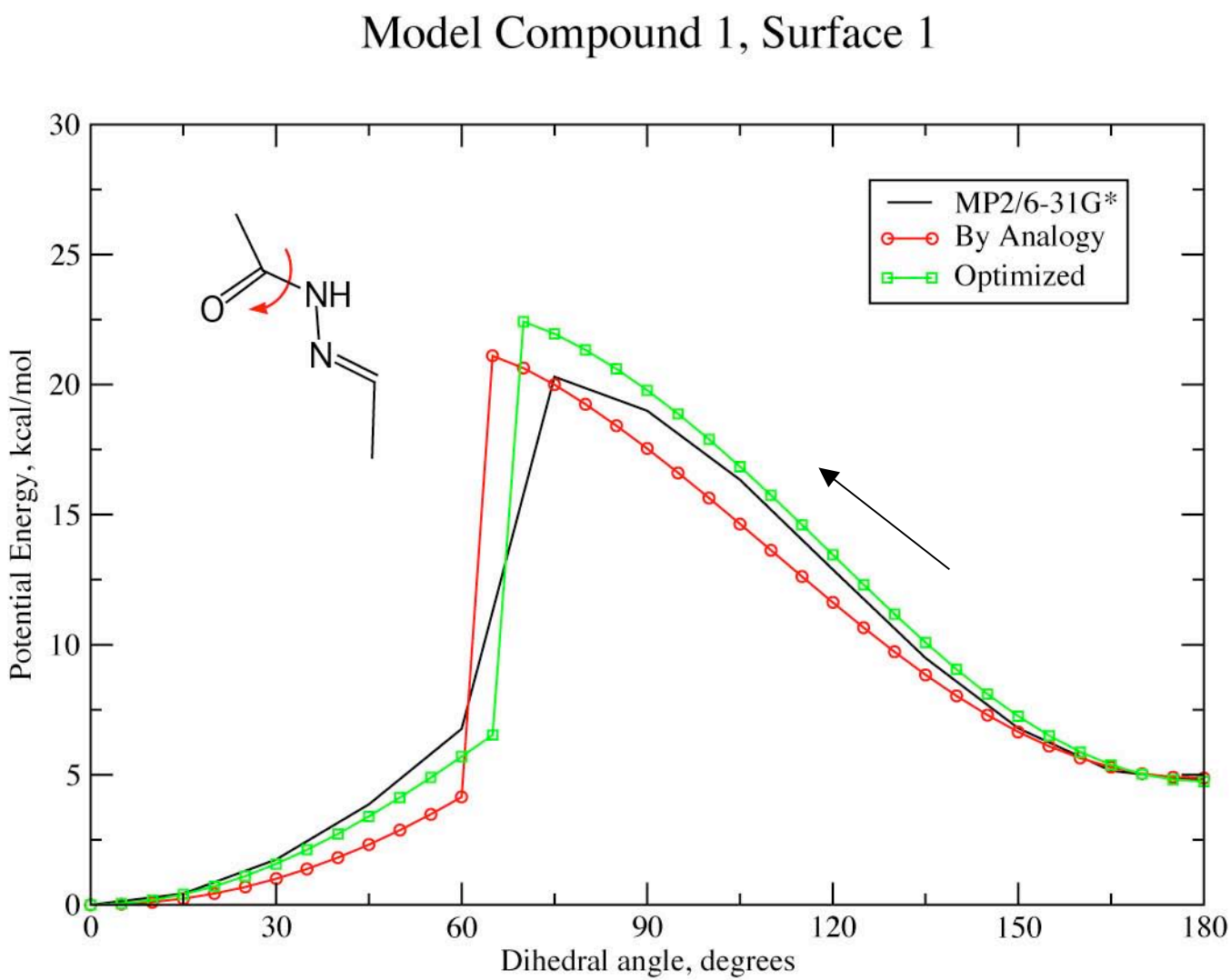
From MacKerell

Potential energy surfaces on compounds with multiple rotatable bonds



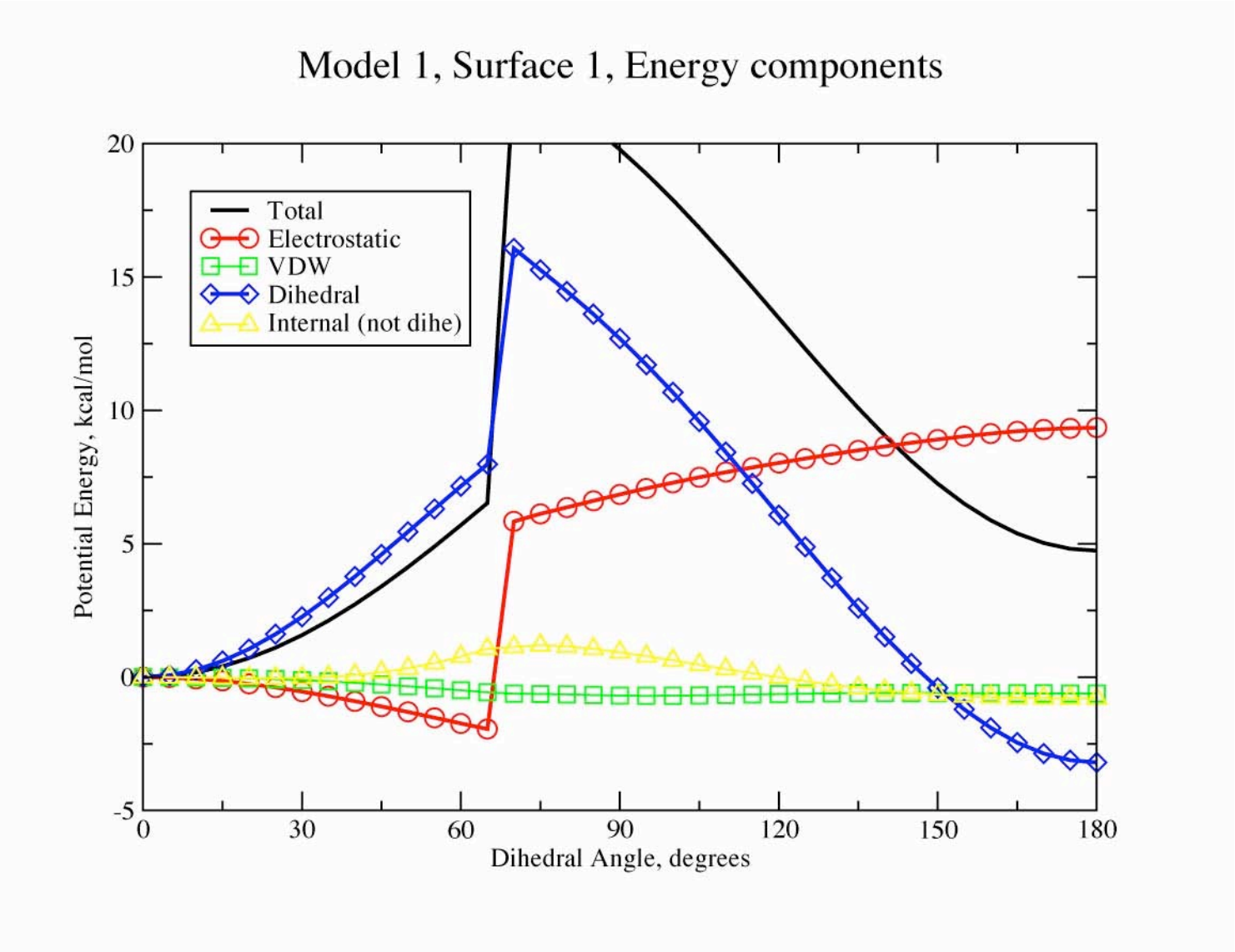
1) Full geometry optimization
2) Constrain n-1 dihedrals to minimum energy values or trans conformation
3) Sample selected dihedral surface
4) Repeat for all rotatable bonds dihedrals
5) Repeat 2-5 using alternate minima if deemed appropriate

From MacKerell



Note that the potential energy surface about a given torsion is the sum of the contributions from ALL terms in the potential energy function, not just the dihedral term

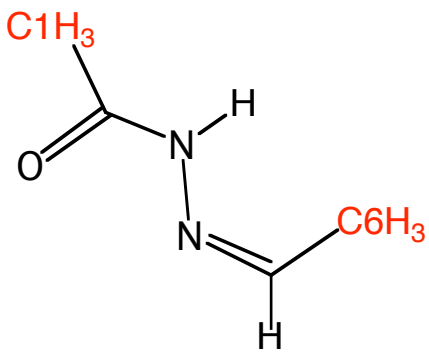
From MacKerell

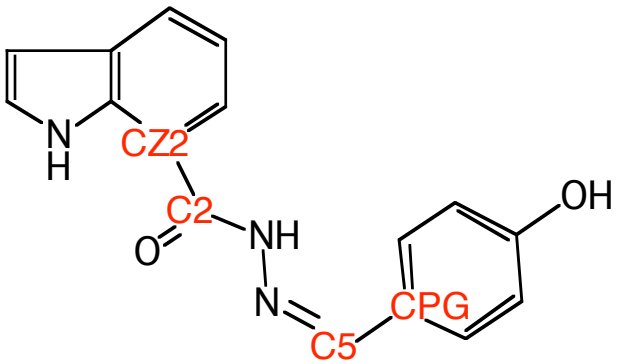


From MacKerell

Creation of full compound

1) Obtain indole and phenol from top_all22_model.inp
2) Rename phenol atom types to avoid conflicts with indole (add P)
3) Delete model 1 terminal methyls and perform charge adjustments
   i) Move HZ2 charge (0.115) into CZ2 (-0.115 -> 0.000) total charge on deleted C1 methyl (0.00) onto CZ2 (0.00 -> 0.00)
   ii) Move HPG charge (0.115) into CPG (-0.115 -> 0.000) and move total charge on the C6 methyl (0.18) onto CPG (0.00 -> 0.18)
4) Add parameters by analogy (use CHARMM error messages)
5) Generate IC table (IC GENErate)
6) Generate cartesian coordinates based on IC table (check carefully!)





