Advanced machine learning for
Innovative Drug Discovery (AIDD)

3D-Structure Refinement

Simple Principles for Meaningful Structures

Dr. Filipe Menezes,
Institute Structural Biology (STB), Popowicz group
Helmholtz-Zentrum München

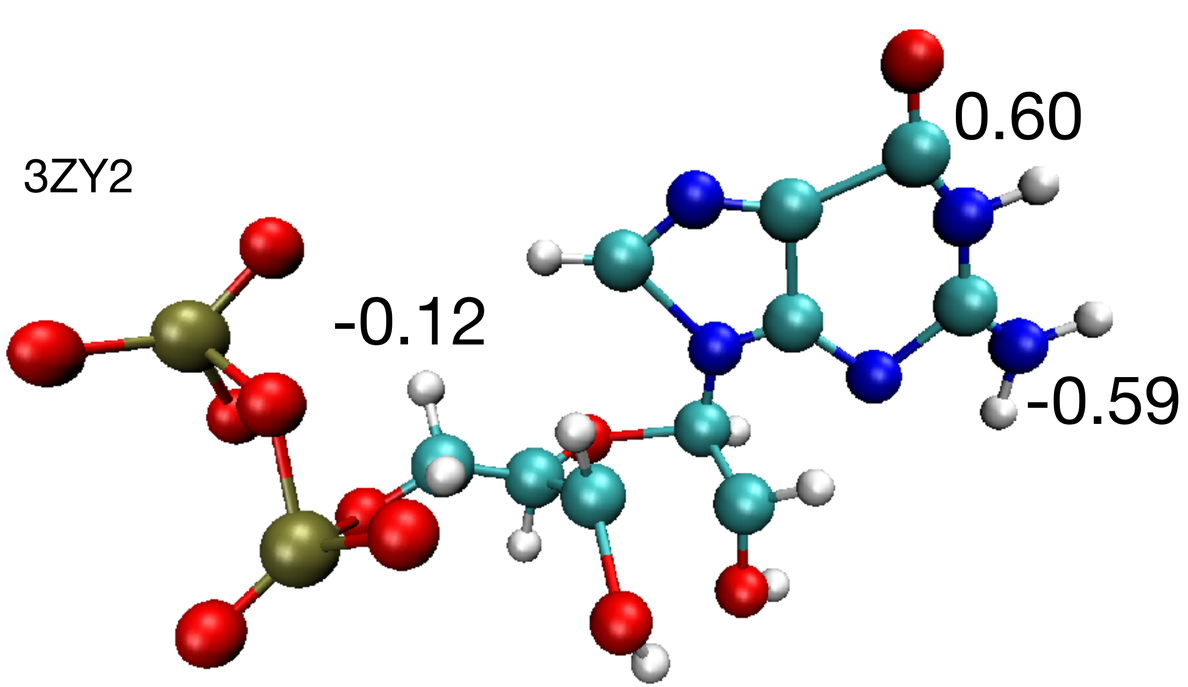
HELMHOLTZ MUNICH

Institute of Structural Biology

Bayerisches NMR Zentrum

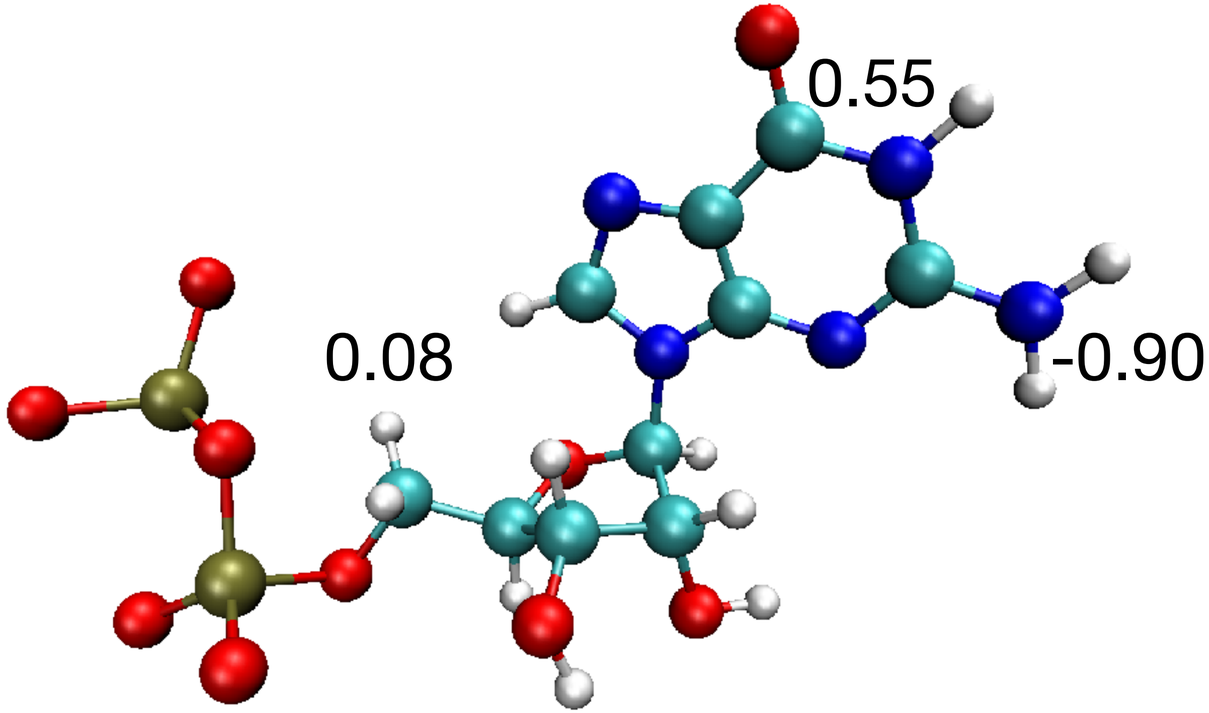
Why do we need good structures?

3ZY2



$$\Delta_{solv}G = -566.32 \quad kcal/mol$$

$$\mu = 62.8 \quad e\text{Å}$$

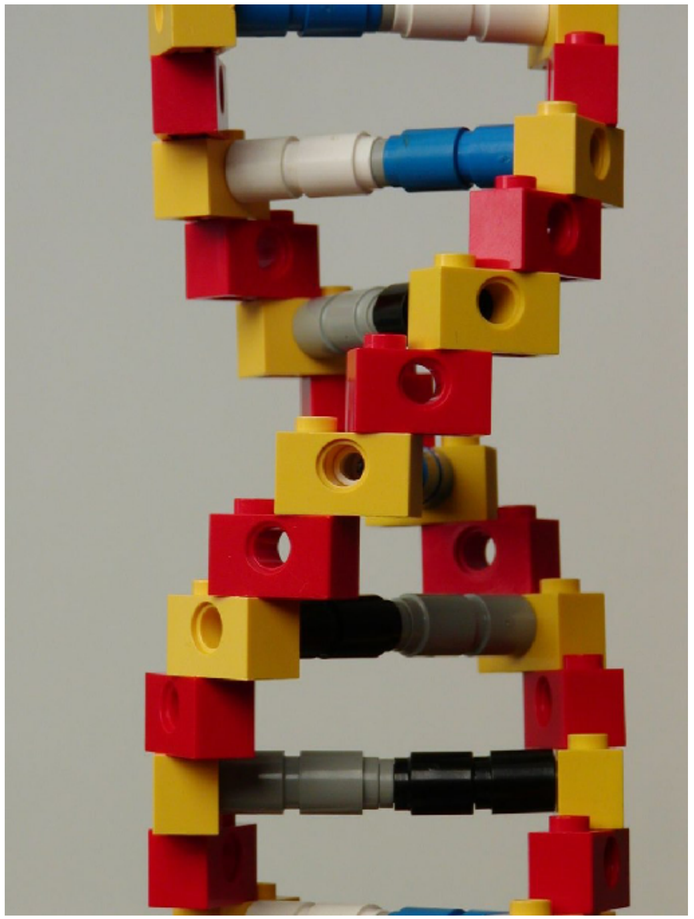


$$\Delta_{solv}G = -513.39 \quad kcal/mol$$

$$\mu = 51.5 \quad e\text{Å}$$

Part 1

Assembling Molecules

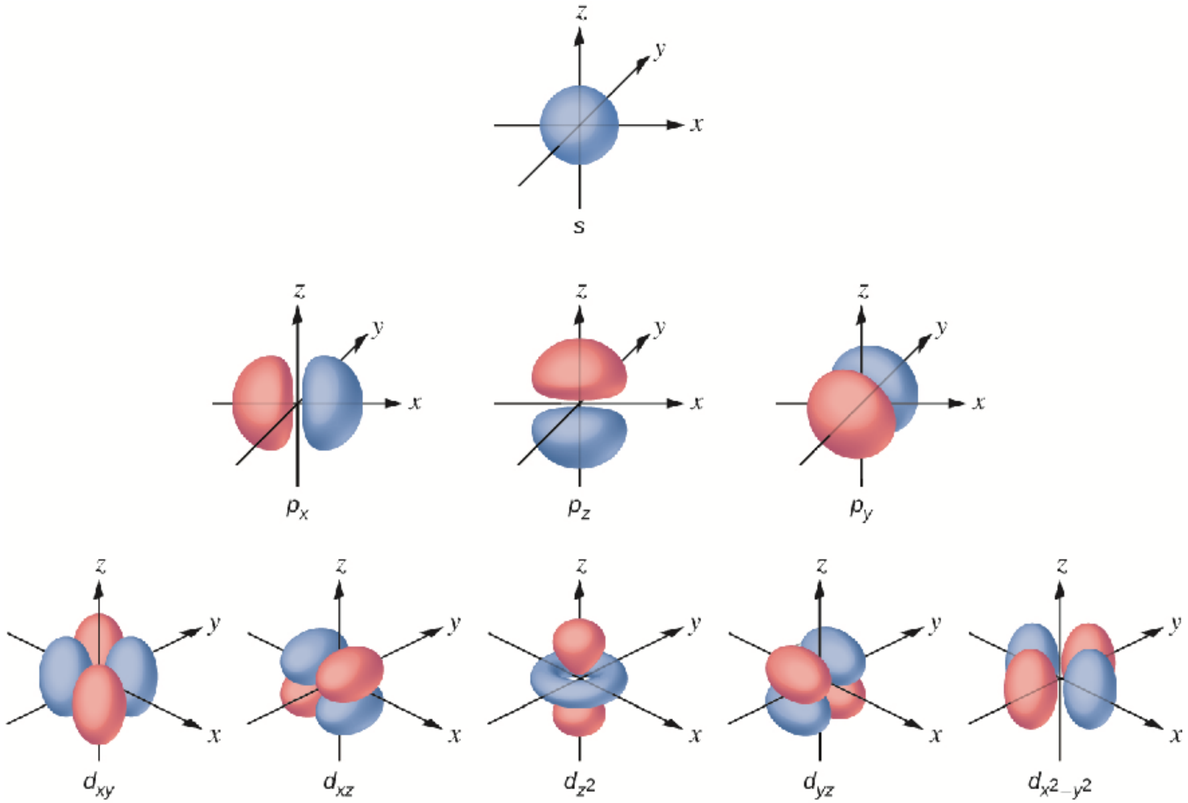


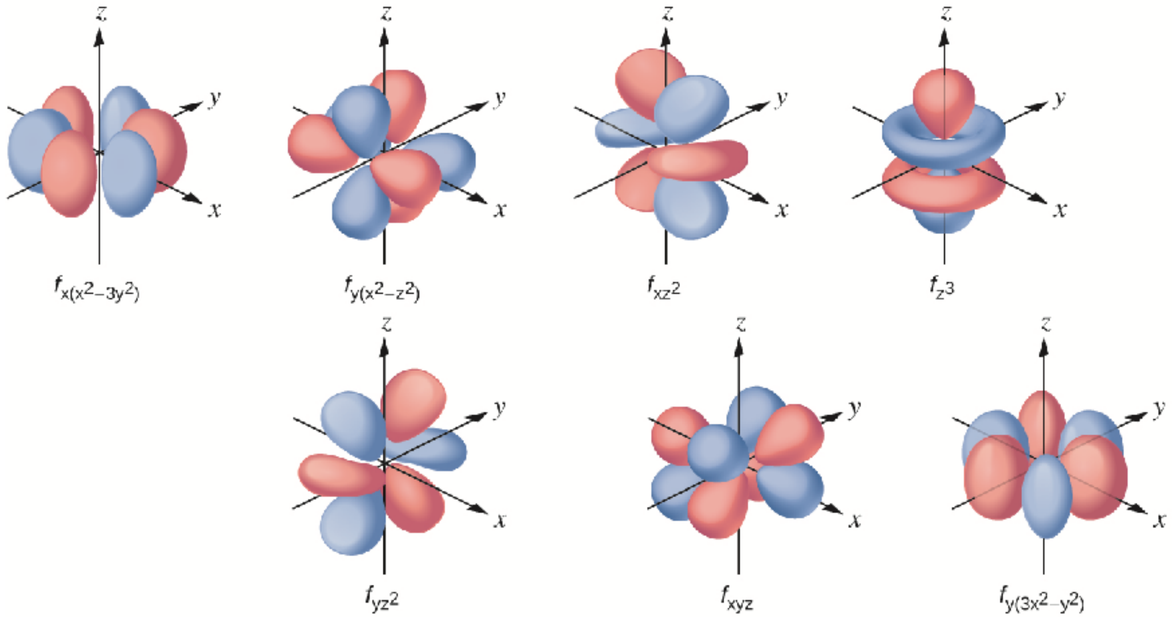
Link

The Schrödinger Equation

Erwin Schrödinger

$$(\hat{T} + \hat{V})\Psi = E\Psi$$

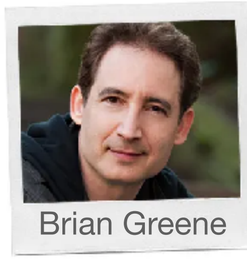




Link

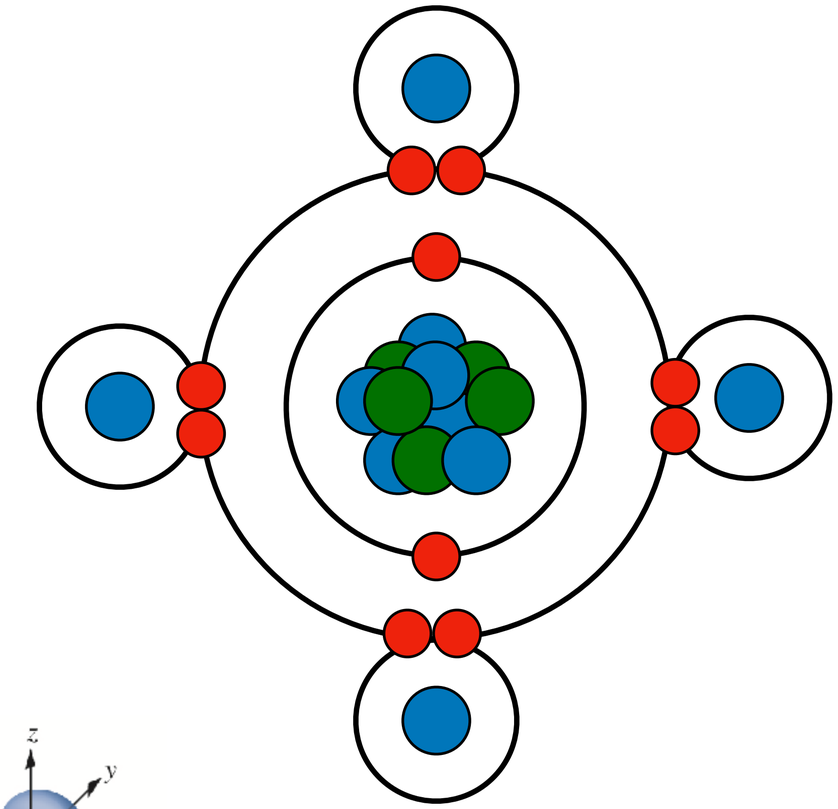
Base Principle

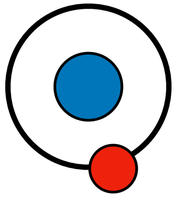
Atoms have OCD (obsessive compulsive disorder)

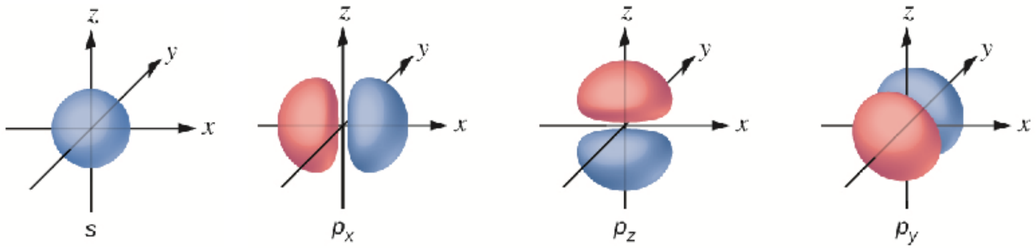

Brian Greene

Sharing is Caring

rather share electrons than having incomplete shells







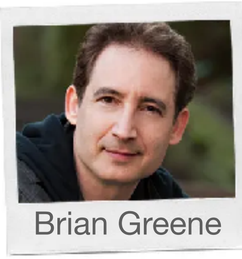
Proton

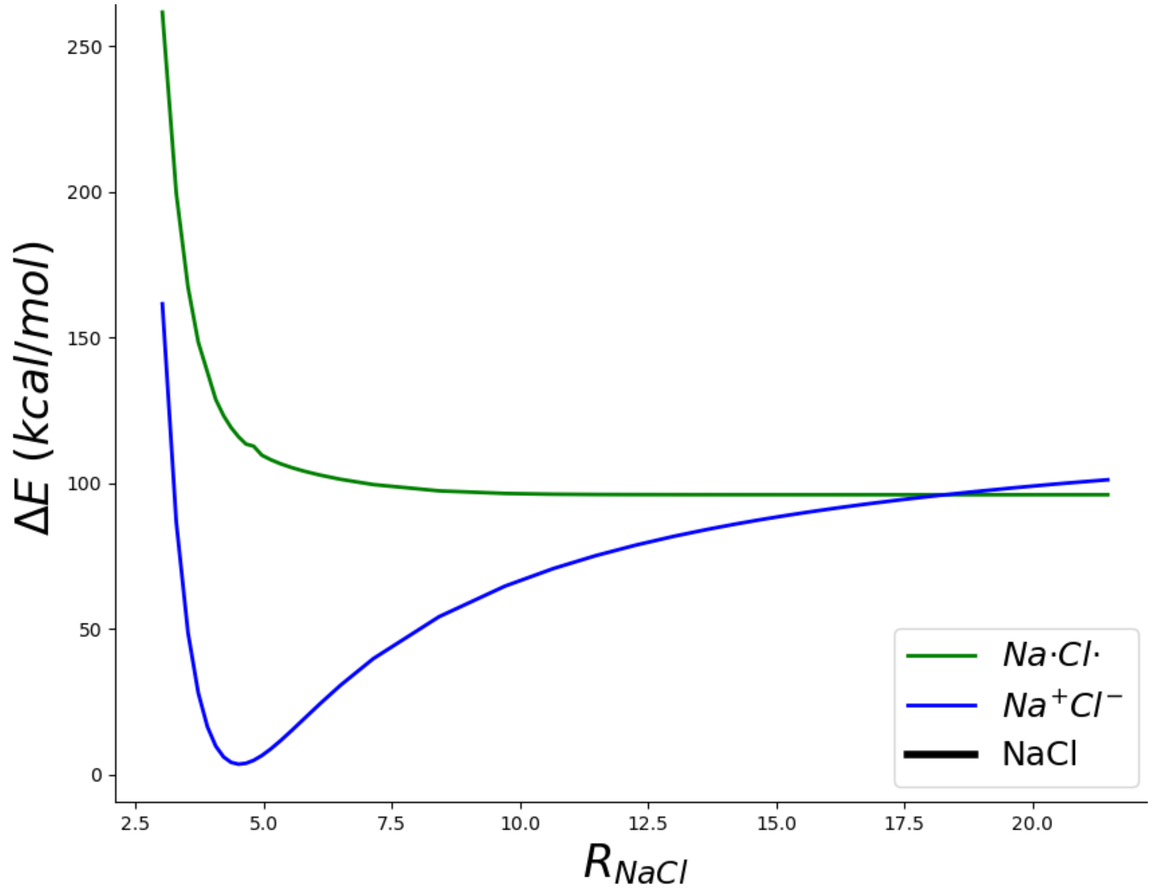
Neutron

Electron

Base Principle

Atoms have OCD (obsessive compulsive disorder)