From MacKerell

Chemistry of Thioesters

Most important example in biology of a thioester is acetyl coA, an intermediate used by nature in the biosynthesis of numerous organic compounds.

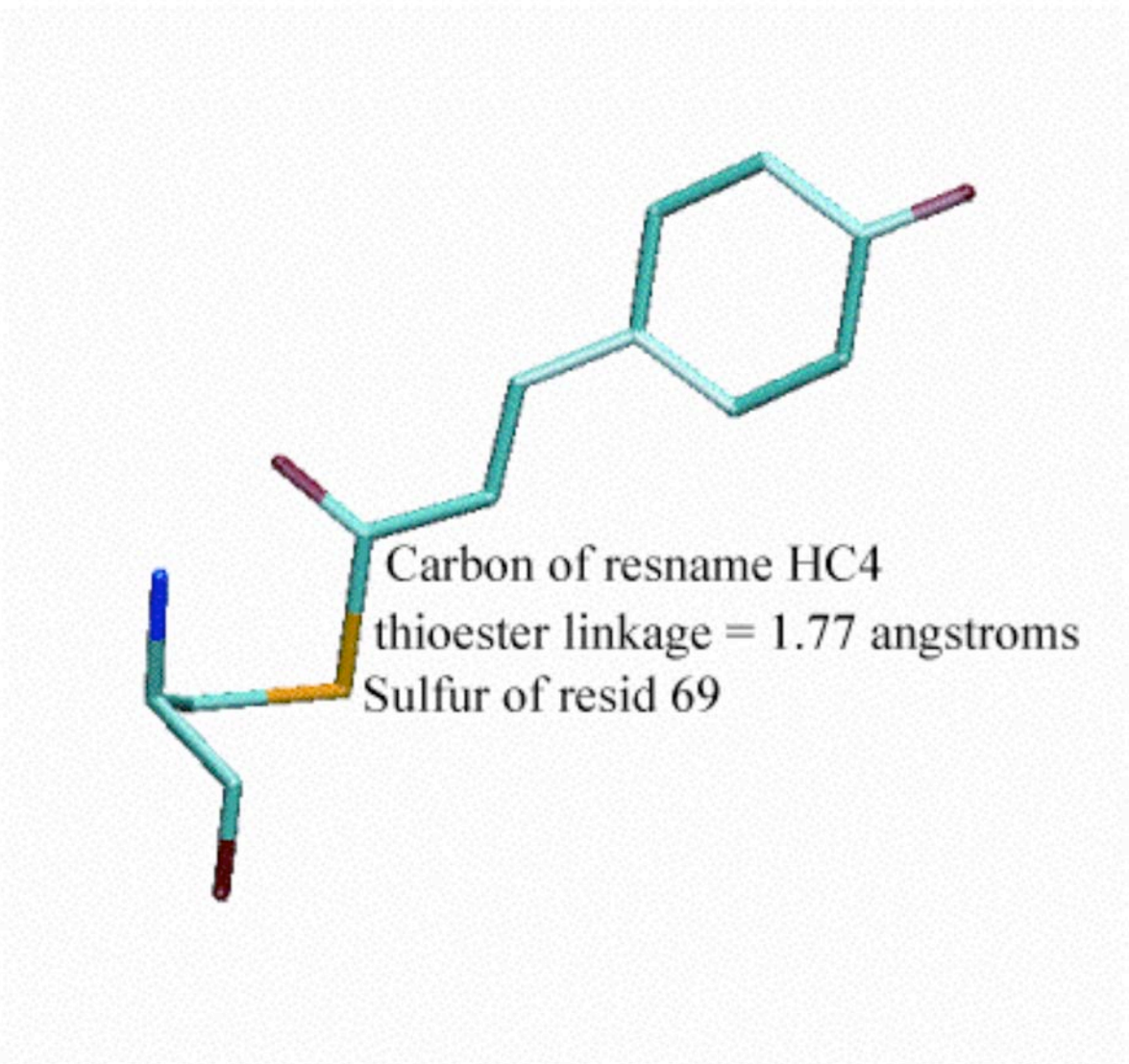
```
 O
 ||
-C – S - C-
```

Experimental Data*

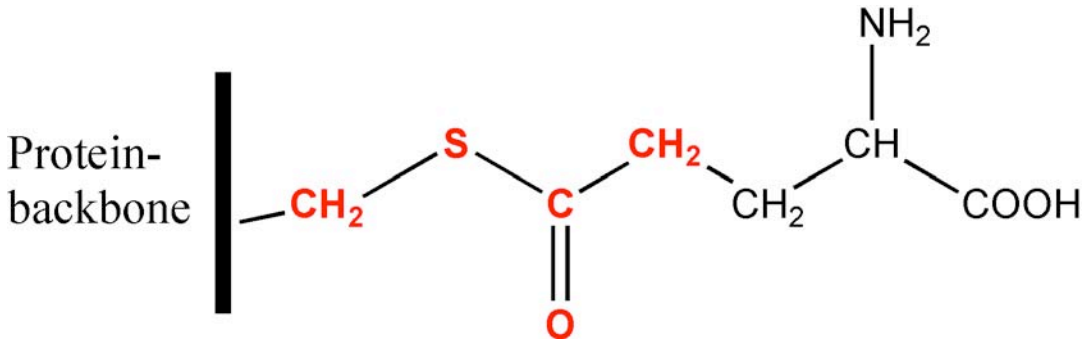
C-S (1.75 A), O=C-S-C (~-4), C-S-C-H (low barrier)
R-C-S (~113), R-C-O (~123), S-C-O (~124)

***Arch.Bioch.Biophys. Zacharias et al. v222,22-34,1983**

Thioester Linkage in Photoactive Yellow Protein - PDB



```
RESI CYG 0.00
GROUP
ATOM N     NH1     -0.47  !      |
ATOM HN    H        0.31  !  HN-N
ATOM CA    CT1      0.07  !      |   HB1
ATOM HA    HB       0.09  !      |   |
GROUP                     !  HA-CA--CB--SG
ATOM CB    CT2     -0.11  !      |   |   |
ATOM HB1   HA       0.09  !      |   HB2 |
ATOM HB2   HA       0.09  !   O=C       |
ATOM SG    S       -0.07  !      |        \
!ATOM HG1  HS       0.16 !                 \
GROUP                     !                  \
ATOM CDG   CC       0.55 !                   \
ATOM OE1   O       -0.55 !                    \
GROUP                      !     HN2G          \
ATOM CGG   CT2     -0.18   !      |             \
ATOM HG1G  HA       0.09   ! HN1G-NG   HB1G HG1G\
ATOM HG2G  HA       0.09   !      |    |     |   \
GROUP                      !  HAG-CAG--CBG--CGG--CDG=OE1
ATOM CBG   CT2     -0.18   !      |    |     |
ATOM HB1G  HA       0.09   !      |    HB2G HG2G
ATOM HB2G  HA       0.09   !  O1G=CG
GROUP                      !      |
ATOM CG    CD       0.75   !      O2G-HO2G
ATOM O1G   OB      -0.55
ATOM O2G   OH1     -0.61
ATOM HO2G  H        0.44
ATOM CAG   CT1     -0.12
ATOM HAG   HB       0.09
ATOM NG    NH3      -0.62
ATOM HN1G  HC        0.31
ATOM HN2G HC       0.31
GROUP
ATOM C     C        0.51
ATOM O     O       -0.51
```



HG1 deleted from CYS and the charge was moved to SG (-0.23 +0.16=0.07) so that the SG charge becomes 0.07 in final compound and the group remains neutral.

This can be improved!!

Quantum Chemical Calculations for New CF Parameters

Classical Potentials:

$$V(\vec{R}) = \sum_{i \in bonds} k_{i,bond}\left(r_{i,bond} - r_{i,bond}^{eq}\right) + \sum_{i \in angles} k_{i,angle}\left(\theta_{i,angle} - \theta_{i,angle}^{eq}\right) +$$

$$+\tfrac{1}{2}\sum_{\substack{i \neq j \\ i,j \in atoms}} \frac{q_i q_j}{\left|r_i - r_j\right|} + \tfrac{1}{2}\sum_{\substack{i \neq j \\ i,j \in atoms}} V_{LJ}^{ij}\left(\left|r_i - r_j\right|\right) \quad + \ldots.$$

QM Operators: $\widehat{H}\Psi_{e,n} = E\Psi_{e,n}$ Many particle wavefunction

$$\widehat{H} = \widehat{T}_n + \widehat{T}_e + \widehat{V}_{e,n}$$

Born Oppenheimer Approximation: $\Psi_{e,n} = \chi_n \psi_{e,n}$

$$\widehat{H}_{electronic}(\vec{R})\psi_{i,electronic}(\vec{r}_{electronic};\vec{R}) = E_{i,electronic}(\vec{R})\psi_{i,electronic}(\vec{r}_{electronic};\vec{R})$$

Electronic Hamiltonian

$$\hat{H}_{electronic} = \sum_{i} \hat{T}(i) - \sum_{iA} \frac{Z_A}{r_{iA}} + \sum_{i,j} \frac{1}{r_{ij}} + \sum_{A,B} \frac{Z_A Z_B}{R_{AB}}$$

Kinetic Energy of electrons

Electron-nucleus attraction

Electron-electron repulsion

Nuclear-nuclear repulsion

The Never-Ending Contraction

$$\chi_k^{AO,CBF} = \sum_{p=1}^{N_k} c_{pk}^{\mathrm{Prim}->CBF} \tilde{\chi}_p^{AO,primitive}(r,R)$$

Every atomic orbital is a fixed contraction of Gaussians

One-particle basis set

$$\phi_i^{MO} = \sum_{k=1}^{N_{CBF}} c_{ik}^{MO}(R)\chi_k^{AO,CBF}(r,R)$$

Molecular orbitals are orthogonal contractions of AOs

$$\psi_\mu(r,R) = A\left[\prod_i \phi_i^{MO}(r,R)\right]$$

Antisymmetrized products of MOs

Many-particle Basis set

$$\Psi_{elec}(r,R) = \sum_{\mu=1}^{N_{AP}} c_\mu^{CI}(R)\psi_\mu(r,R)$$

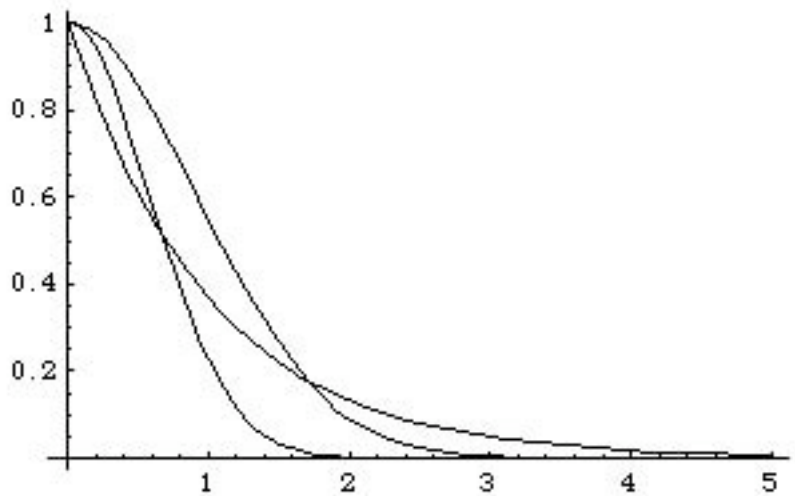
Total electronic wfn is contraction of APs

See T. Martinez, http://www.ks.uiuc.edu/Training/SumSchool/lectures.html

Basis Function Overview

$$\phi_S(r) = \sum_{i=1}^{N} c_{i,S} e^{-\alpha_{i.S} f^2 r^2}$$

S-type basis function is composed of sum of primitive gaussian functions. N is the number gfs and is called the degree-of-contraction. C, α,and f are the contraction coefficients, exponents, and scale factors.



It takes at least 3 Gaussians to approximate an S orbital

$$e^{-r}, e^{-0.6r^2}, e^{-1.5r^2}$$

Basis Set Pople Classification

Minimal Basis Set - STO-3G

One BF per occupied orbital on an atom

e.g. H 1s, 1st row elements 1s,2s,2px,2py,2pz like orbitals

Split Valence Basis Set - Inner and Outer' basis functions

3-21G: 1s - 3G; 2s, 2p - 2G; 2s',2p' - 1G

6-31G: 1s - 6G; 2s, 2p - 3G; 2s', 2p' - 1G

Polarization

Add higher angular momentum functions

e.g. 6-31G* = 6-31G+d for 1st row

6-31G** = 6-31G+(d,p) 1st row and +p for H

Semiempirical Methods and Number of Parameters in Method

MNDO Modified Neglect of Differential Overlap 10
AM1 Austin Model 1 13
PM3 Parametric Model number 3 13

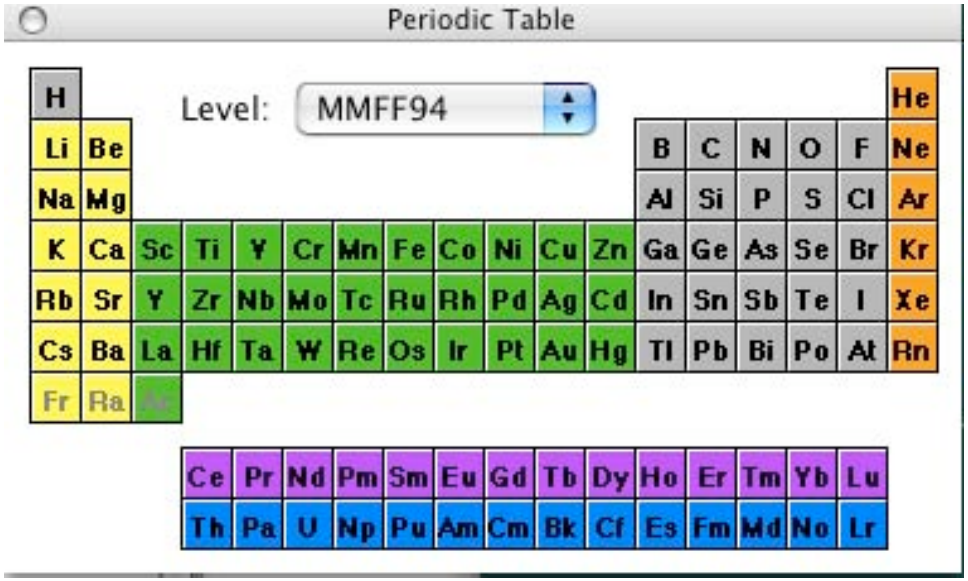
Perturbation Method to Treat Electron Correlation: MP2

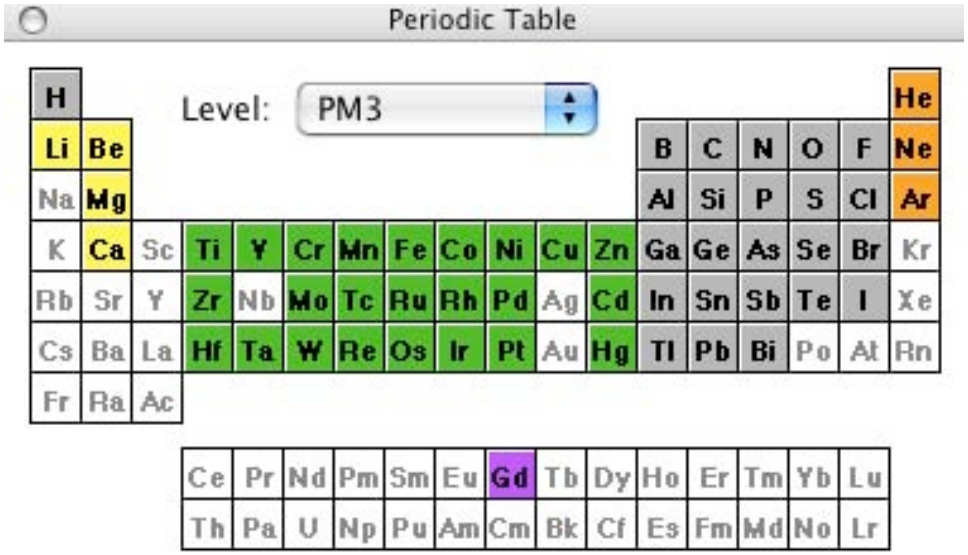
Improvement over HF, RHF and UHF

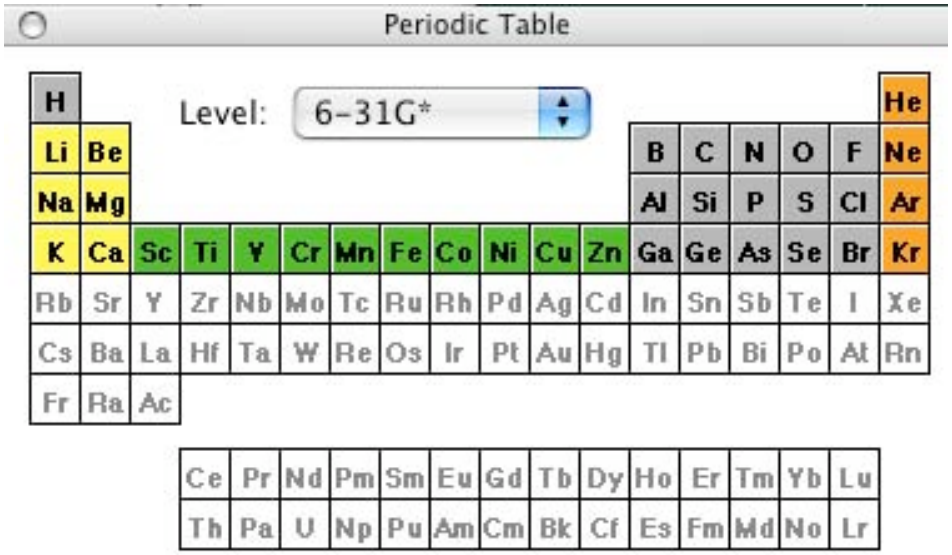
Molecular Properties from Wavefunctions

Dipole moment, partial charges, vibrations, ….