Brian Greene



They will do anything to avoid unpaired electrons

- “Philanthropist atoms”:

$$Na \rightleftharpoons Na^{+} + e^{-}$$

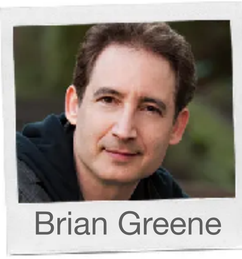
- “Looter atoms”:

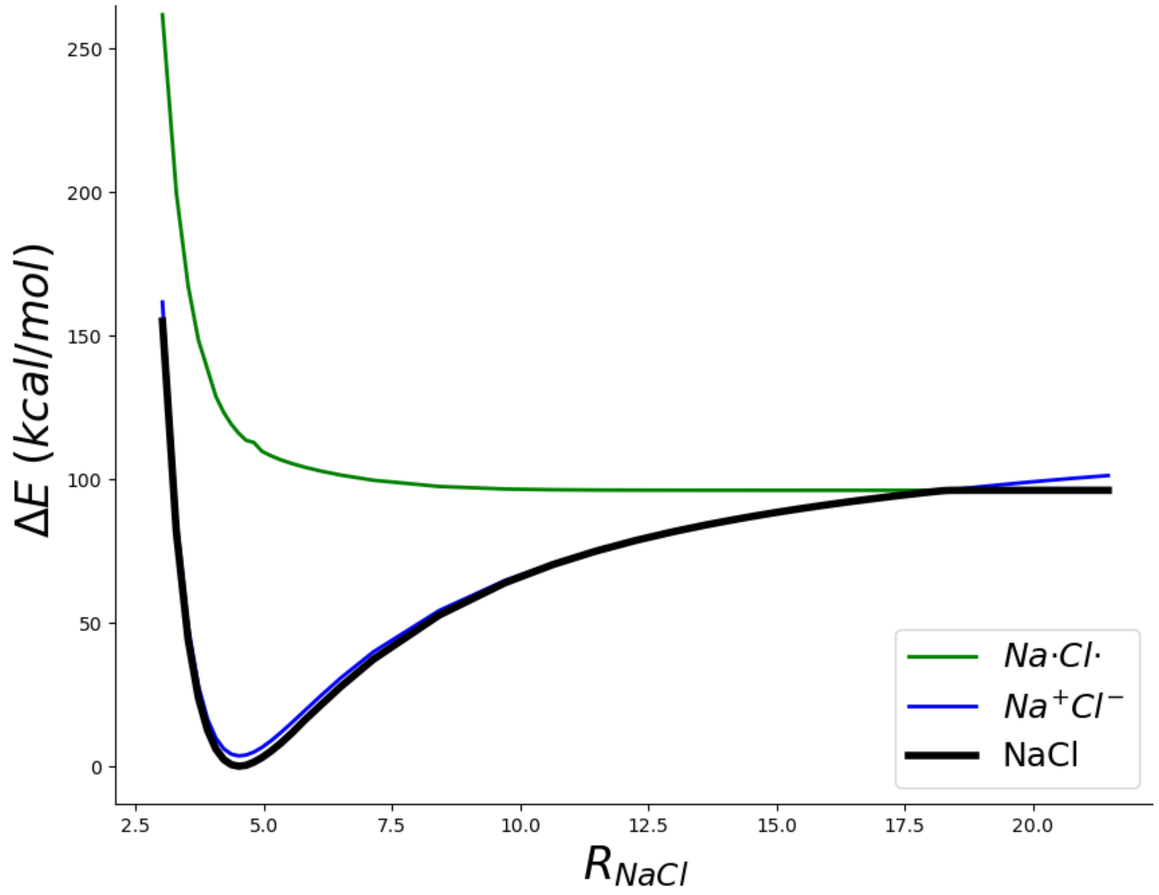
$$Cl + e^{-} \rightleftharpoons Cl^{-}$$

10.1063/1.1690232

Base Principle

Atoms have OCD (obsessive compulsive disorder)


Brian Greene



They will do anything to avoid unpaired electrons

- “Philanthropist atoms”:

$$Na \rightleftharpoons Na^{+} + e^{-}$$

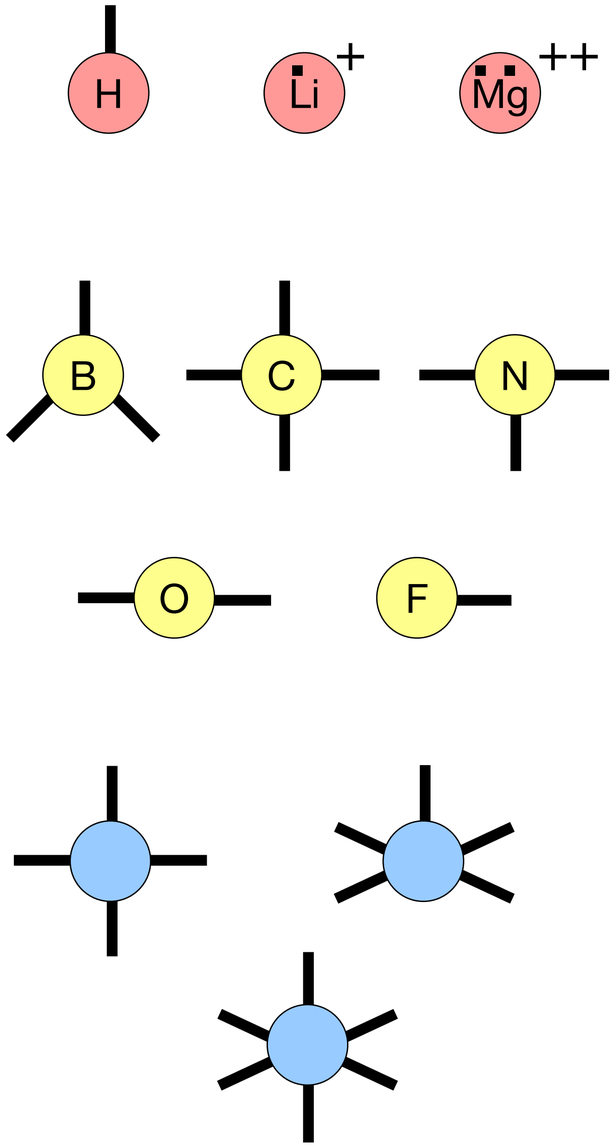
- “Looter atoms”:

$$Cl + e^{-} \rightleftharpoons Cl^{-}$$

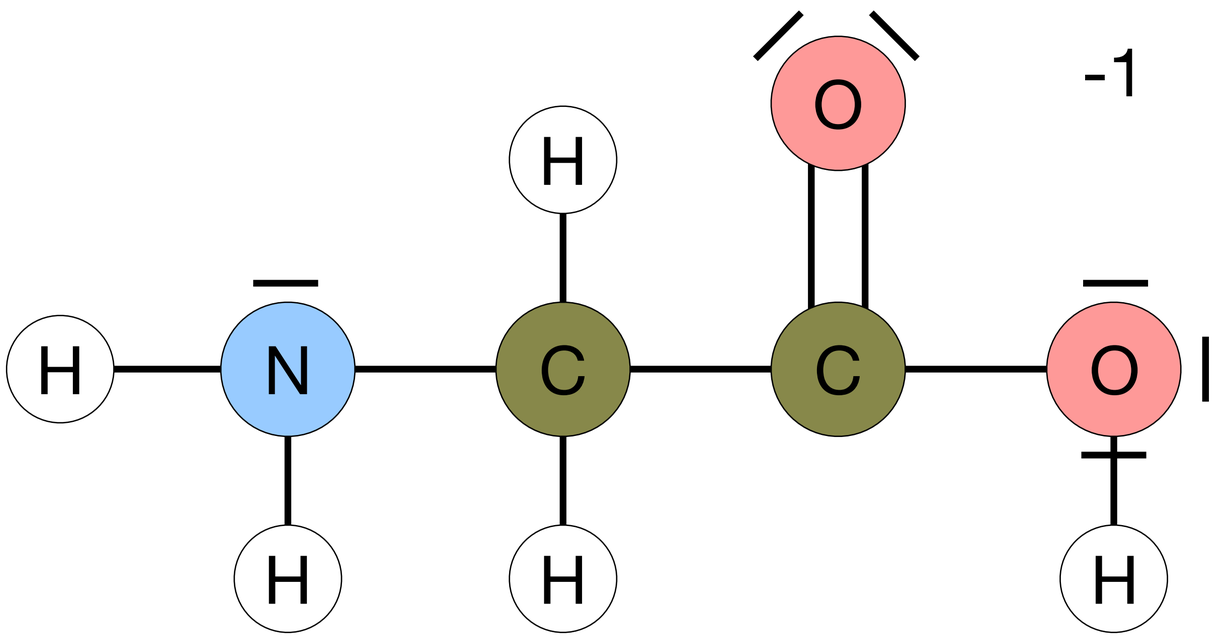
10.1063/1.1690232

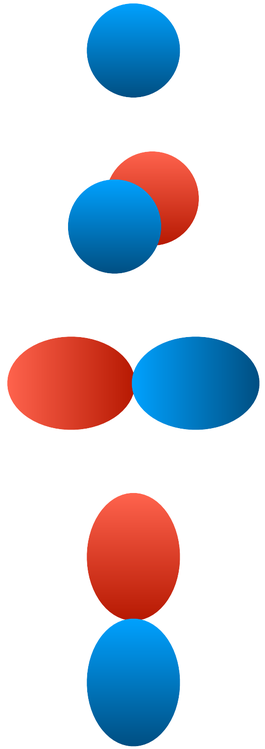
Octet And 18-Electron Rules

Group → / Period ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	* 71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	* 103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
			* 57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb		
			* 89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No		