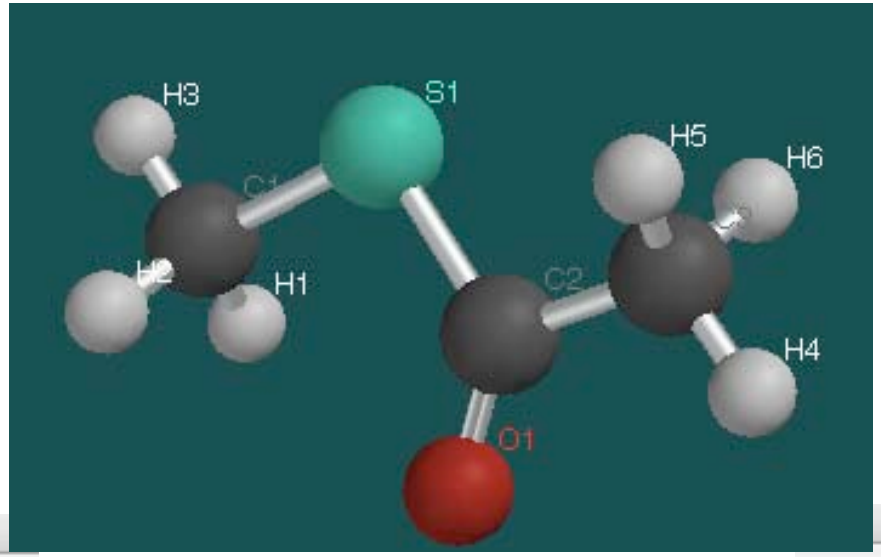
Choice of Basis Set in Spartan







PM3
Spartan



Vibrations

Frequency	Type
86	A"
108	A"
117	A"
180	A'
323	A'
483	A'
534	A"
630	A'
756	A'
931	A"
935	A'
988	A"
995	A'
1142	A'

Amplitude: 3.00 Steps: 7

Make List

Bond angle C-CO-S bend

Bond stretch C=O

CH3 stretches ↗

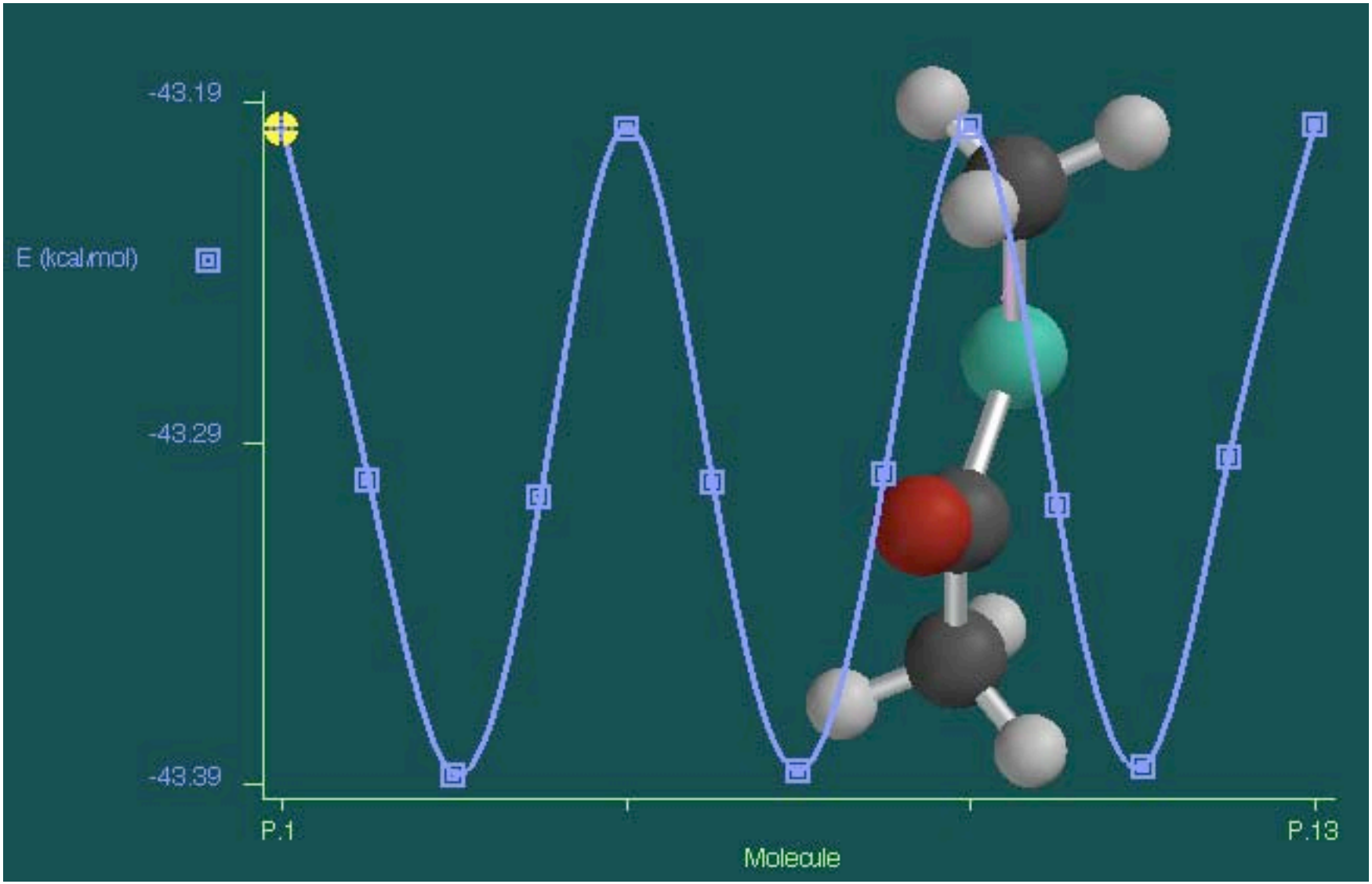
Vibrations

Frequency	Type
756	A'
931	A"
935	A'
988	A"
995	A'
1142	A'
1327	A'
1363	A'
1376	A"
1384	A"
1389	A'
1393	A'
1937	A'
3081	A"

Amplitude: 3.00 Steps: 7

Make List

Dihedral Draw to Check Barriers



Atom type	Mulliken 6-31G**	ESP 6-31G**	Charmm
CT2	-0.63	-0.08	-0.11
HA	0.22	0.07	0.09
HA	0.20	0.07	0.09
HA	0.22	0.12	
S	0.11	-0.29	-0.07
CC	0.36	0.64	0.55
O	-0.50	-0.50	-0.55
CT2	-0.58	-0.46	-0.18
HA	0.20	0.14	0.09
HA	0.21	0.14	0.09
HA	0.22	0.14	

No waters in QM calc.

Results on Fragment

Bonds	Ab initio 6-31G*	Minimization
CT2 – S	1.81 Å	1.85 Å
CC – S	1.78 Å	1.79 Å
CC – CT2	1.51 Å	1.54 Å
CC – O	1.19 Å	1.23 Å
Angles		
S – CC – CT2	114.3 º	117.1 º
CC – S – CT2	100.2 º	107.7 º
S – CC – O	122.2 º	122.9 º

Classical Force Fields

- Physics-based, full atom FF like CHARMM, AMBER, OPLS - Mechanisms, function, protein/RNA/DNA interactions…

✓ Knowledge-based, coarse-grained model -

Protein structure prediction, folding kinetics, docking,…

✓ Hybrid force fields like Go + full atom FF -

Mutational effects on protein folding kinetics,….

- Hybrid QM/MM approaches

Enzyme reactions, bond breaking and making, excited states,…

Each problem has a different goal and time scale!

Protein Structure Prediction

1-D protein sequence

3-D protein structure

SISSIRVKSKRIQLG….