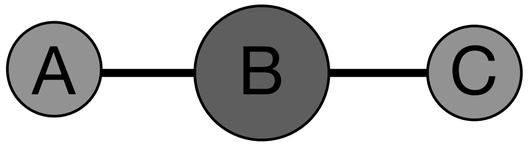


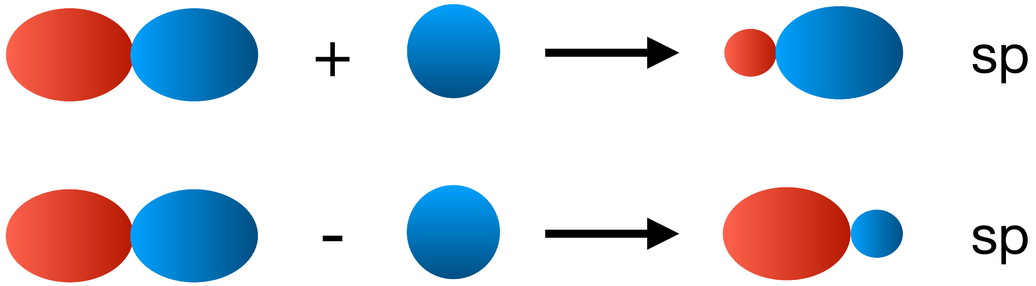
Lewis Structures

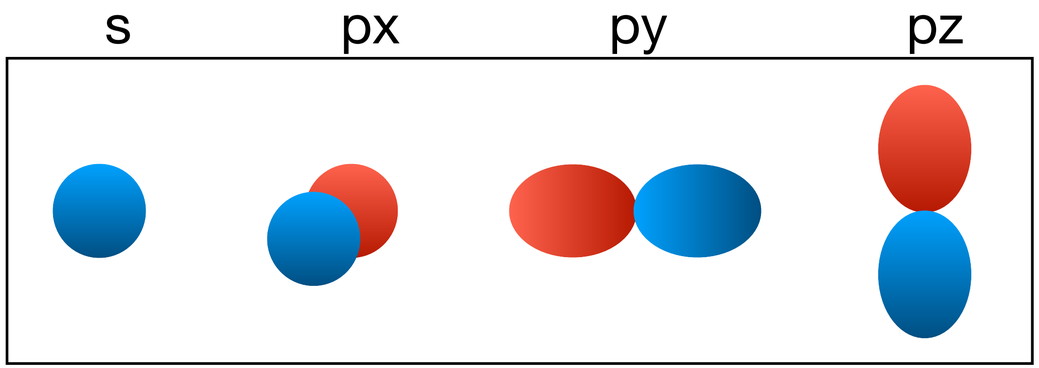




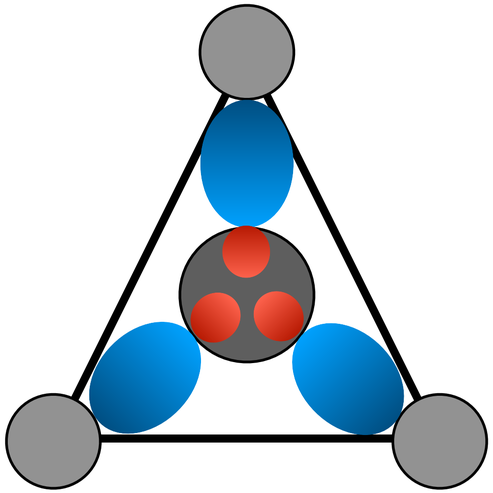
Orbital Hybridization



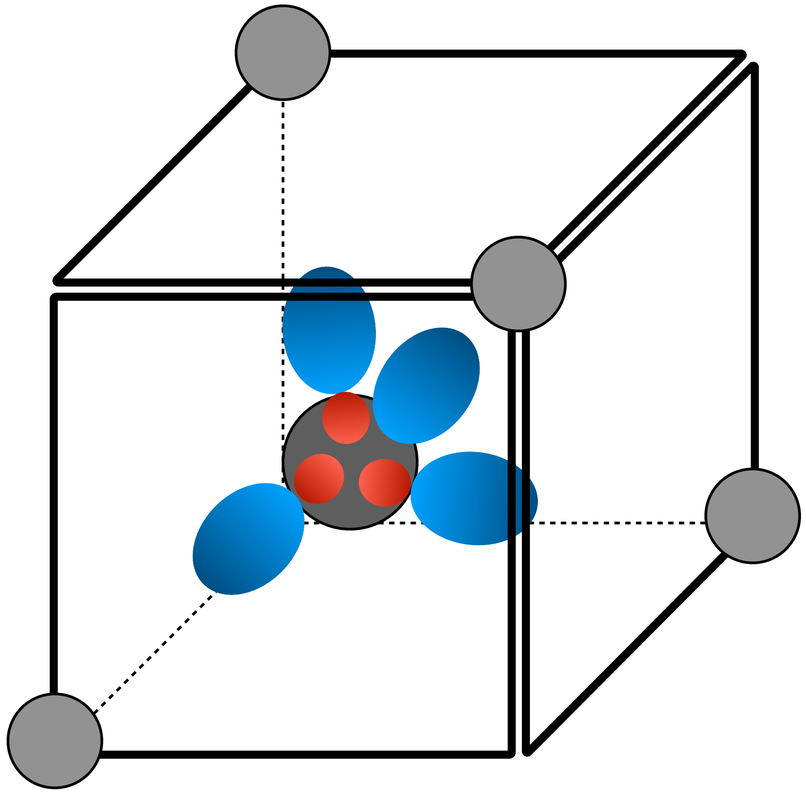




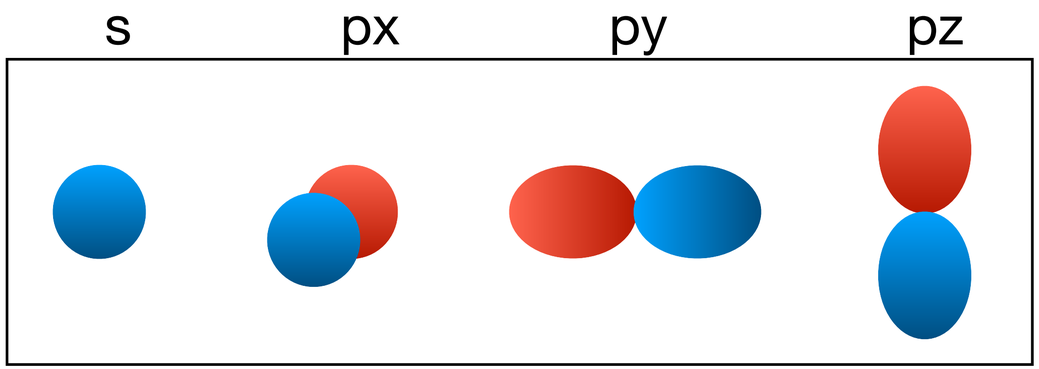
Orbital Hybridization



sp2

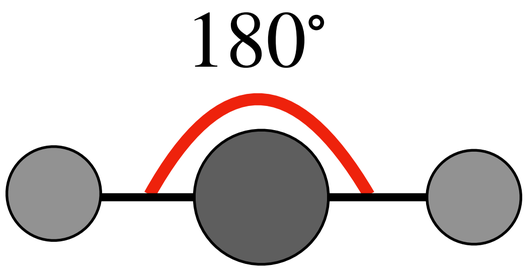


sp3

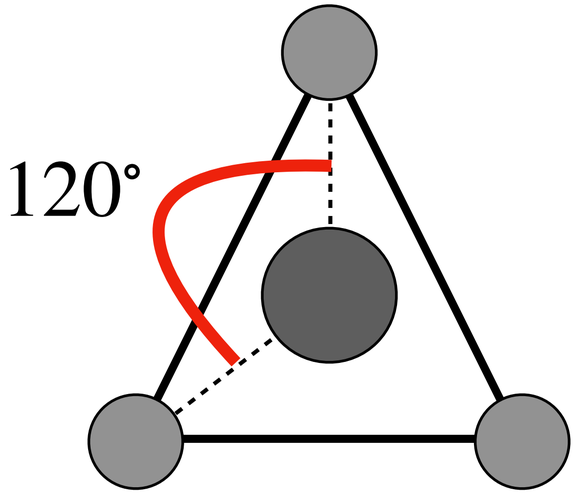


Orbital Hybridization

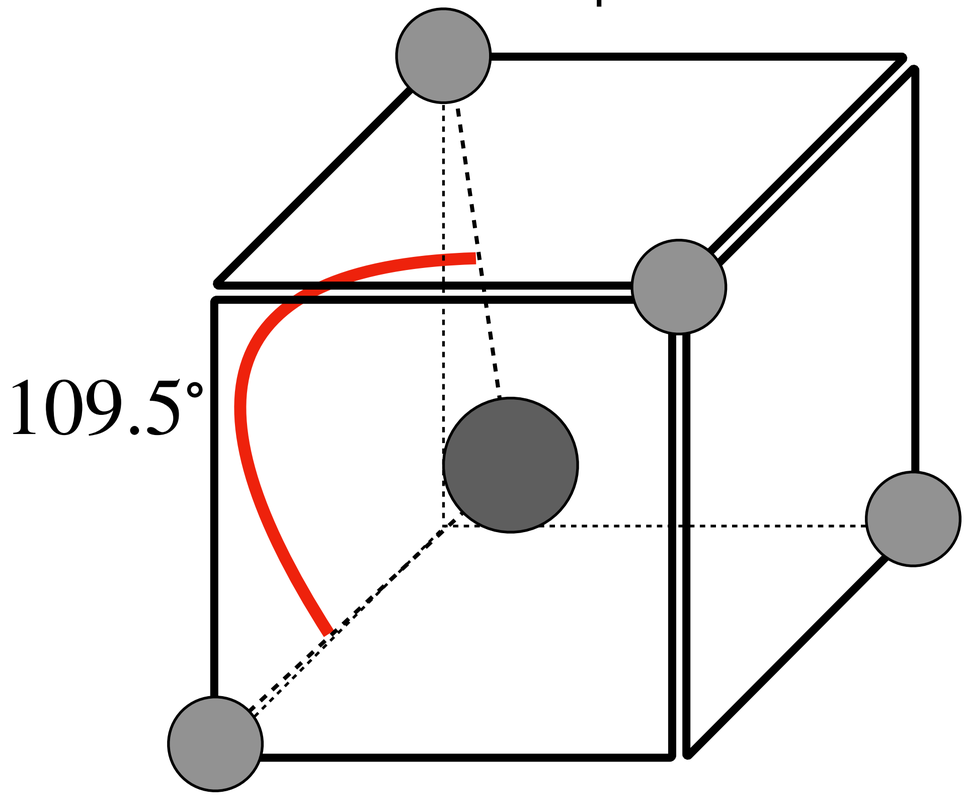
sp



sp2

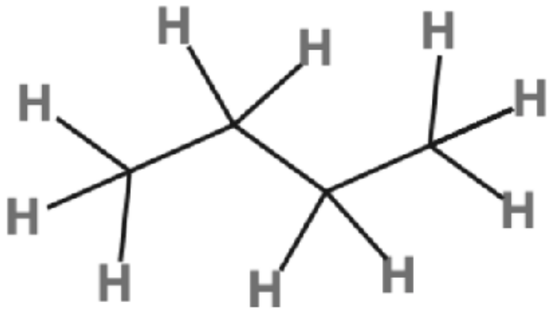


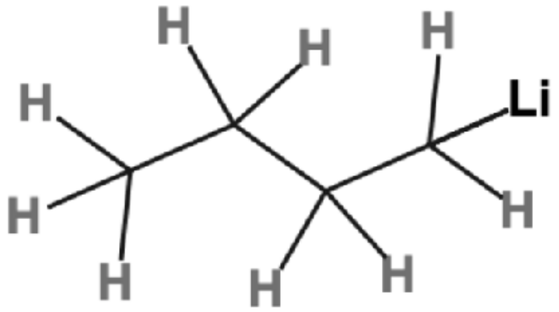
sp3

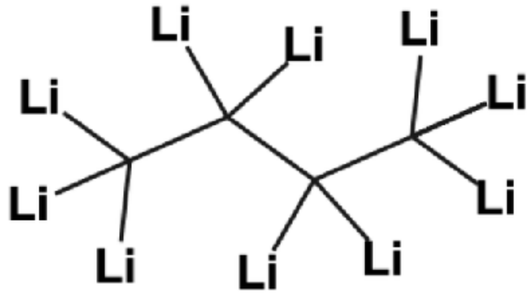


To Make It Complicated:

Atoms are picky with whom they hang out





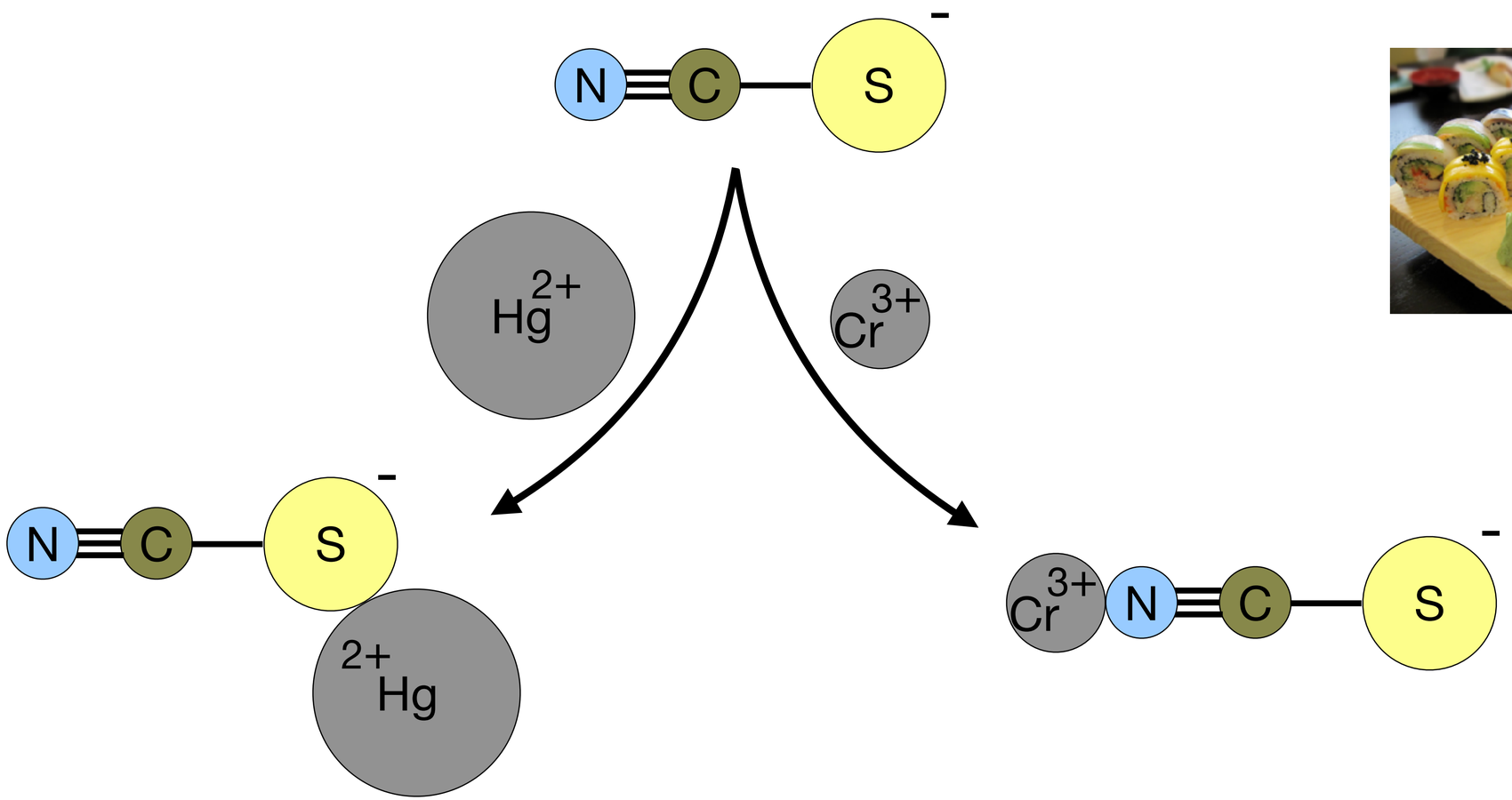


Link

Link

Hard-Soft Acid-Base

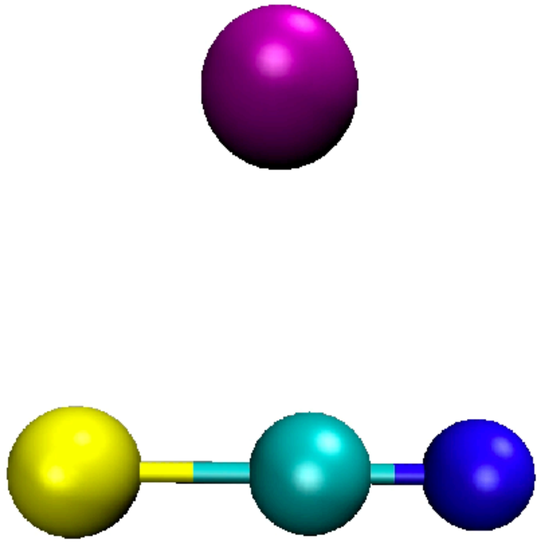
Atoms are picky with whom they hang out



Link

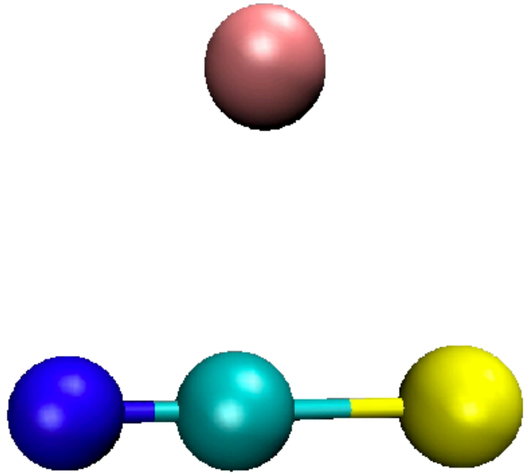
Hard-Soft Acid-Base

Atoms are picky with whom they hang out



$$(s_N | s_{Hg}) = 0.18$$

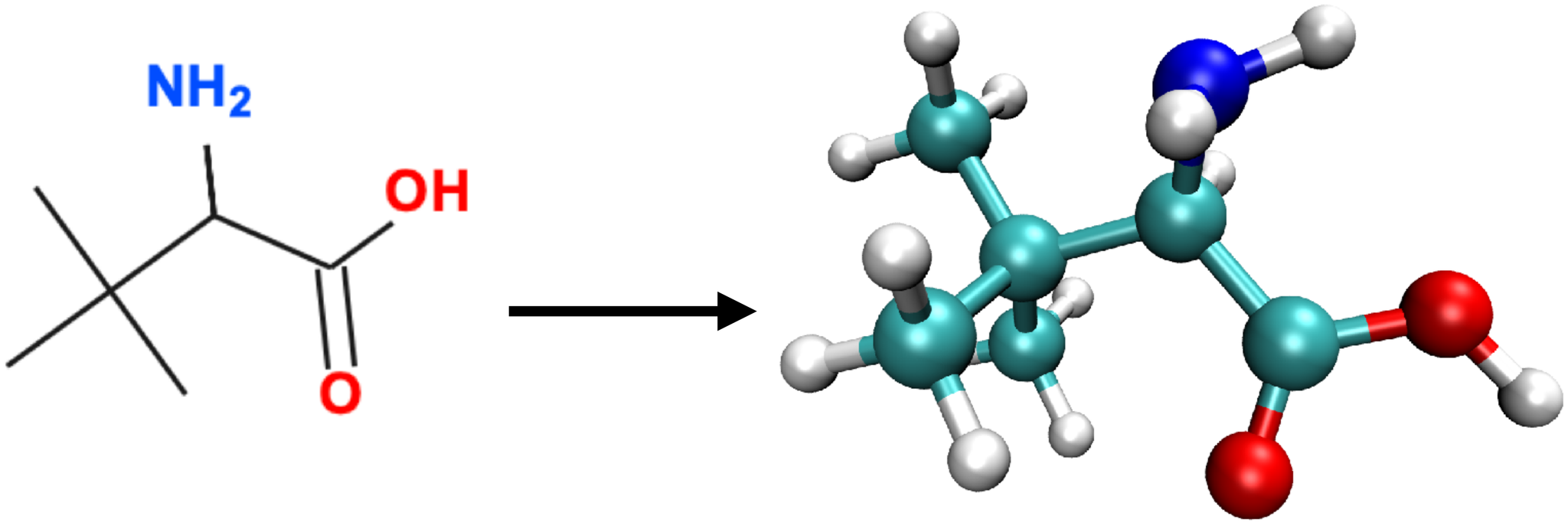
$$(s_S | s_{Hg}) = 0.22$$



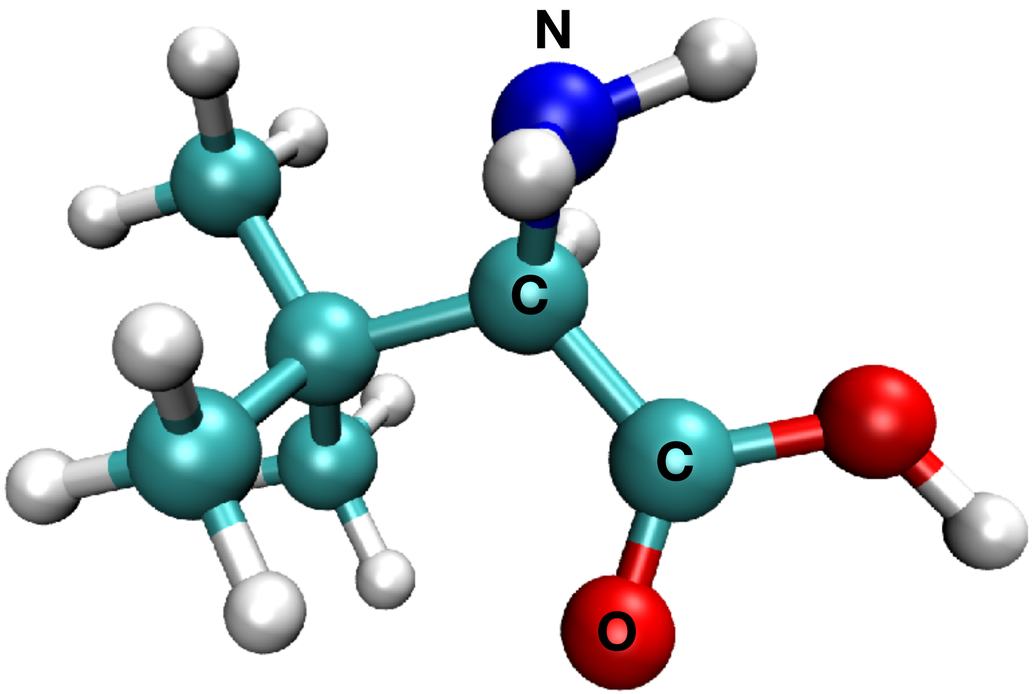
$$(s_N | s_{Cr}) = 0.30$$

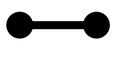
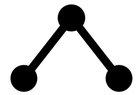
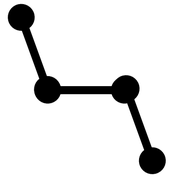
$$(s_S | s_{Cr}) = -$$

From 2D to 3D

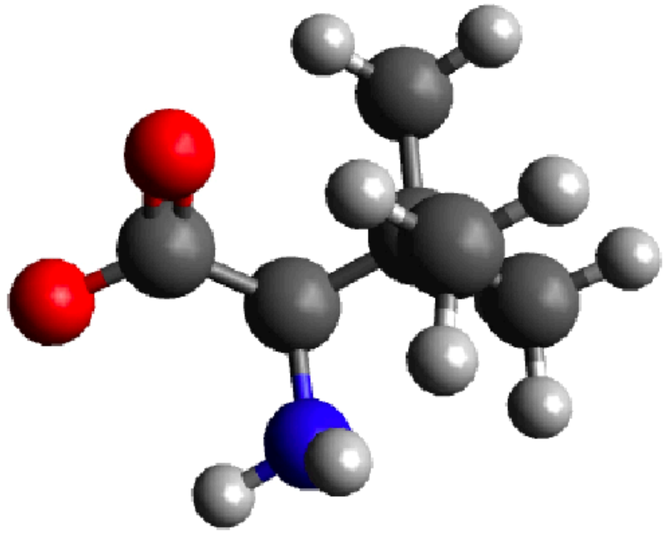


Z-Matrix



El	#		#		#	
C	1					
O	1	1.20				
C	1	1.55	2	120		
N	3	1.45	1	109.5	2	120
…						

Z-Matrix



C	1					
O	1	1.20				
C	1	1.55	2	120.0		
N	3	1.45	1	109.5	2	120.0
C	3	2.50	1	150.0	2	360.0
C	5	1.55	3	35.0	1	0.0
H	5	1.10	3	90.0	1	230.0
H	5	1.10	3	90.0	1	130.0
H	5	1.10	3	150.0	1	0.0
C	6	1.55	5	109.5	3	240.0
C	6	1.55	5	109.5	3	120.0
H	10	1.10	6	109.5	5	60.0
H	10	1.10	6	109.5	5	290.0
H	10	1.10	6	109.5	5	180.0
H	11	1.10	6	109.5	5	290.0
H	11	1.10	6	109.5	5	180.0
H	11	1.10	6	109.5	5	60.0
H	3	1.10	1	109.5	2	240.0
O	1	1.35	2	120.0	3	180.0
H	4	1.10	3	109.5	1	290.0
H	4	1.10	3	109.5	1	60.0
H	19	0.10	1	120.0	2	0.0

Link

Automated 3D-Structure Generation

RDKit

Open-Source Cheminformatics
and Machine Learning

c1ccccc1

↓

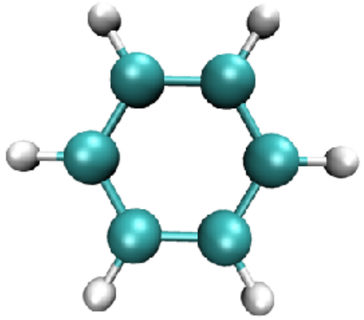
Use rules above to connect atoms

↓

(use fragments to simplify)

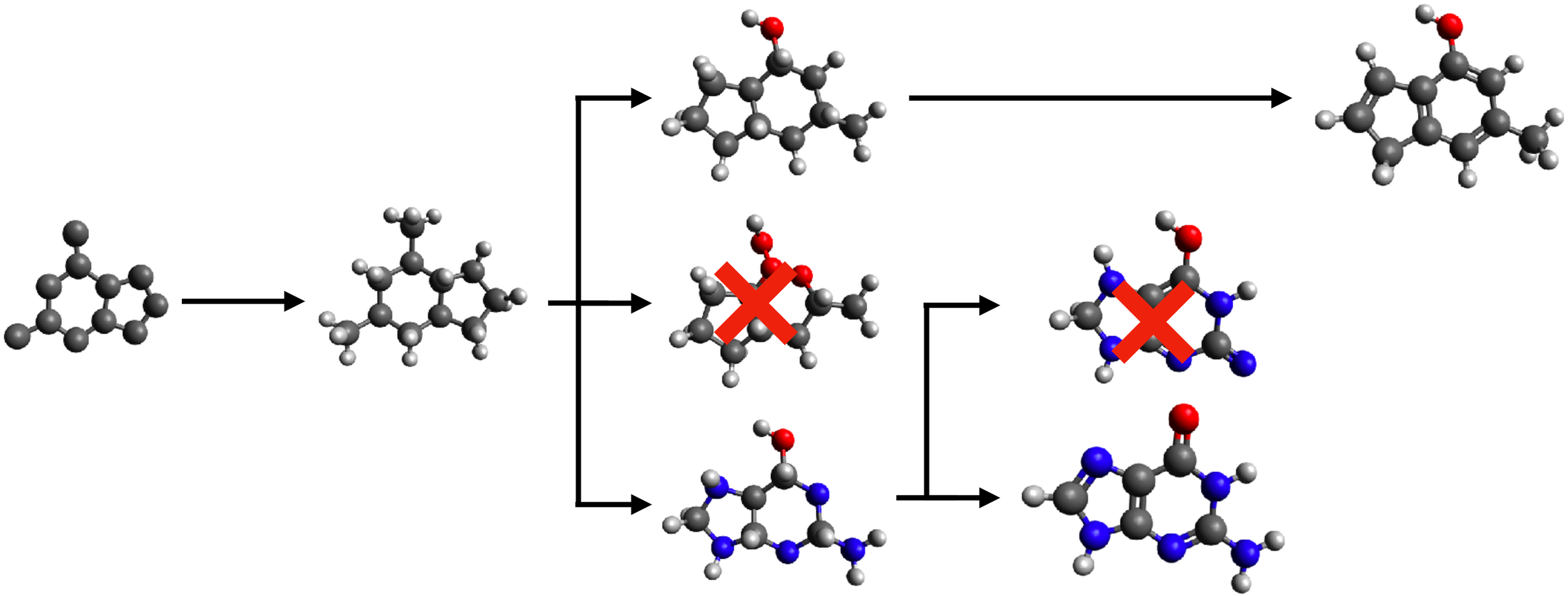
↓

Optimize the Geometry (FF or QM) →

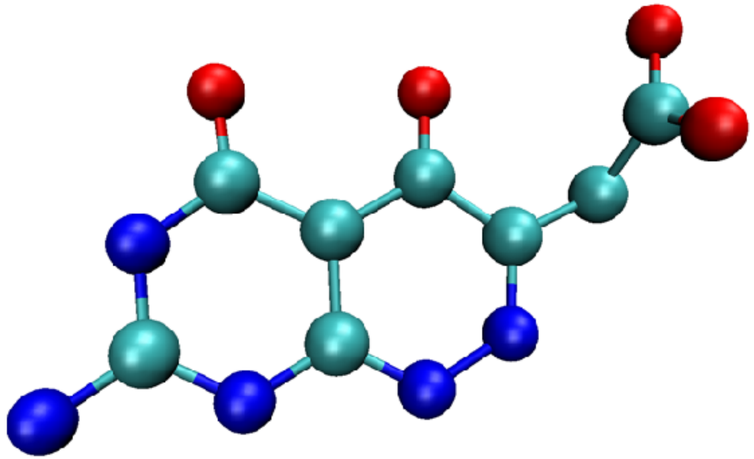


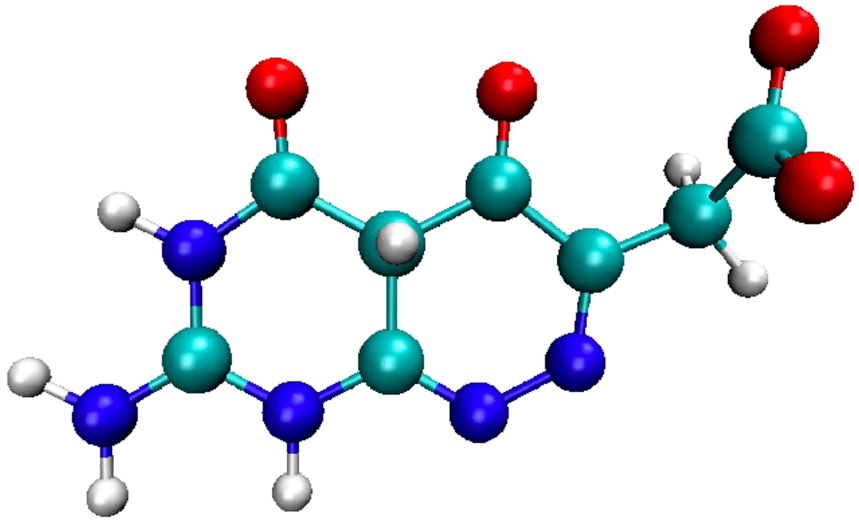
Automated 3D-Structure Generation

QUANTUM-MACHINE.ORG

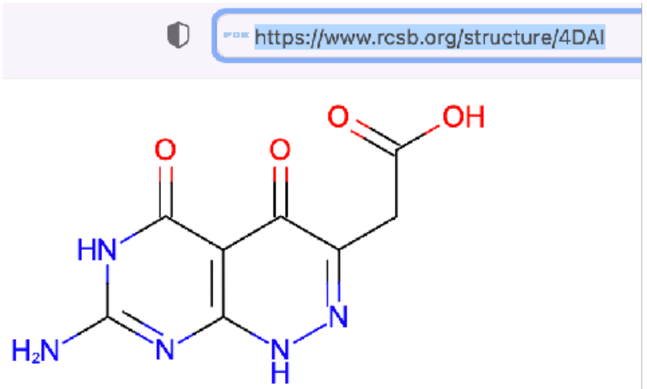


What can go wrong… will go wrong

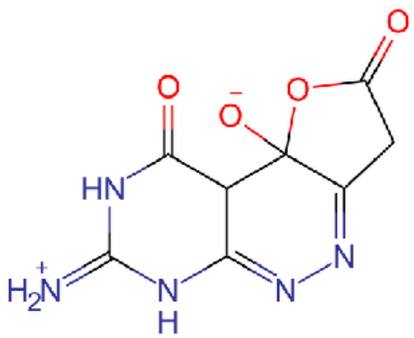




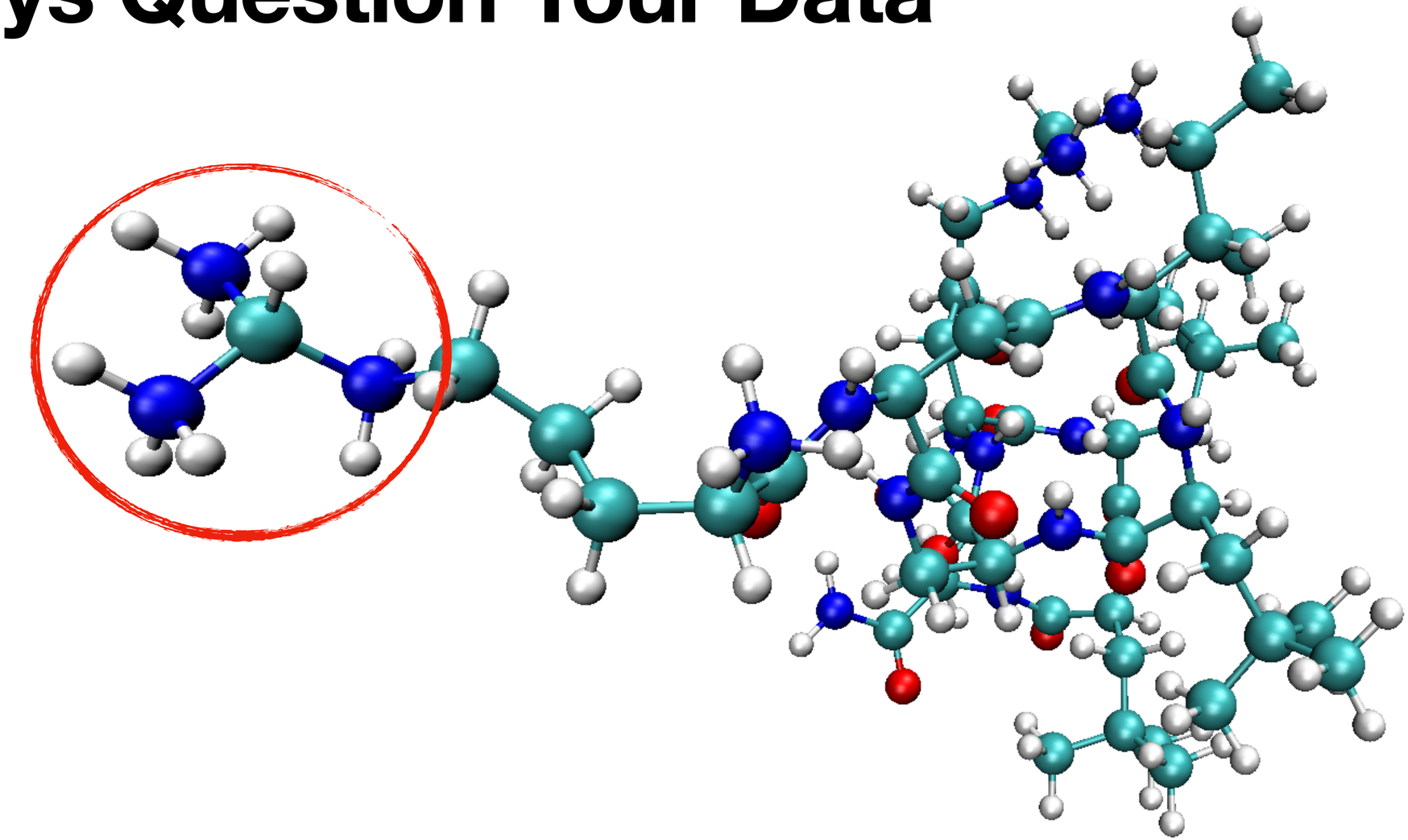
PDBbind



4DAI

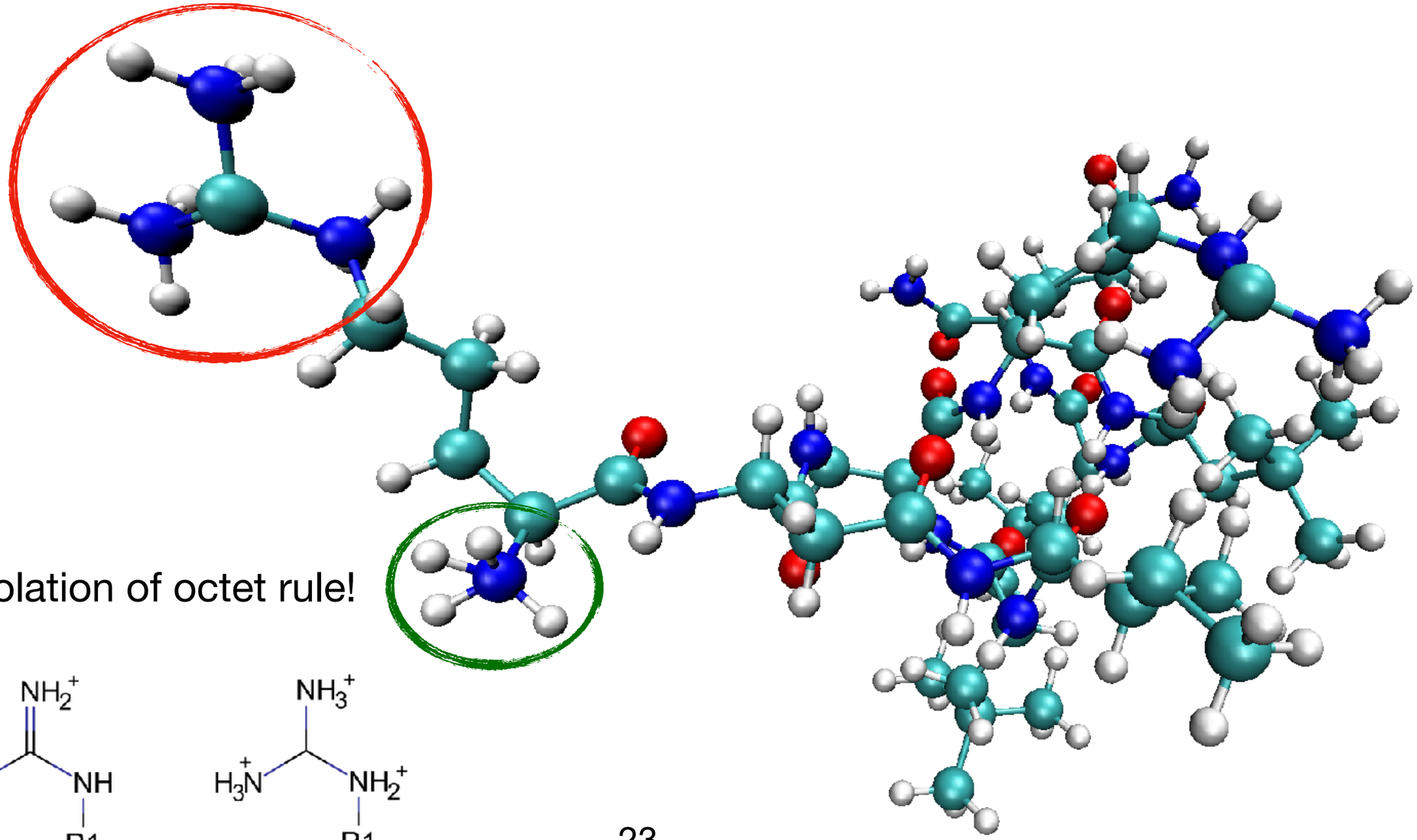


Always Question Your Data



5GTR

So fix it with widely used software

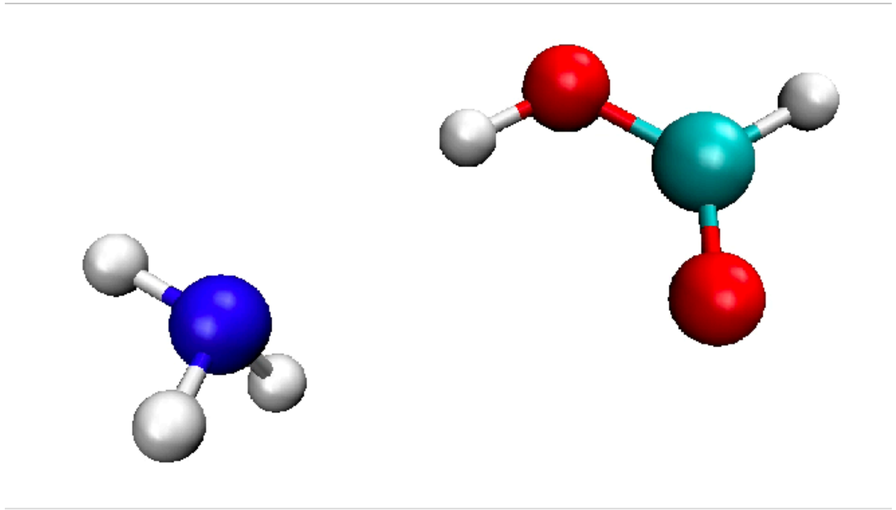


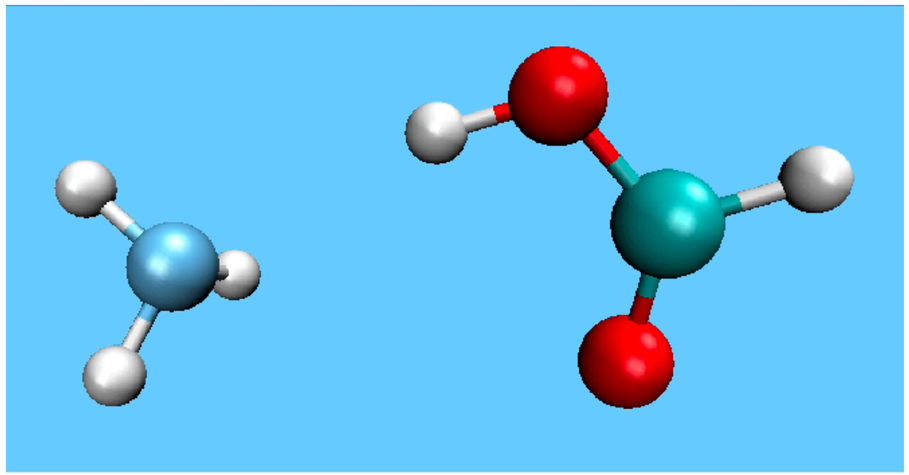
Explicit violation of octet rule!

5GTR

Tipps and Tricks

- Follow the rules above and the structure will be reasonable (at least in gas).
- Some groups like to be charged. Use pKa and pKb tables.