***Ab Initio* protein folding**

Sequence Alignment

Target protein of unknown structure → SISSRVKSKRIQLGLNQAELAQKV------GTTQ···

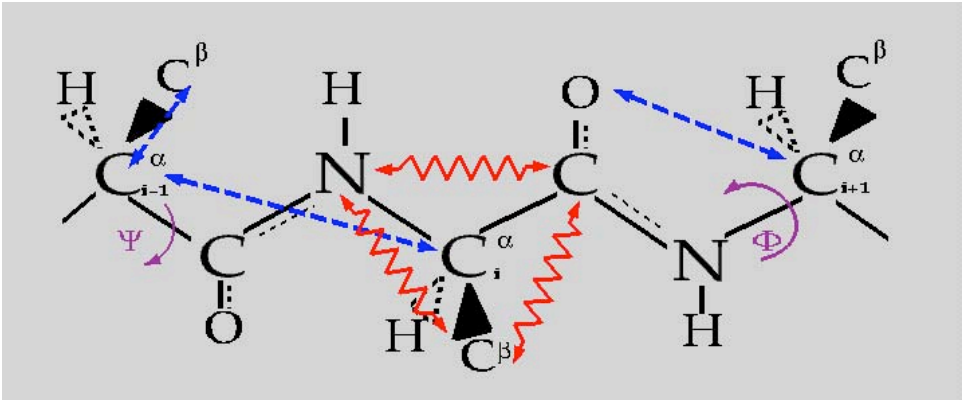
Homologous/analogous protein of **known** structure → QFANEFKVRRIKLGYTQTNVGEALAAVHGS···

Ab Initio Folding: the Energy Function

$$E = E_{backbone} + E_{residue/residue} \longrightarrow ?$$

Ab initio Structure Prediction – Prediction without Homology

$$V_{back} = V_{SHAKE} + V_{ev} + V_{chain} + V_{chi} + V_{\phi\psi} + V_{HB}$$



- Reduced Representation: C_α,C_β,O.
- Interaction Potentials: AMH and Contact averaged over MS
- Non-additive HB

Ab initio Interaction Potentials

$$E_{AMC} = E_{AM} + E_C = (E_{short} + E_{medium}) + E_{long}$$

Associate similar sequence/structure fragments in protein database

$$E_{AM} = -\frac{\varepsilon}{a}\sum_{\mu=1}^{N_\mu}\sum_{j-12\le i\le j-3}\left\{\gamma_{AM}[P_i, P_j, P_i^\mu, P_j^\mu, x(|i-j|)]\exp\left[\frac{-(r_{ij}-r_{i'j'}^\mu)^2}{2\sigma_{ij}^{\ 2}}\right]\right\}$$

Long range interactions mimic pair distribution functions

$$E_{long} = -\frac{\varepsilon}{a}\sum_{i<j-12}\sum_{k=1}^{3}\left\{\gamma_{long}\left[P_i, P_j, k\right]c_k(N)U\left[r_{\min}(k), r_{\max}(k), r_{ij}\right]\right\}$$

Weights γ *learned from training set*

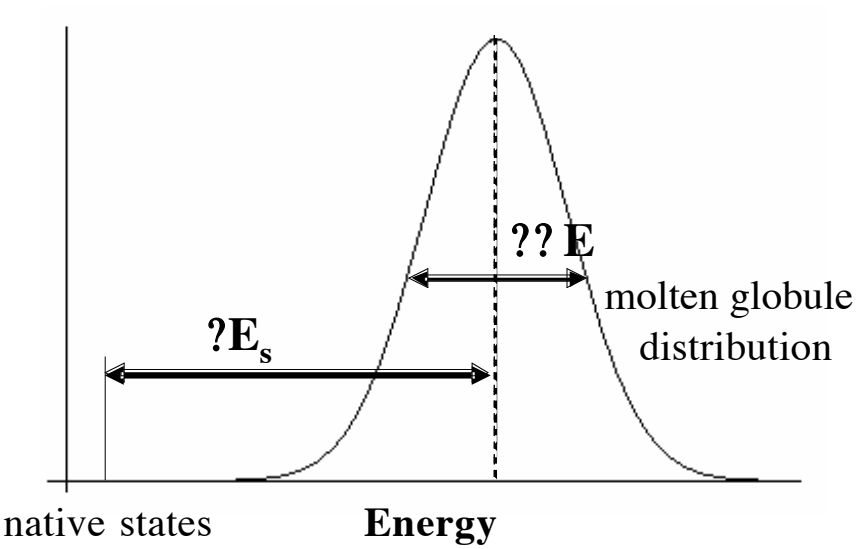
Optimization Strategy for Coarse Grained Energy Function