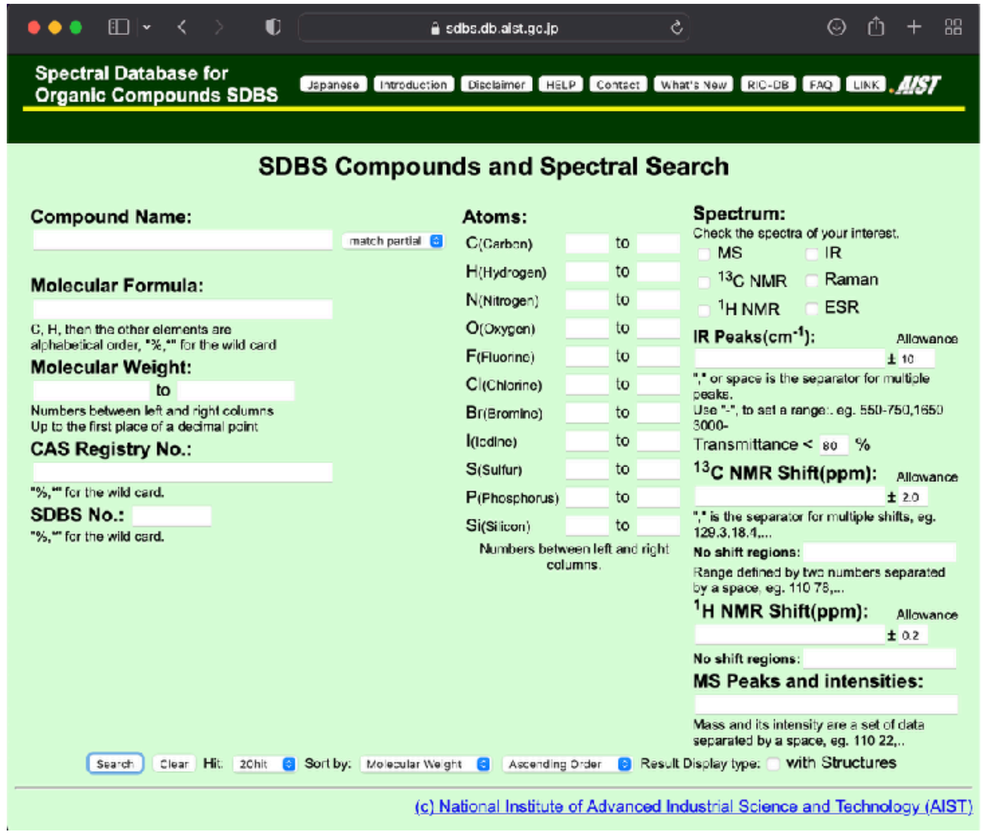




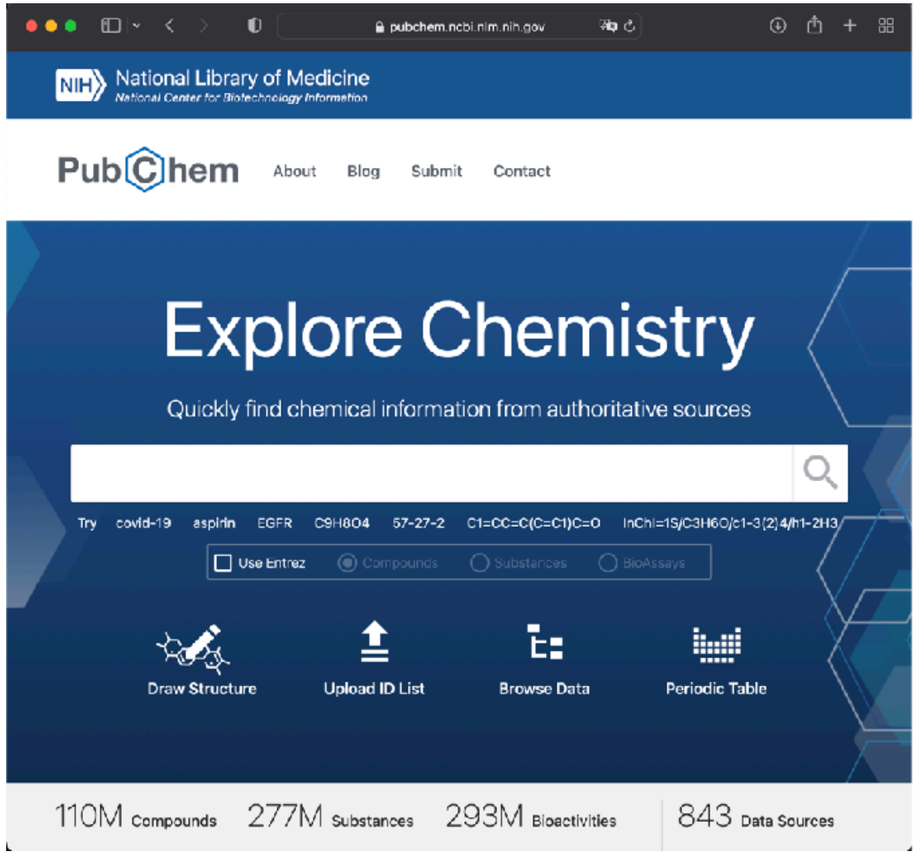
- Rather consistently wrong, than sometimes (inconsistently) correct.

Tipps and Tricks

- Use Databases for double-checking



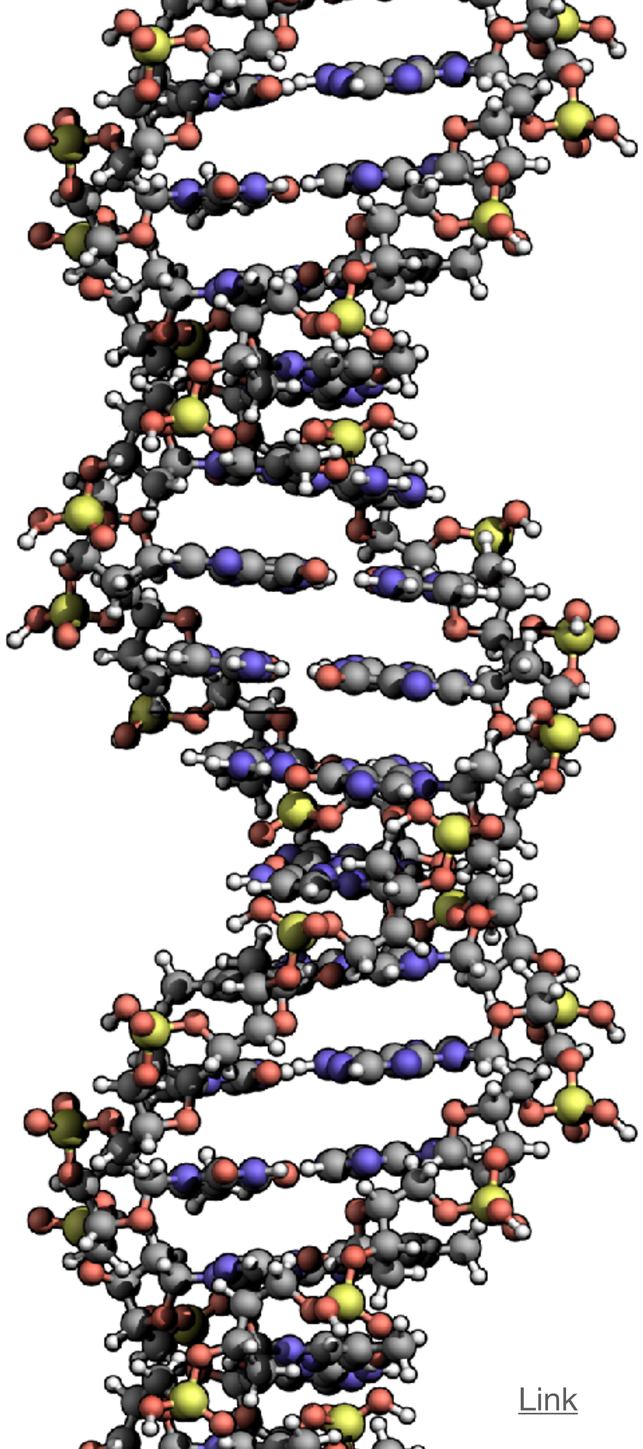
Link



Link

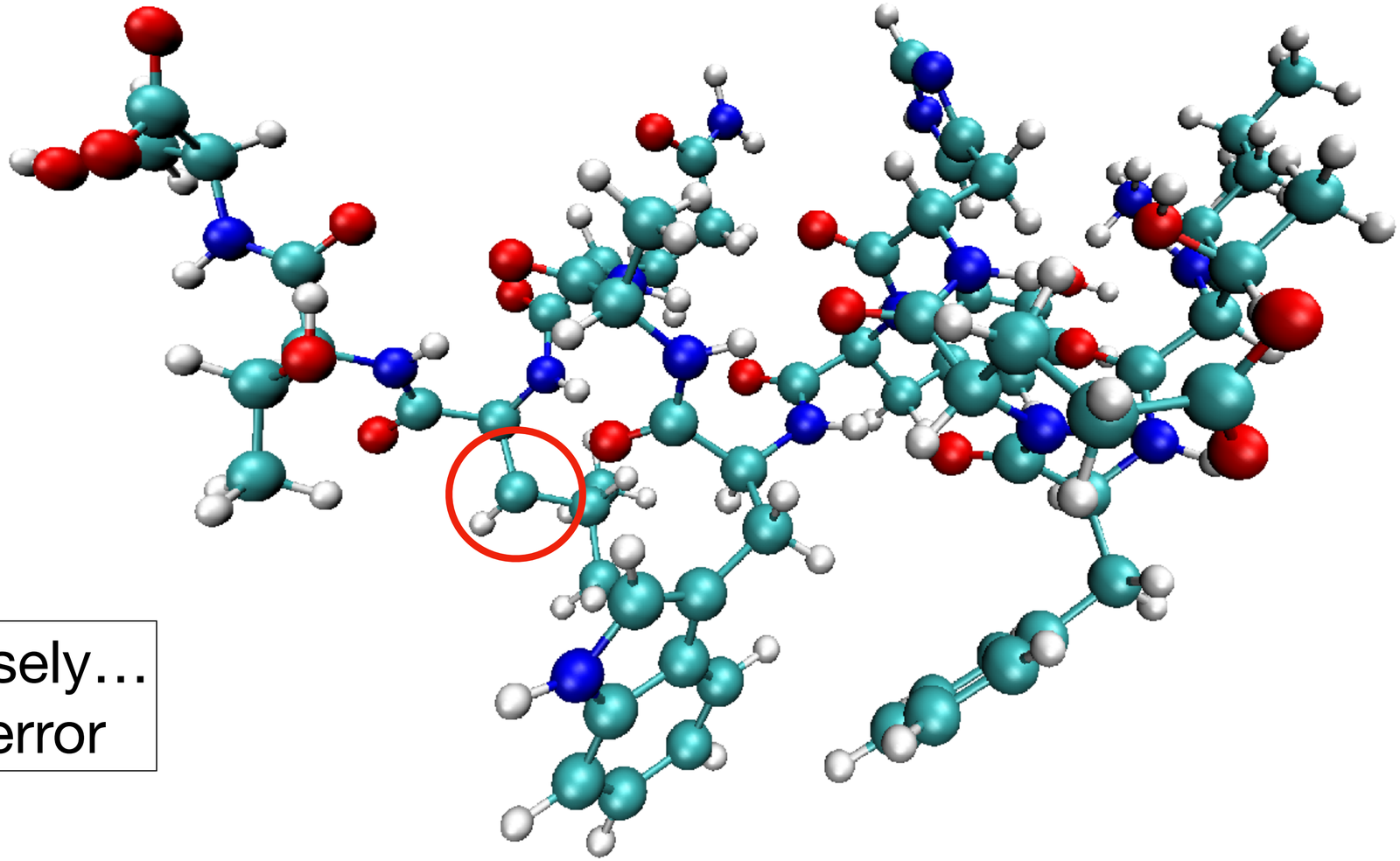
Part 2

Structure Refinement



Link

How to find problems in structures 101

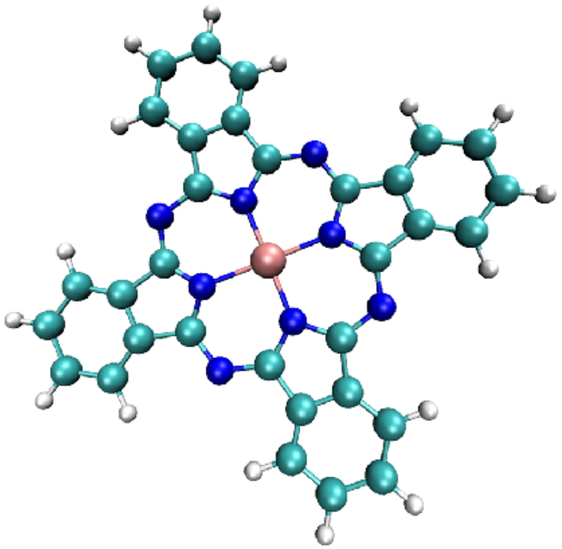


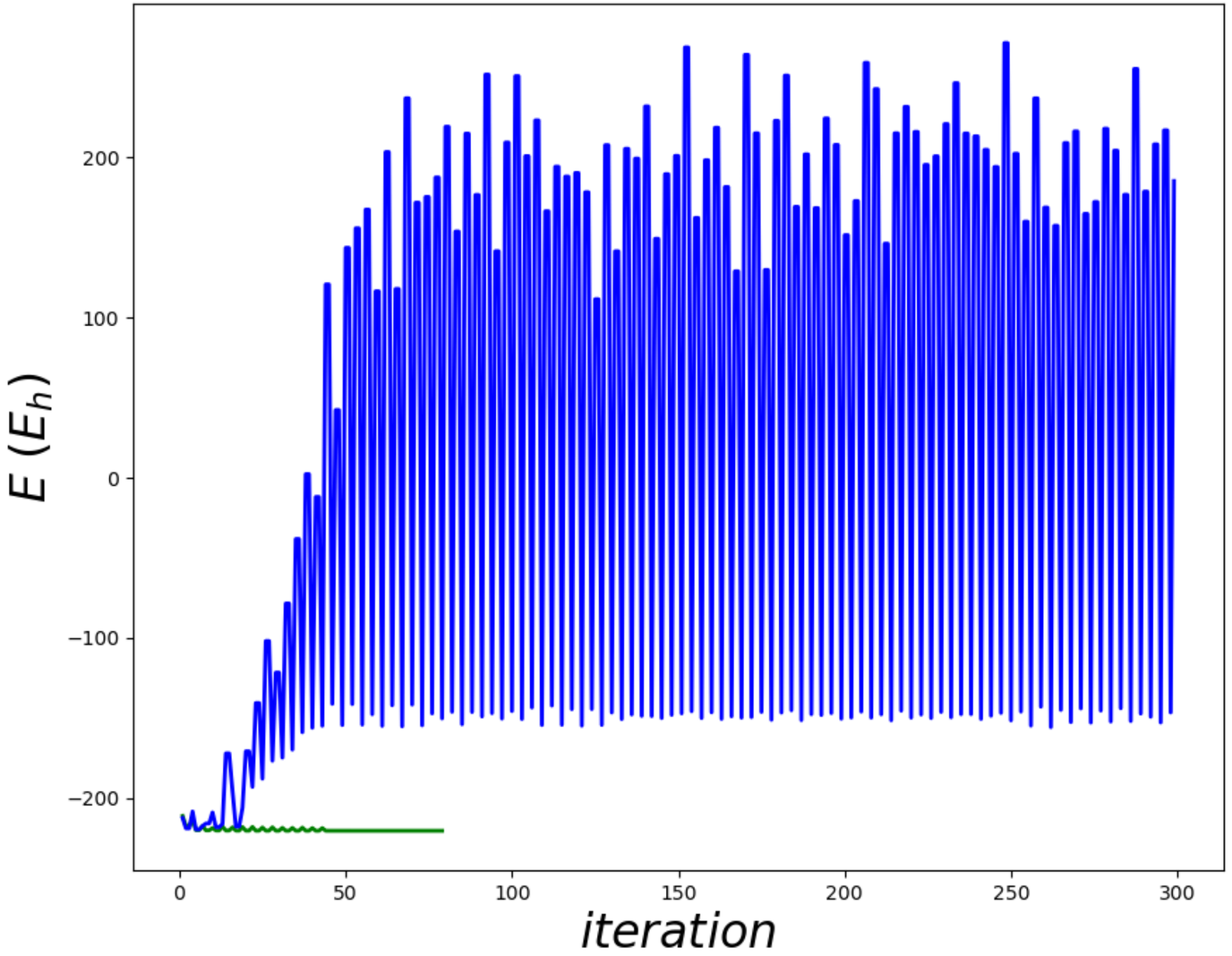
You look very closely…
and you find the error

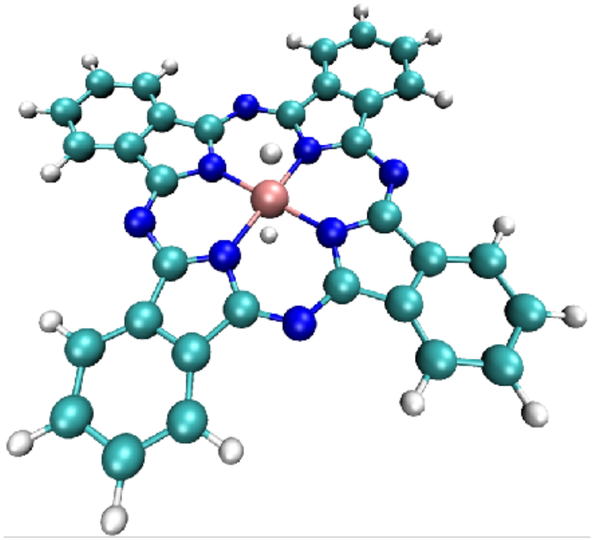
If You don’t Know, Ask Physics

- Physically wrong structures (and electron counting) affects:
    - Force Field Parametrization (fails).
    - Quantum Mechanics convergence (zero to none).
- But also good structures have problems:
    - Dancing Charge problem.
- Population Analysis.
- Geometry Optimization or Dynamics.

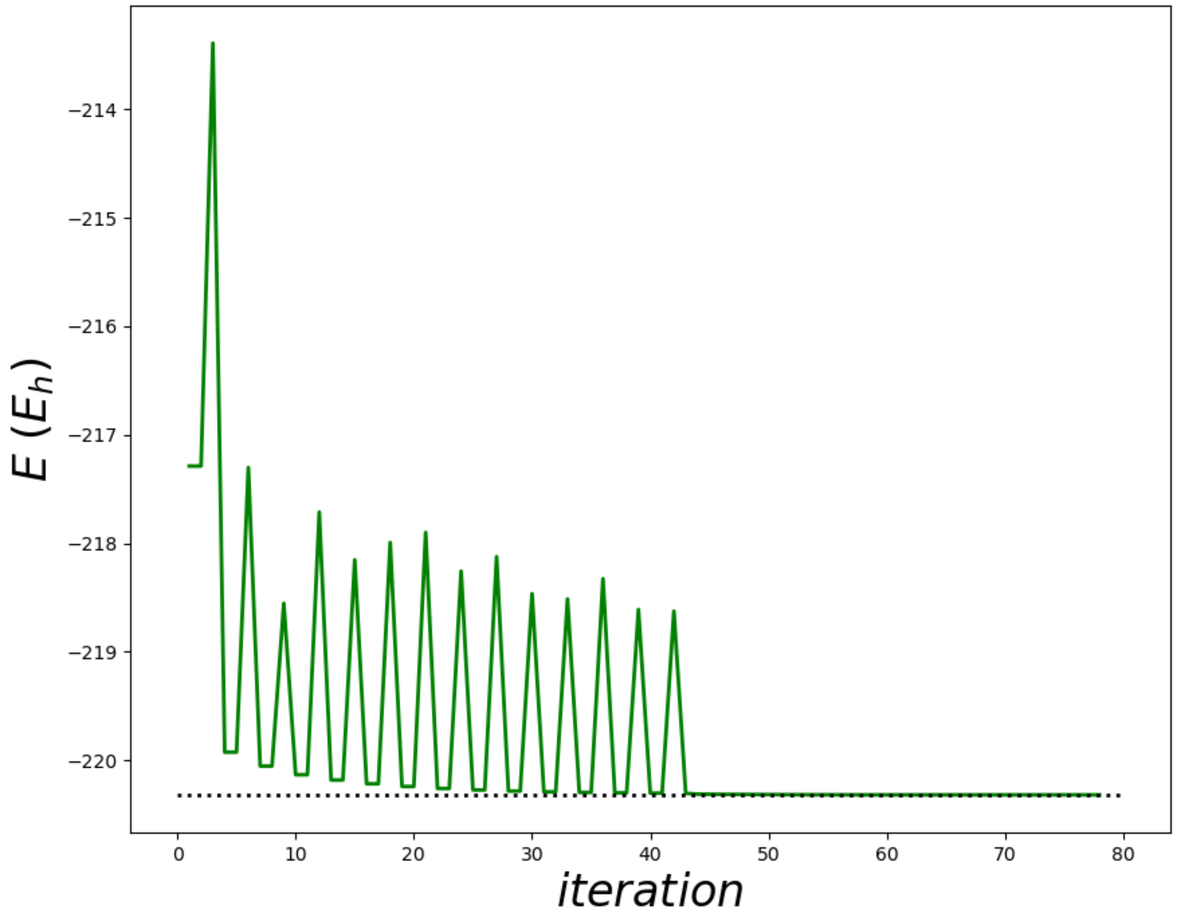
Example of Convergence of Wavefunction

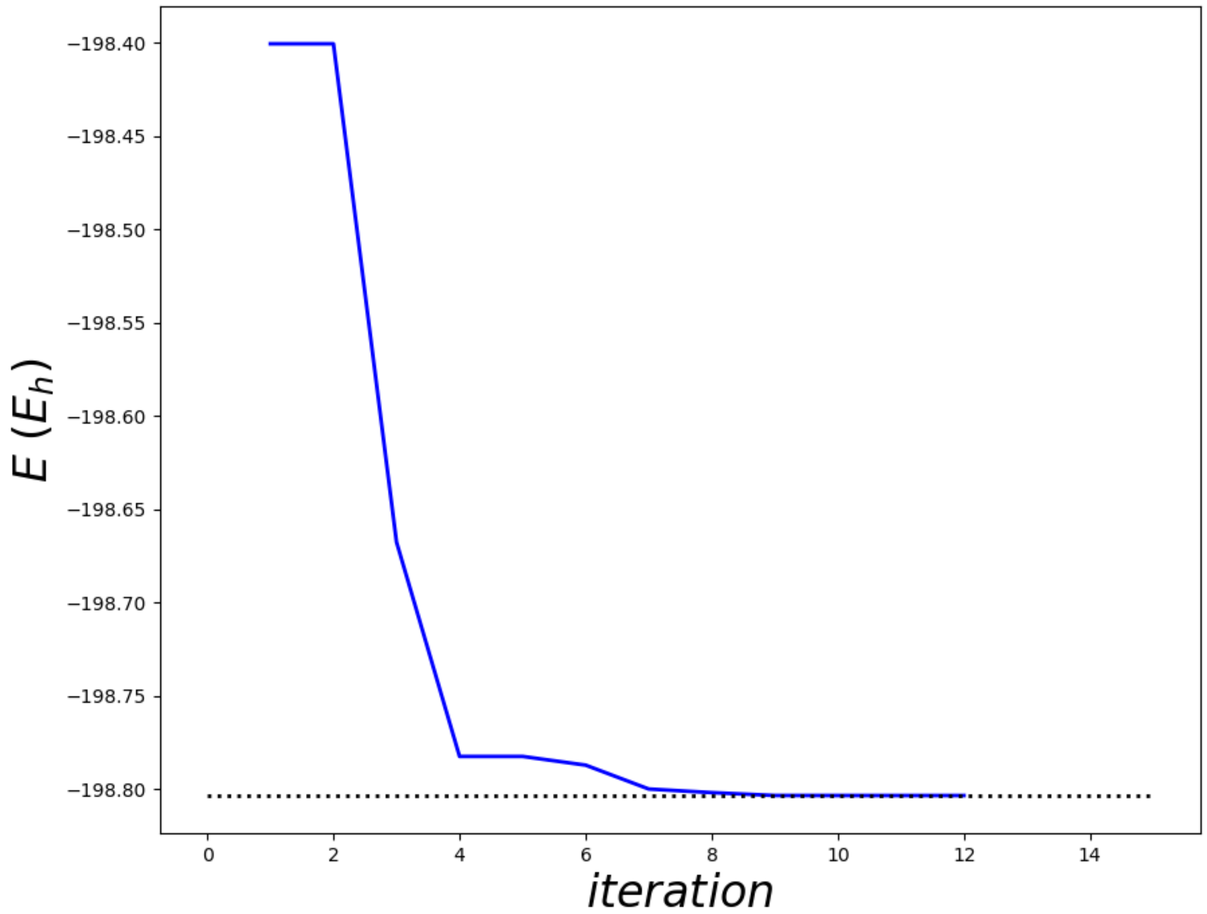




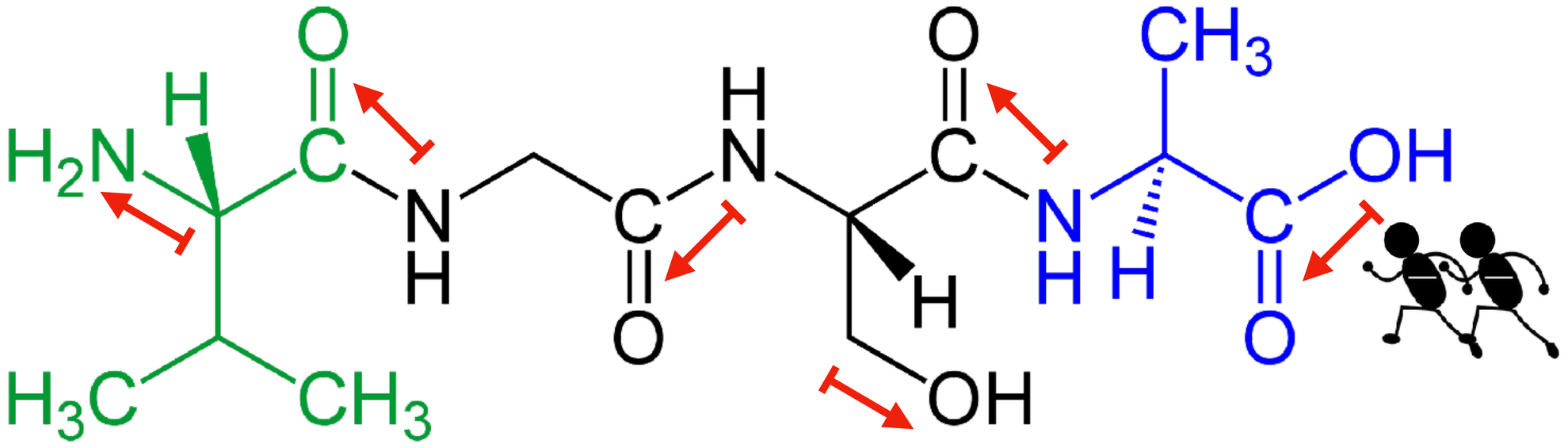


Example of Convergence of Wavefunction



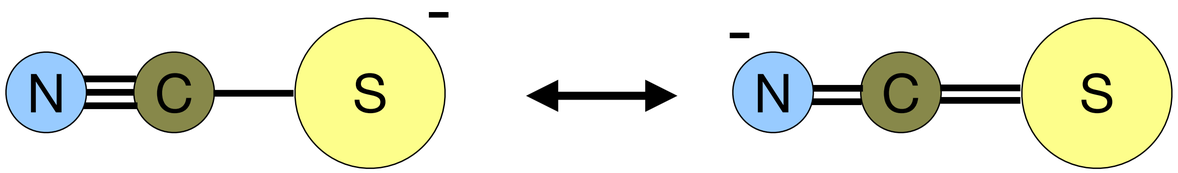


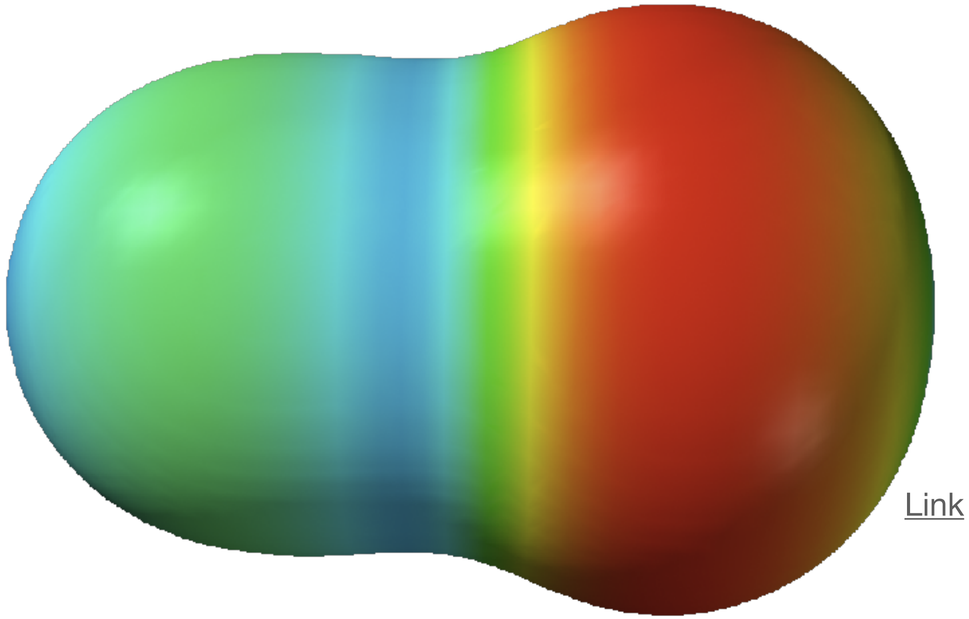
Peptides and Dancing Charges



Link

Population Analysis and Bond Orders

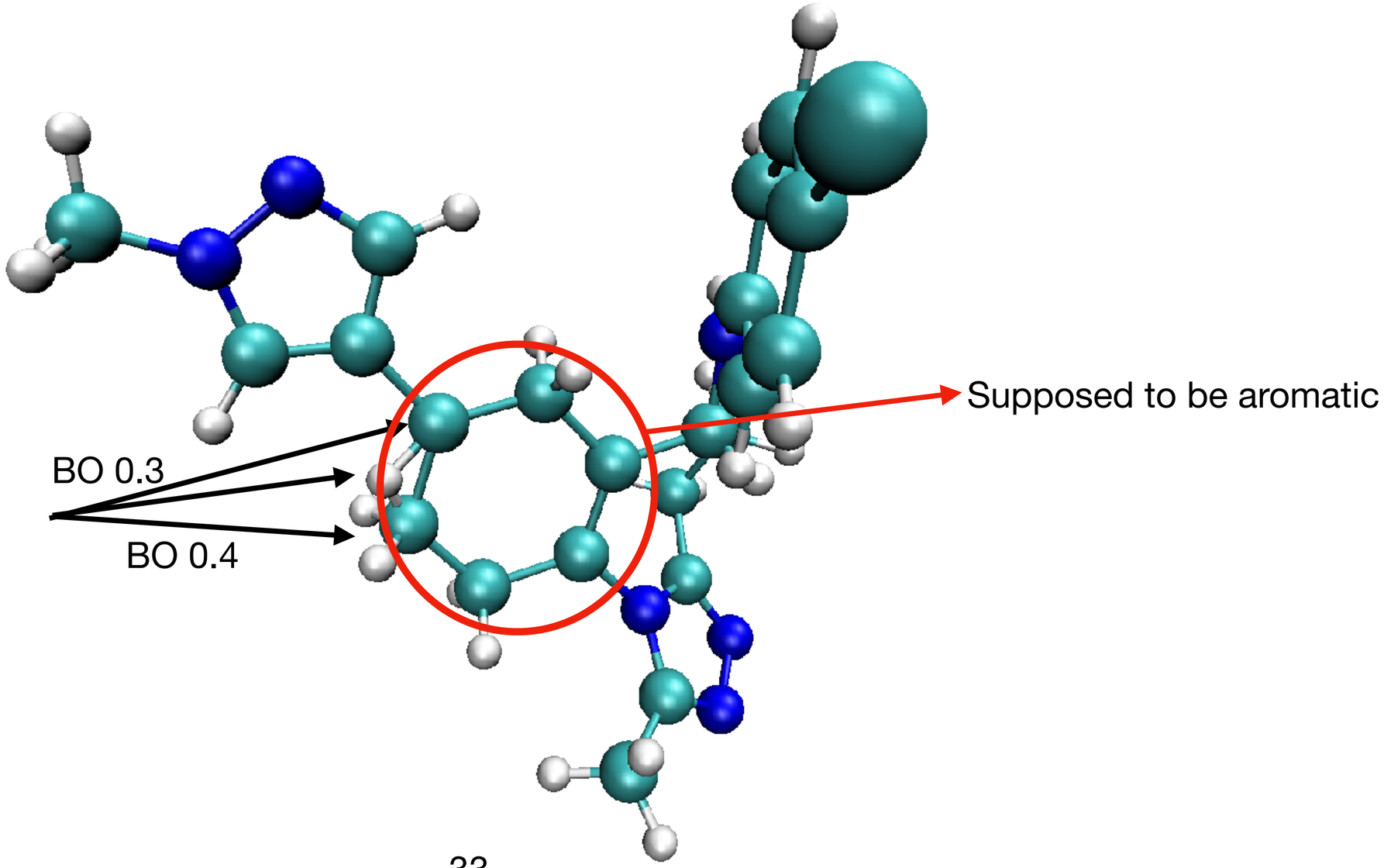




Link

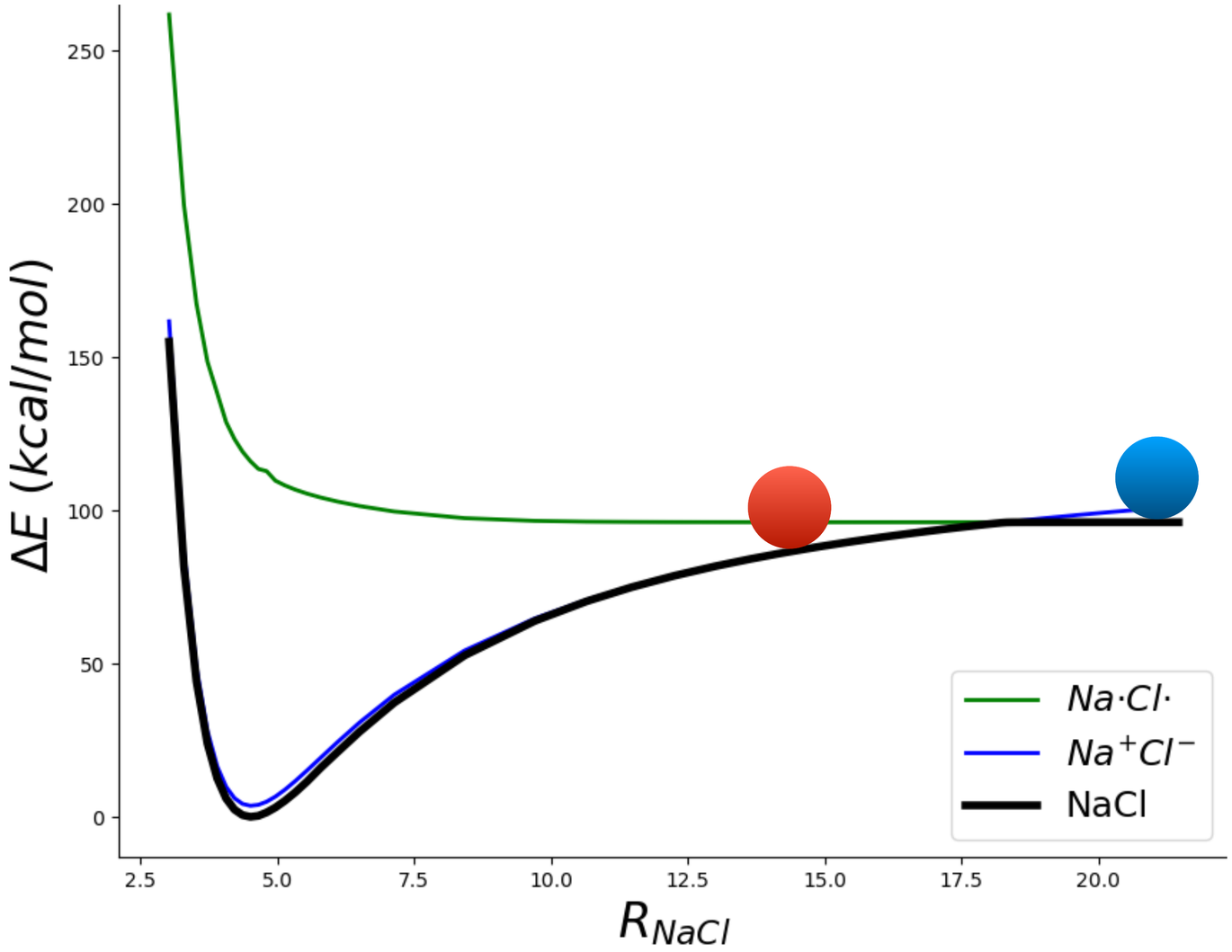
	N	C	S
N	0.000	2.555	0.357
C	2.555	0.000	1.385
S	0.357	1.385	0.000