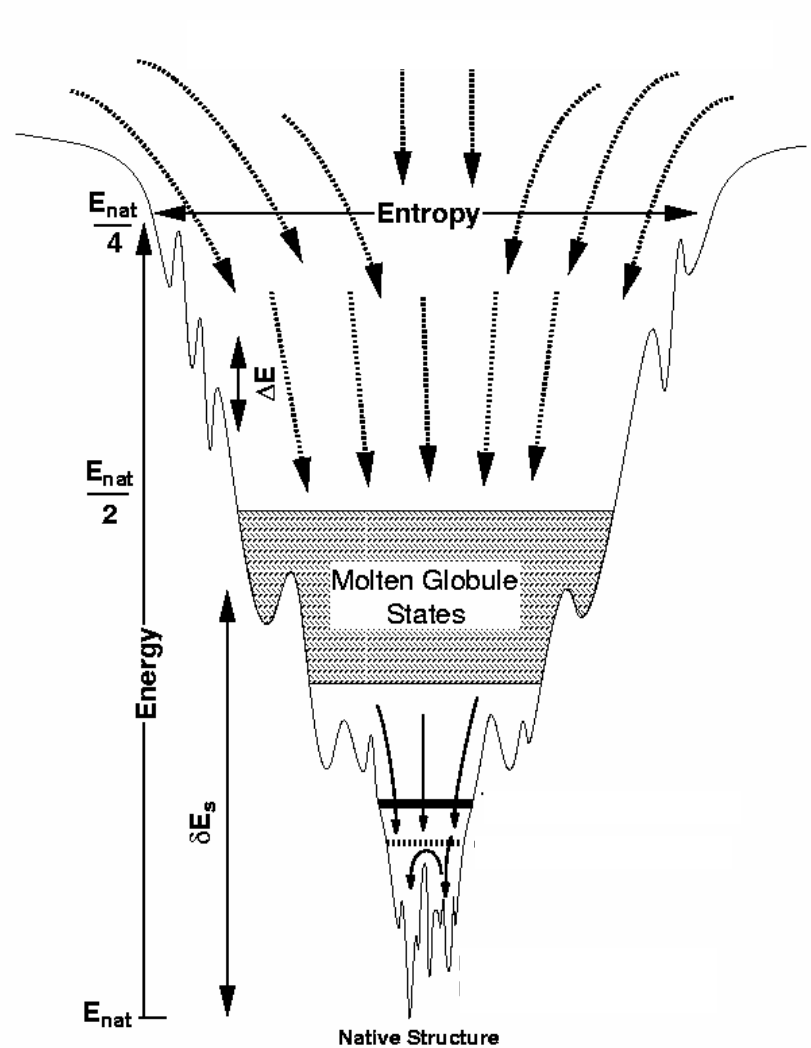
Energy Landscape Theory

When <?E_s /? E> is maximum
the energy landscape is**optimally funneled.**

Vary the parameters to obtain discrimination

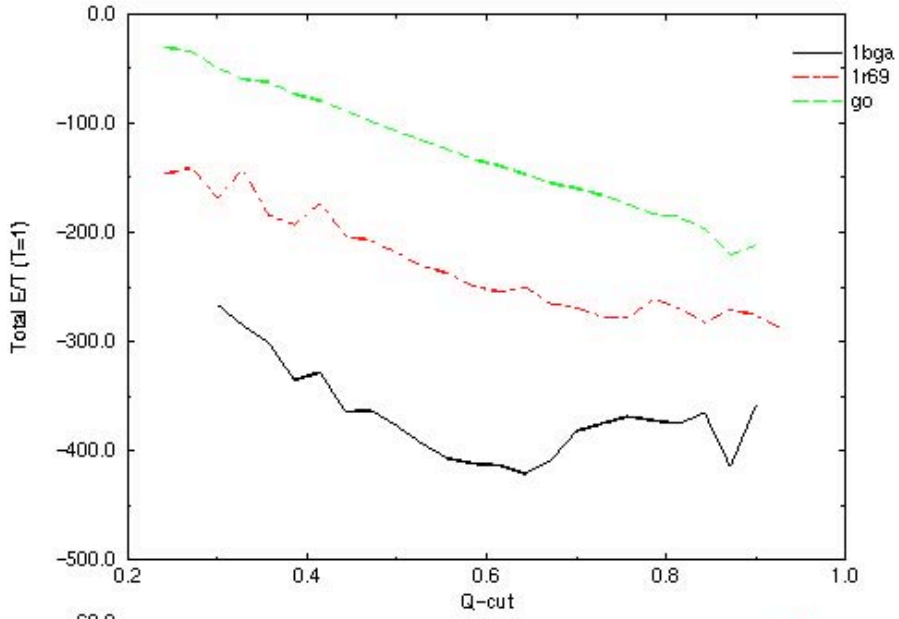
Optimization over an Ensemble of Folds

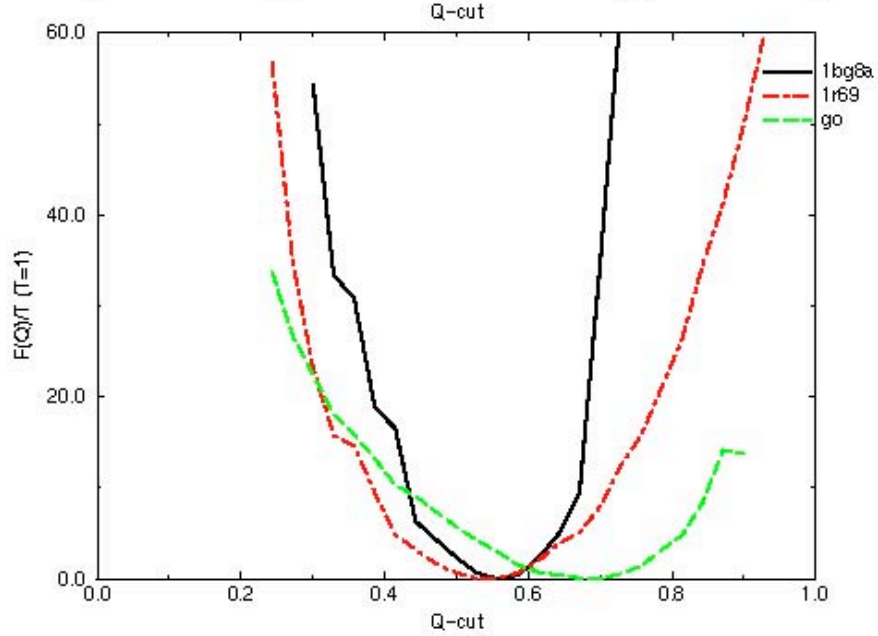




Onuchic, Luthey-Schulten, Wolynes (**1997**) *Annu. Rev. Phys. Chem.* 48:545-600.
Eastwood,Hardin,Luthey -schulten,Wolynes(**2001**)*IBM. J.RES.&DEV.*45:475-497

Energy Landscapes of Prediction Energy Function

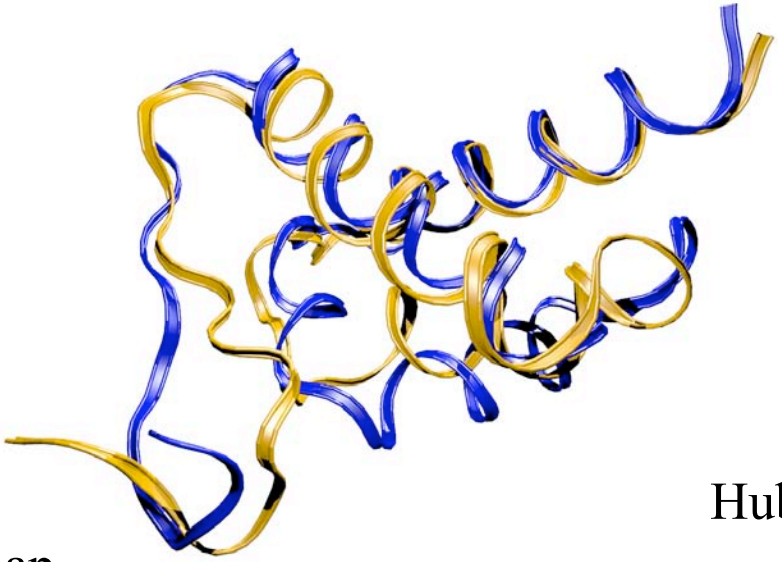




Q fraction of native contacts

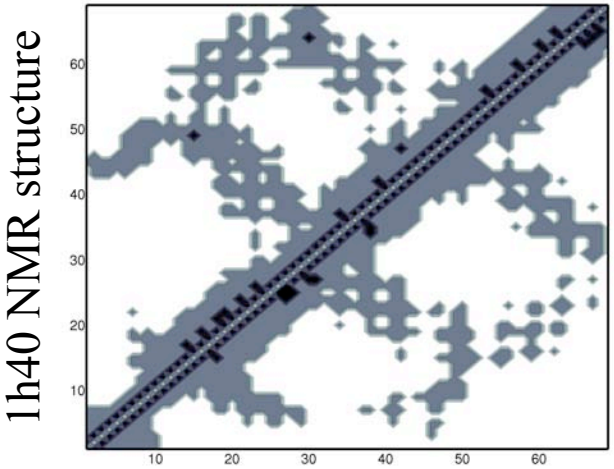
CASP5 Wolynes-Schulten Prediction T0170

FF domain of HYPA/FBP11

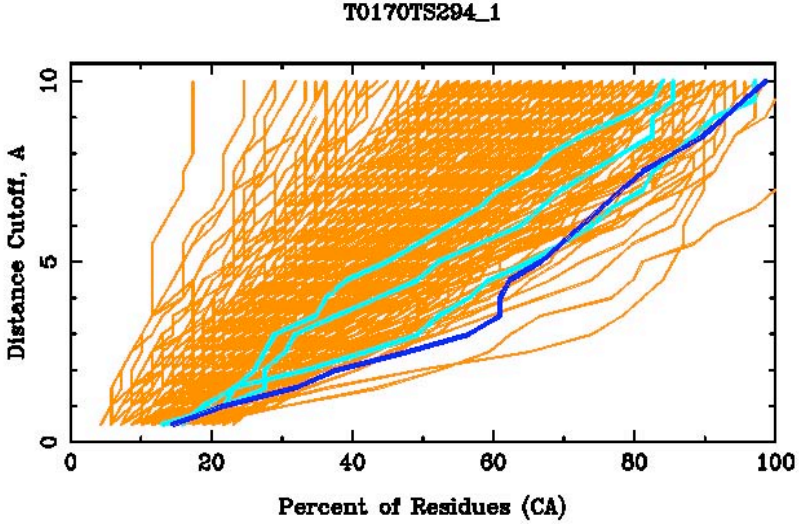


Blue – NMR structure
Orange – Predicted Structure
Global RMSD 2.67 Å.

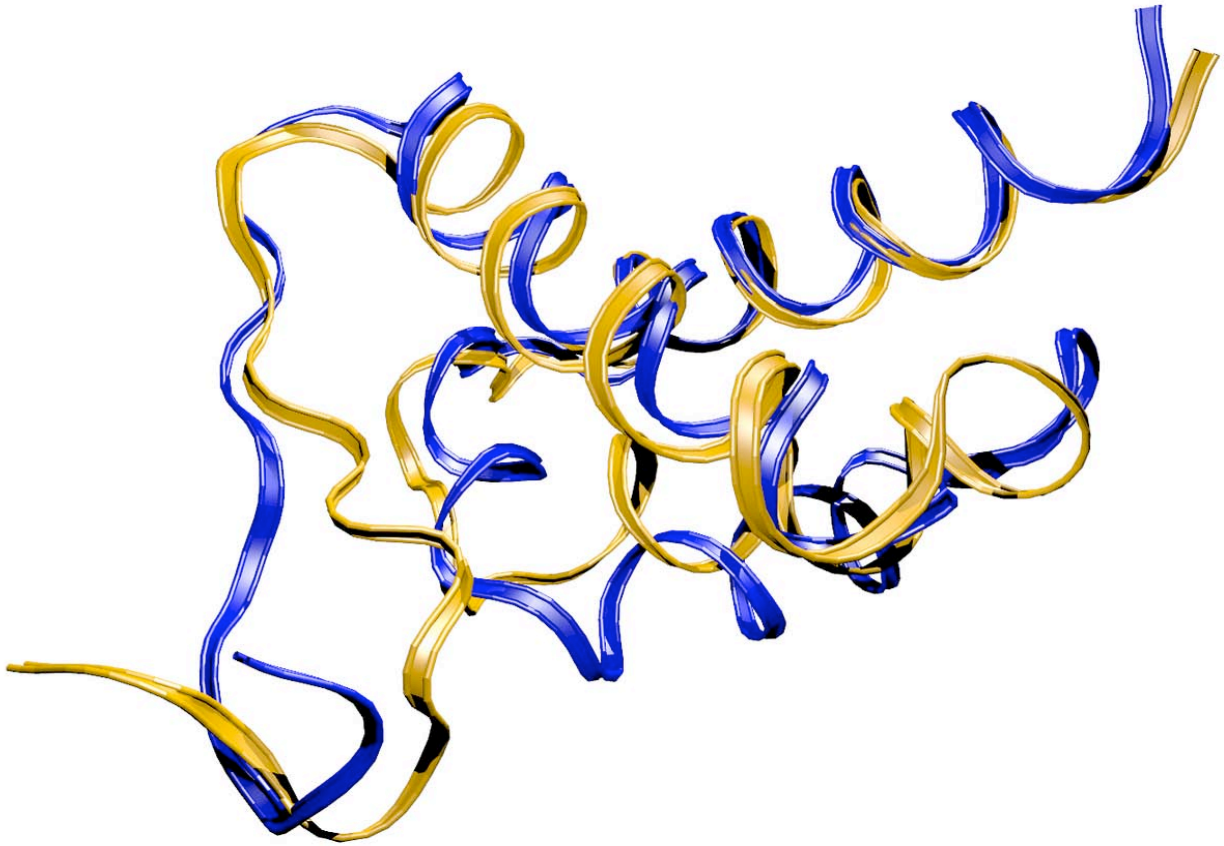
T0170 contact map



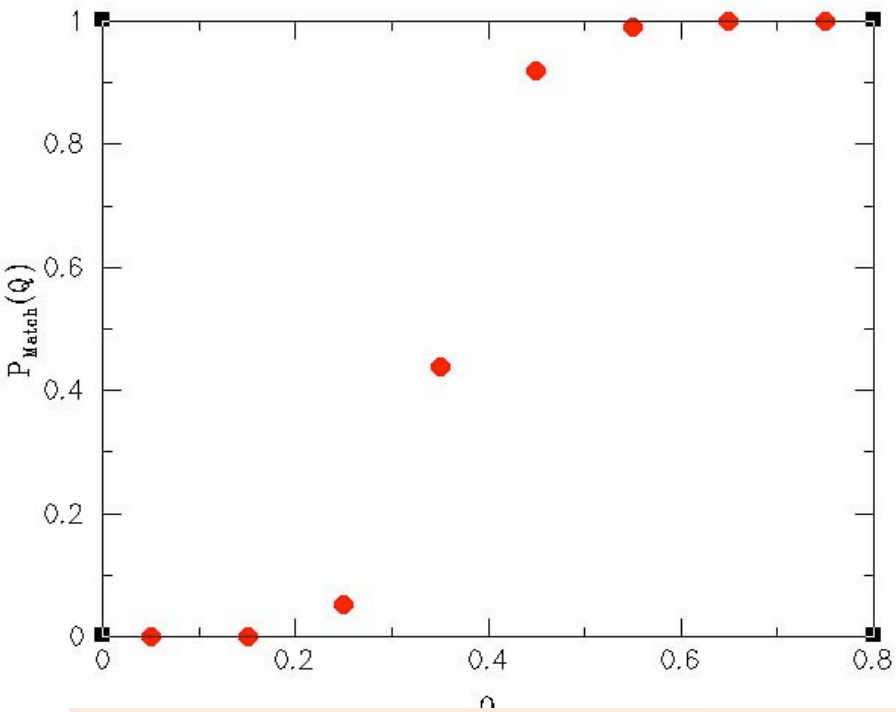
Hubbard Plot



CASP5 Result



Functional Annotation Requires Structures of ~ Q=0.4



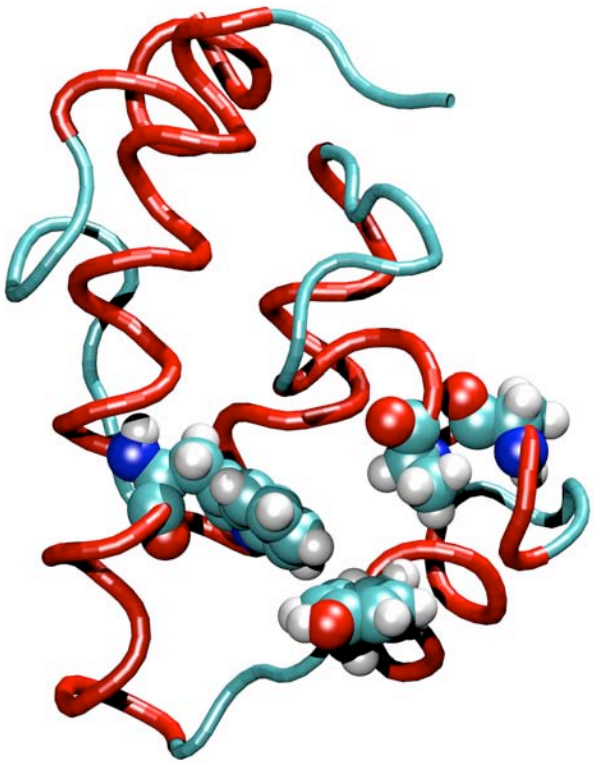
$$Q = \frac{1}{N_{pairs}} \sum_{i,j} \exp\left[-\frac{(r_{ij} - r_{ij}^{N})^2}{2\sigma_{ij}^2} \right]$$

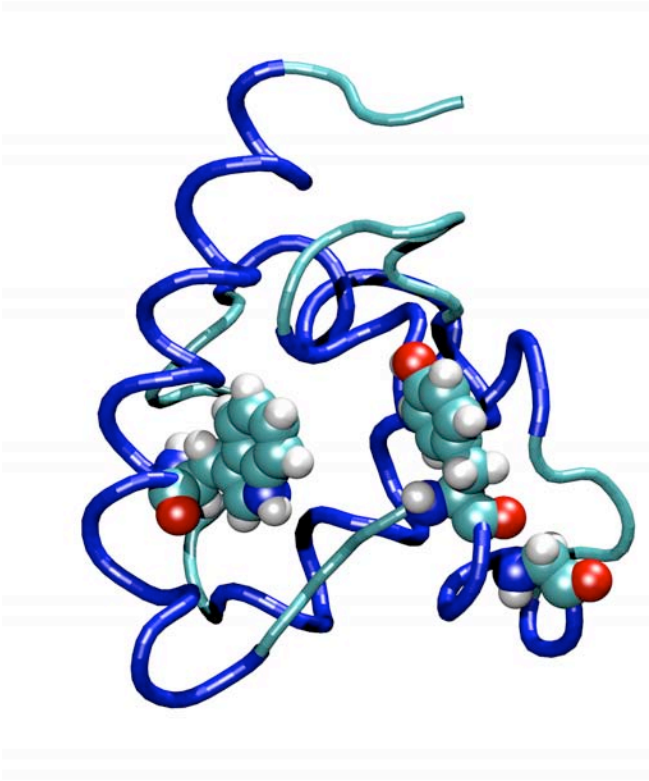
Hybrid Force Field Calculations
Future Directions?

- Protein Folding Kinetics and Mutational Studies
  - Go Potentials
  - Charmm + Go Potentials

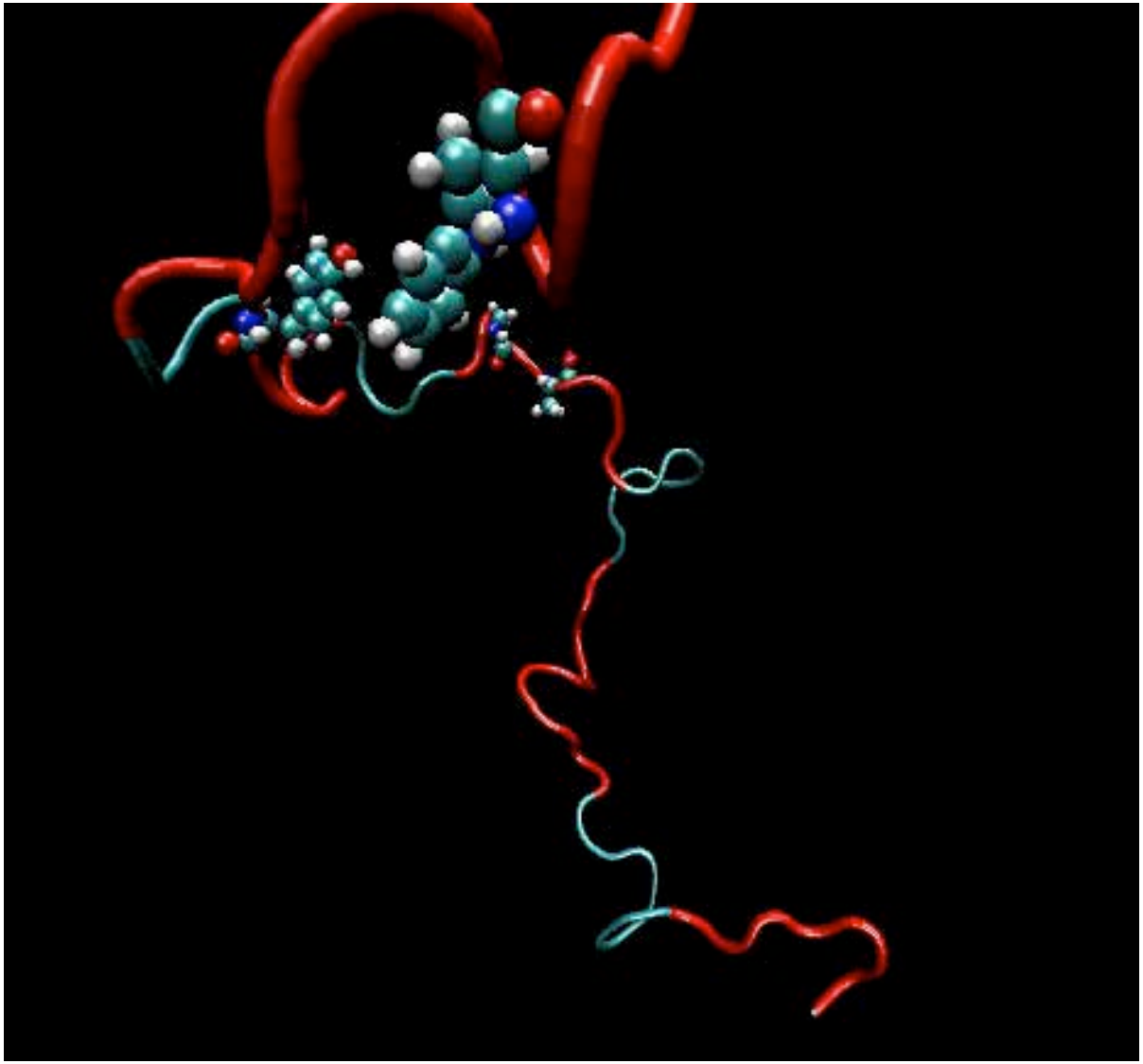
T. Pogorelov and Z. Luthey-Schulten, Biophysical J. 87 (2004)