Bond Order Matrices

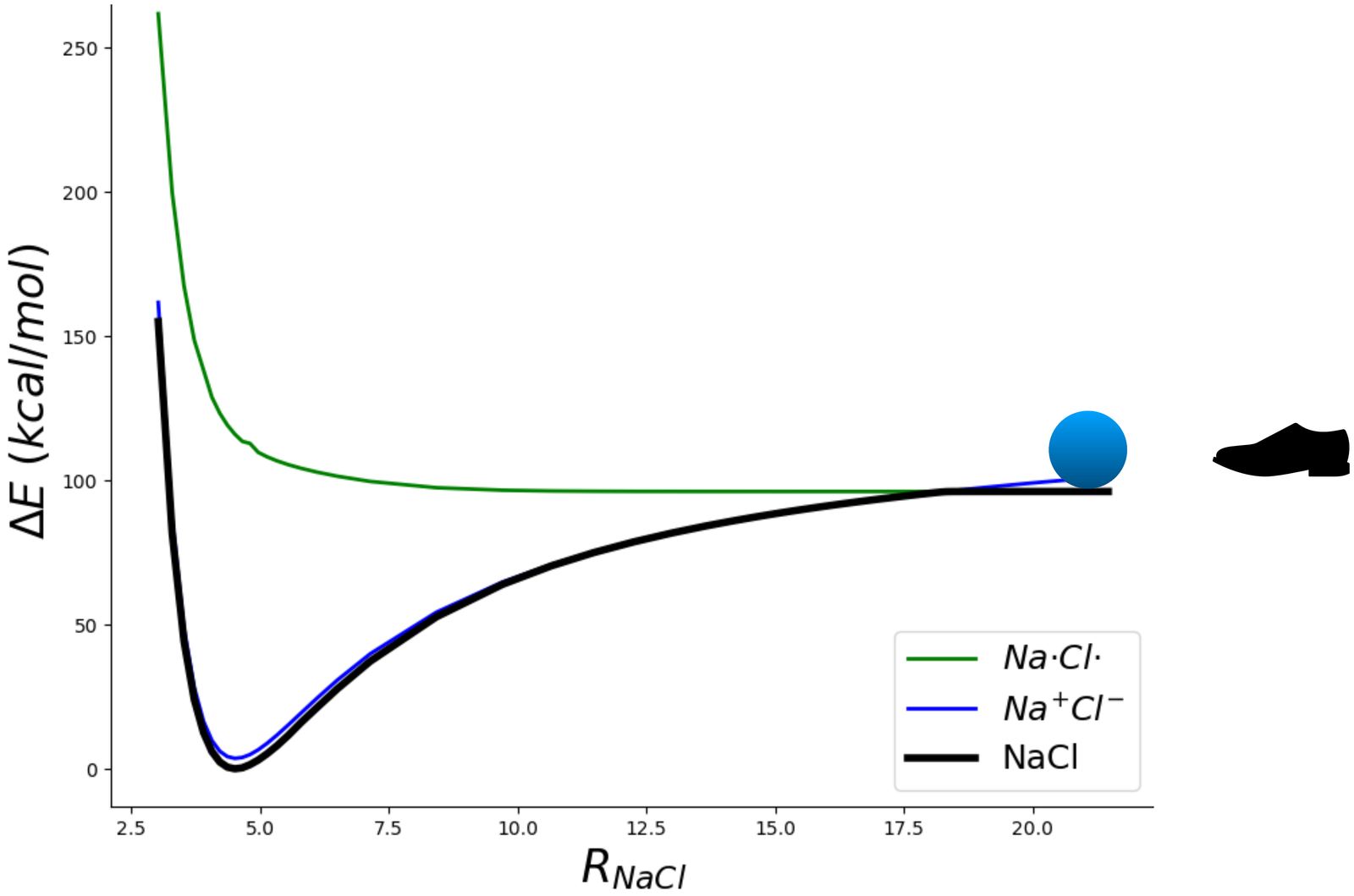


6K05

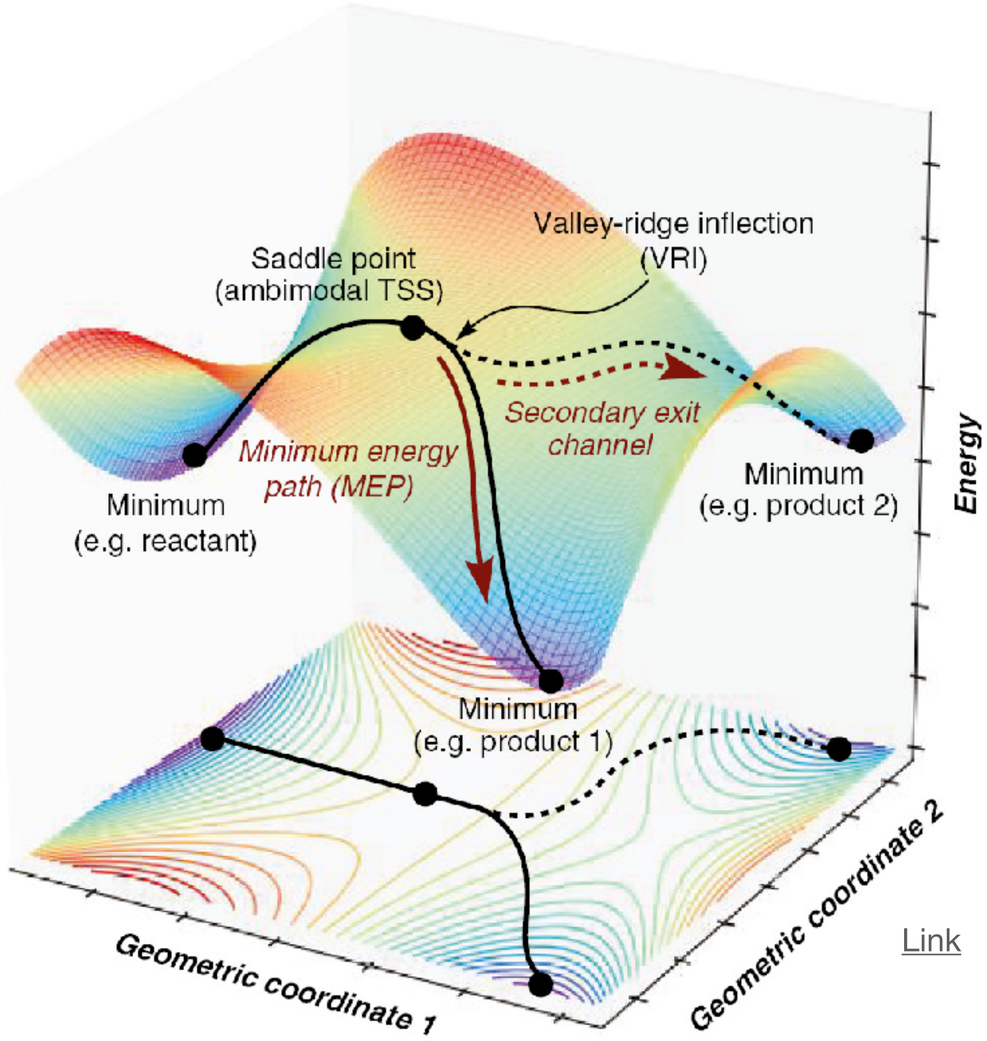
Optg and MD


250
200
150
100
50
0
ΔE (kcal/mol)
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
R_{NaCl}
$Na{\cdot}Cl{\cdot}$
$Na^{+}Cl^{-}$
NaCl

Optg and MD

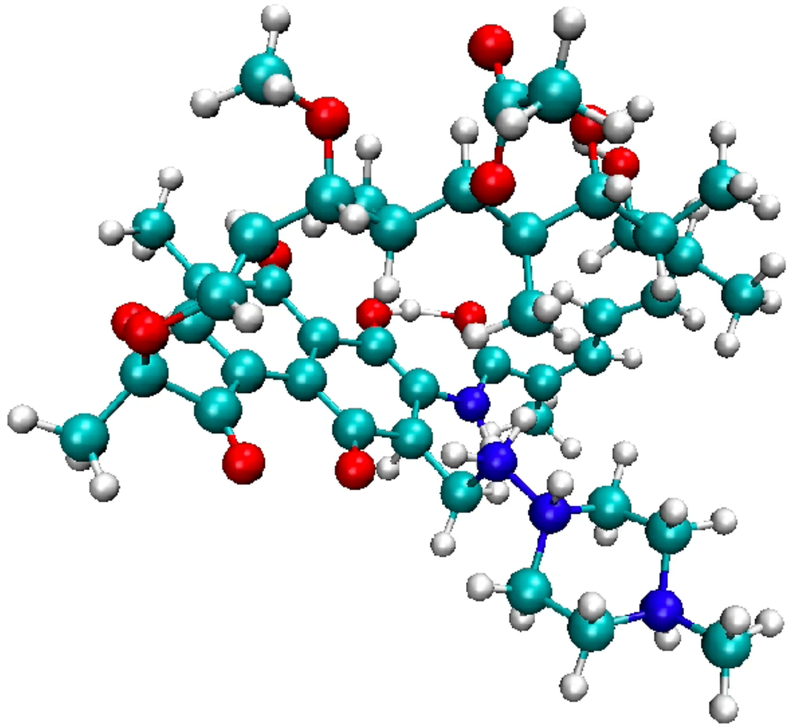

250
200
150
100
50
0
ΔE (kcal/mol)
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
R_NaCl
Na·Cl·
Na⁺Cl⁻
NaCl

Potential Energy Surface

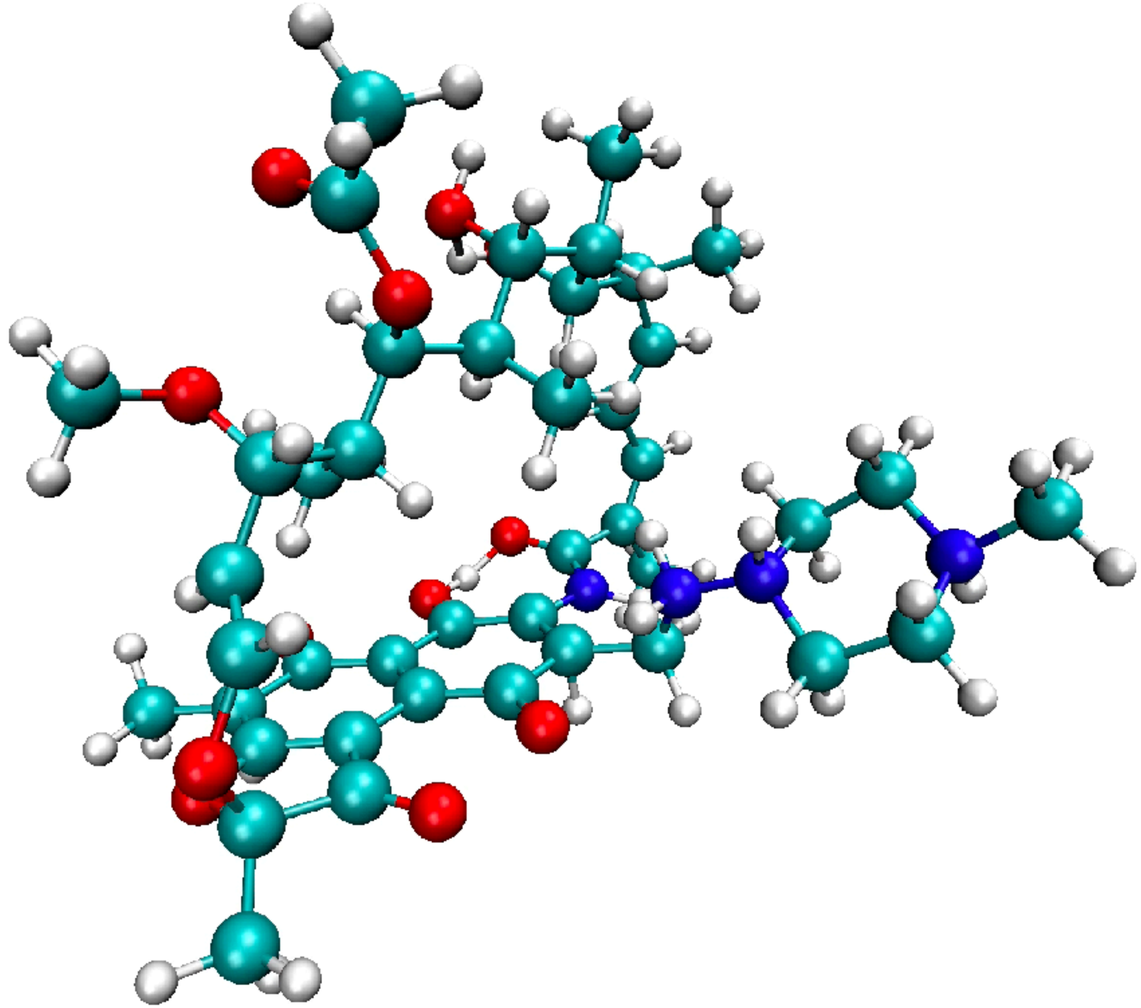


Link

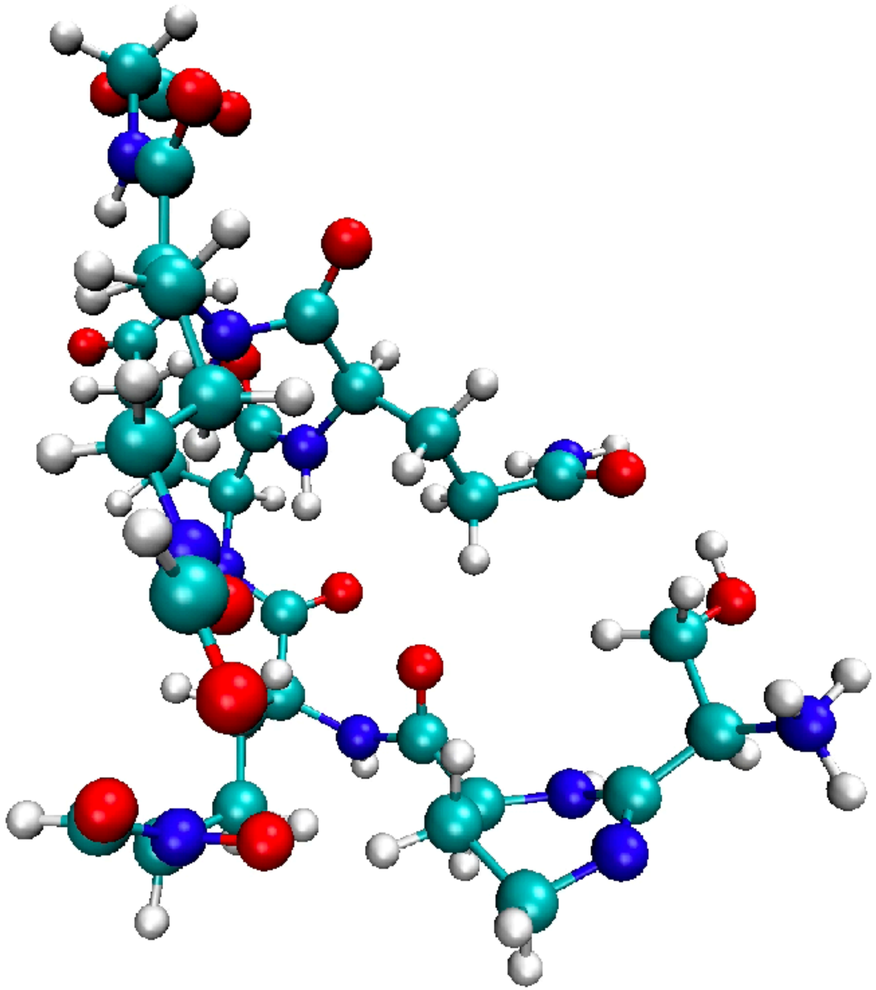
MD of Incorrect Structure



Optg of (same) Incorrect Structure



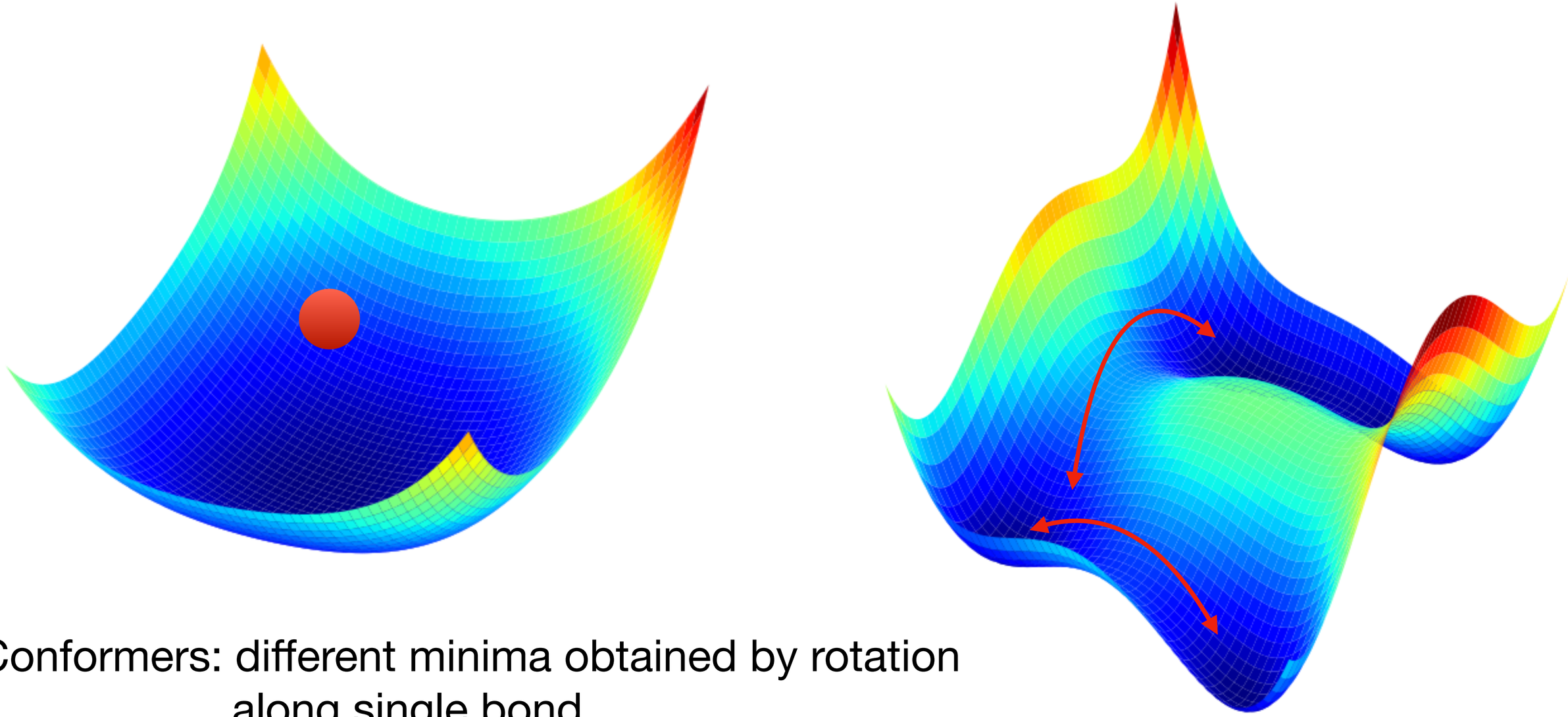
Radicals Force Their Way



- Two electrons were missing (radical).
- Those atoms panicked: incomplete shells
- They did all they could to refill shells

2W76

Molecules as Minima in Surfaces

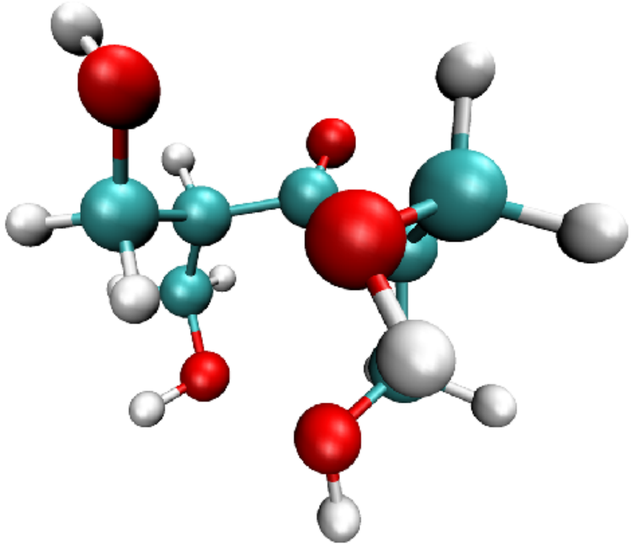


Conformers: different minima obtained by rotation along single bond

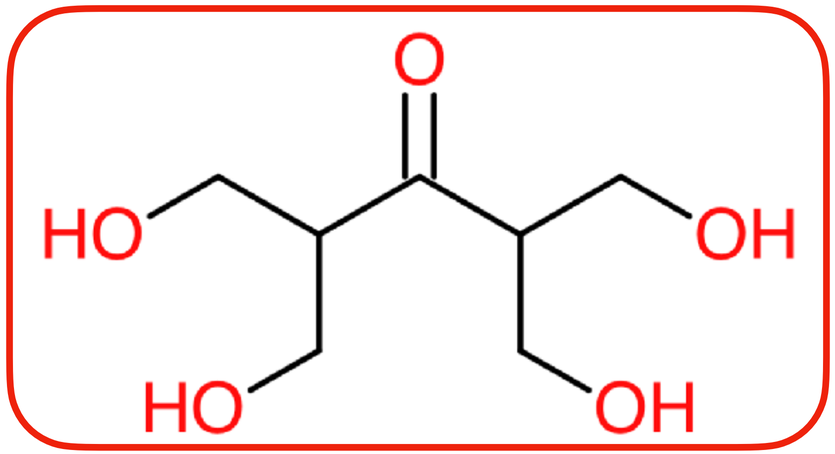
10.1002/9780470125793.ch1

All Conformers are equal… are they?