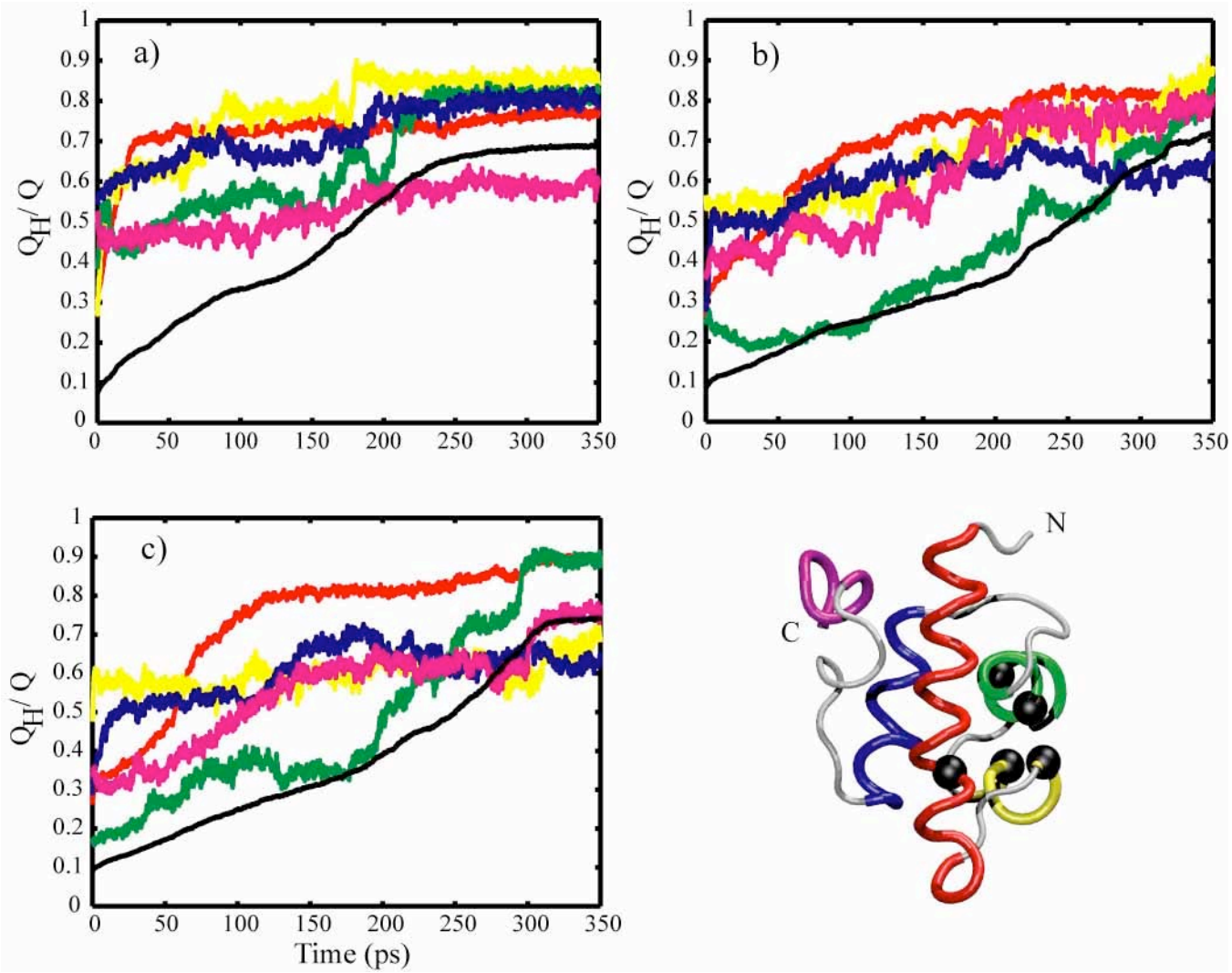
LAMBDA REPRESSOR MUTATIONS Q33Y and A37G





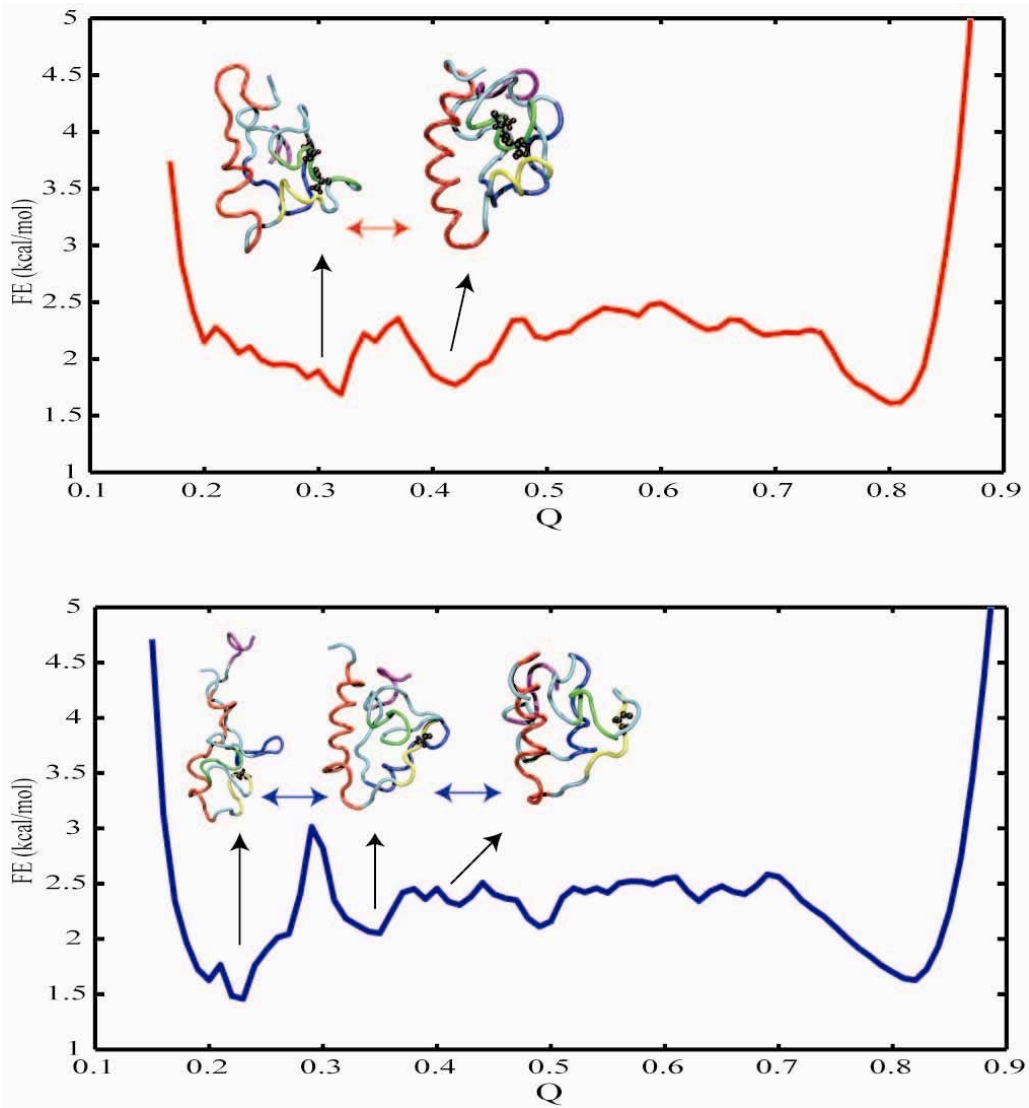
Folding of lambda-repressor





T. Pogorelov and Z. Luthey-Schulten, Biophysical J. 87 (2004)

Free-Energy Potentials for Fast and Slow Mutants



Acknowledgements

- Rommie Amaro – HisH/HisF
- Felix Autenrieth - Cyt c2 Parameterization
- Taras Pogorelo - Go + CHARMM