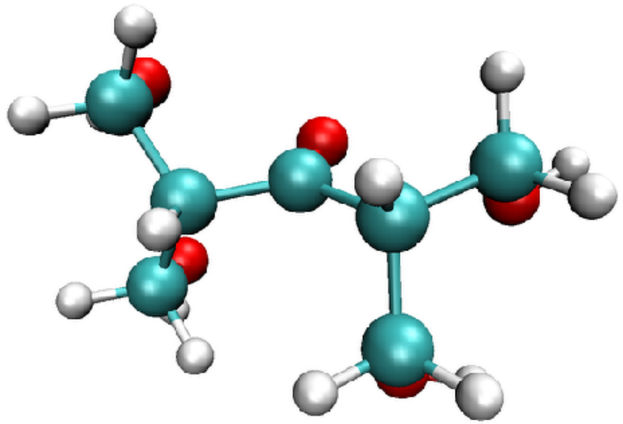
Conformer A



Solvent 1

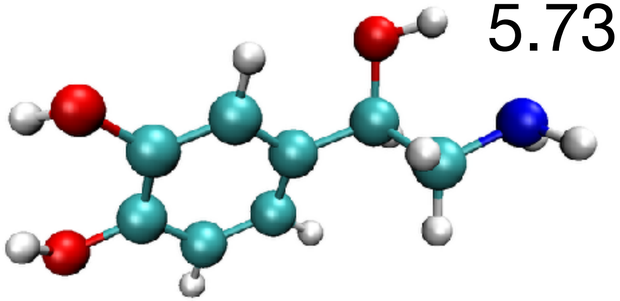


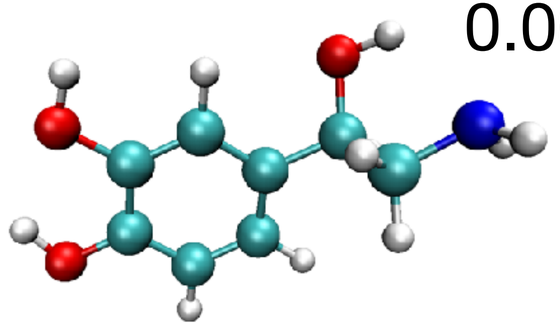
Conformer B

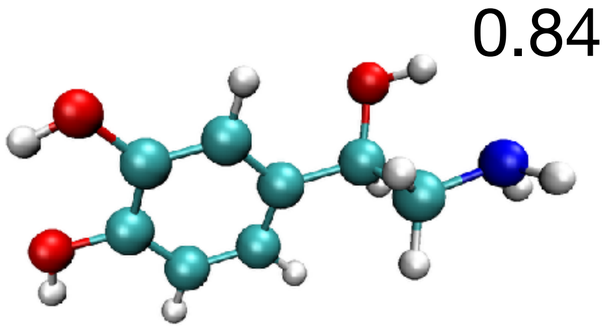


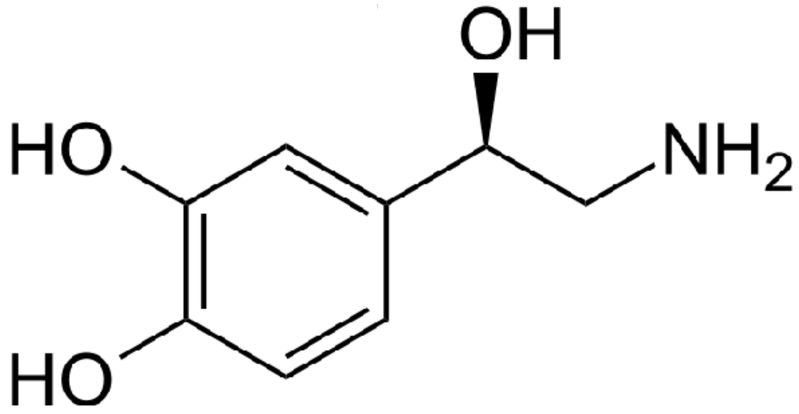
Solvent 2

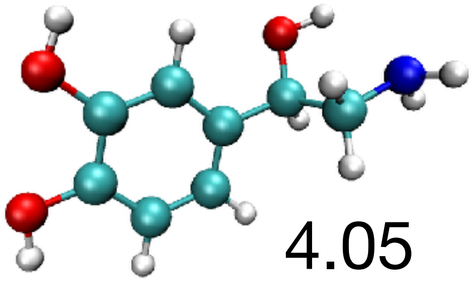
All Conformers are equal…

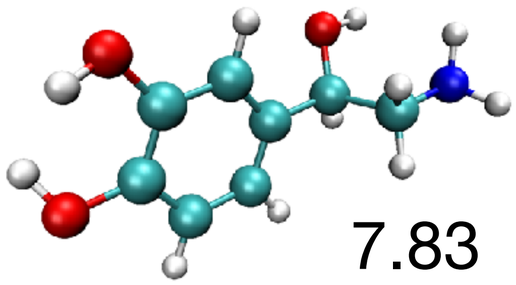


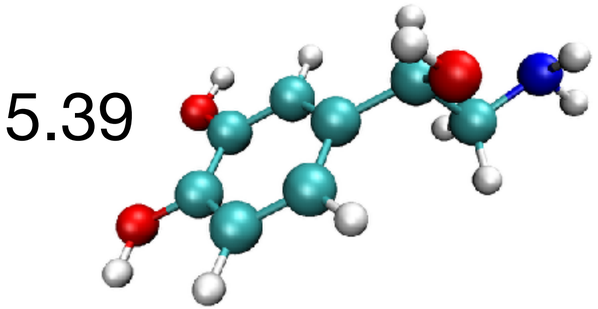


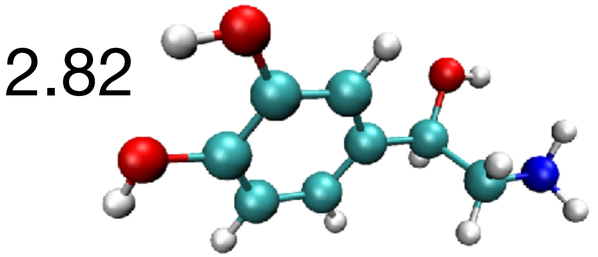


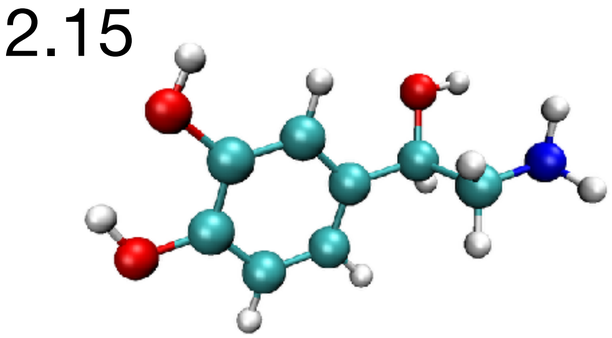






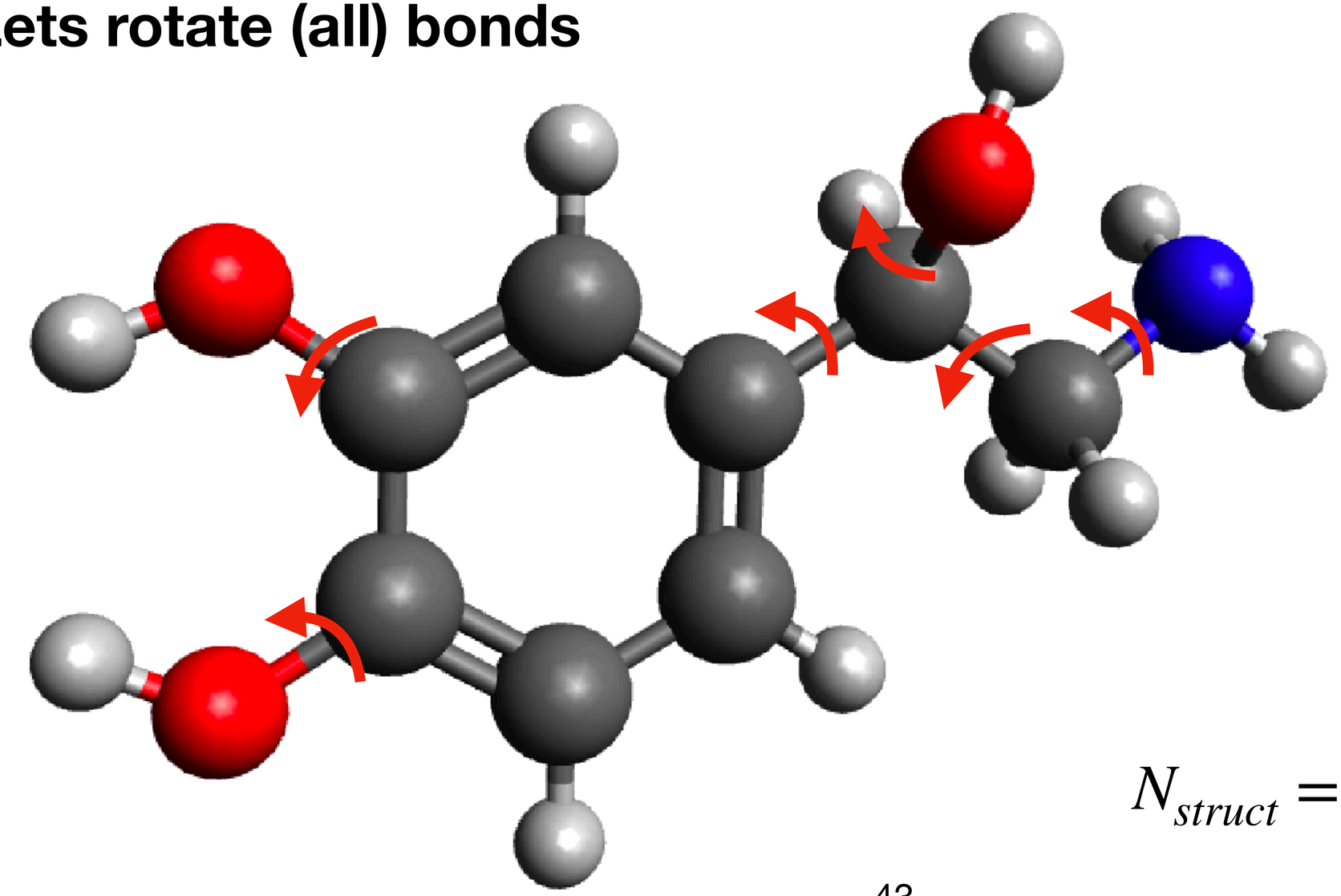






Simple Algorithms for Finding Conformers

Lets rotate (all) bonds

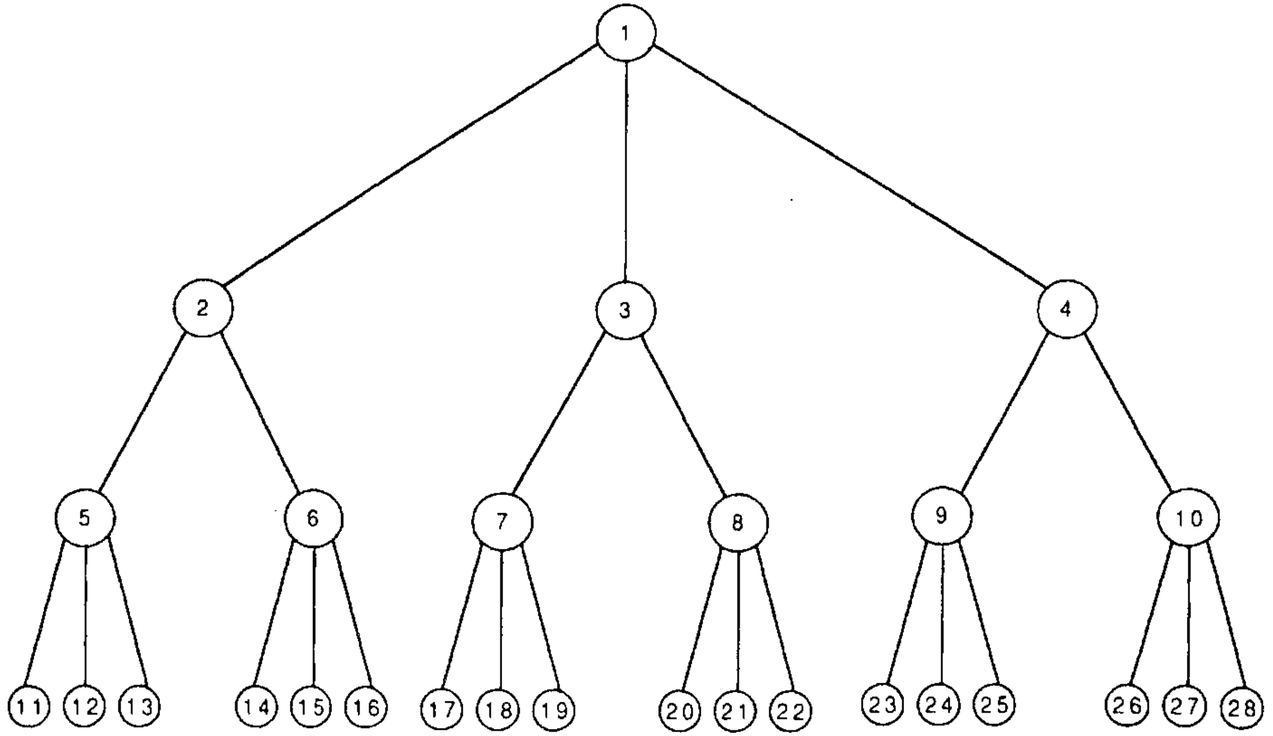


$$N_{struct} = \left(\frac{360}{\theta}\right)^{N}$$

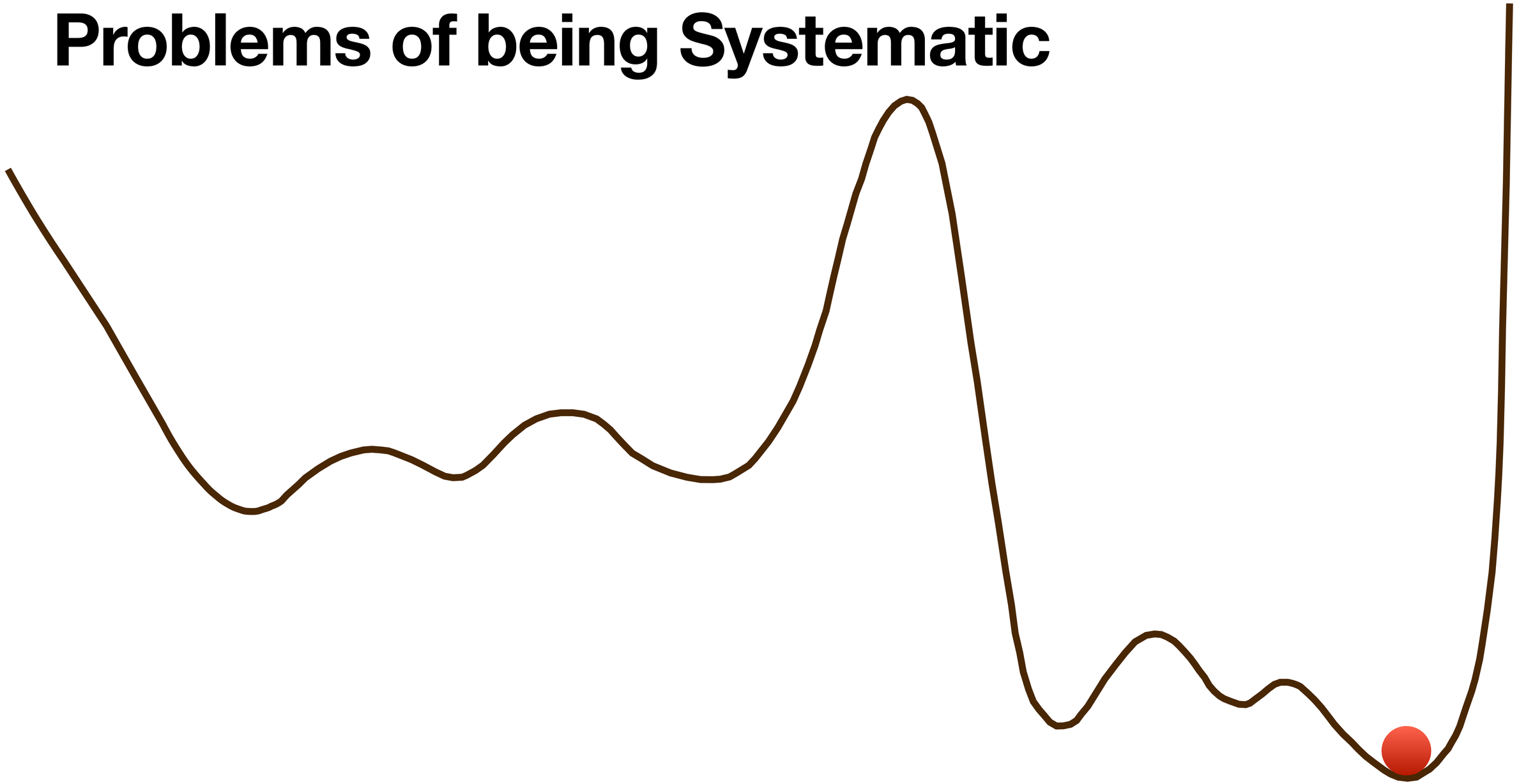
$$N_{struct} = \left(\frac{360}{45}\right)^{6} = 262144$$

Simple Algorithms for Finding Conformers

Lets use tree… searches

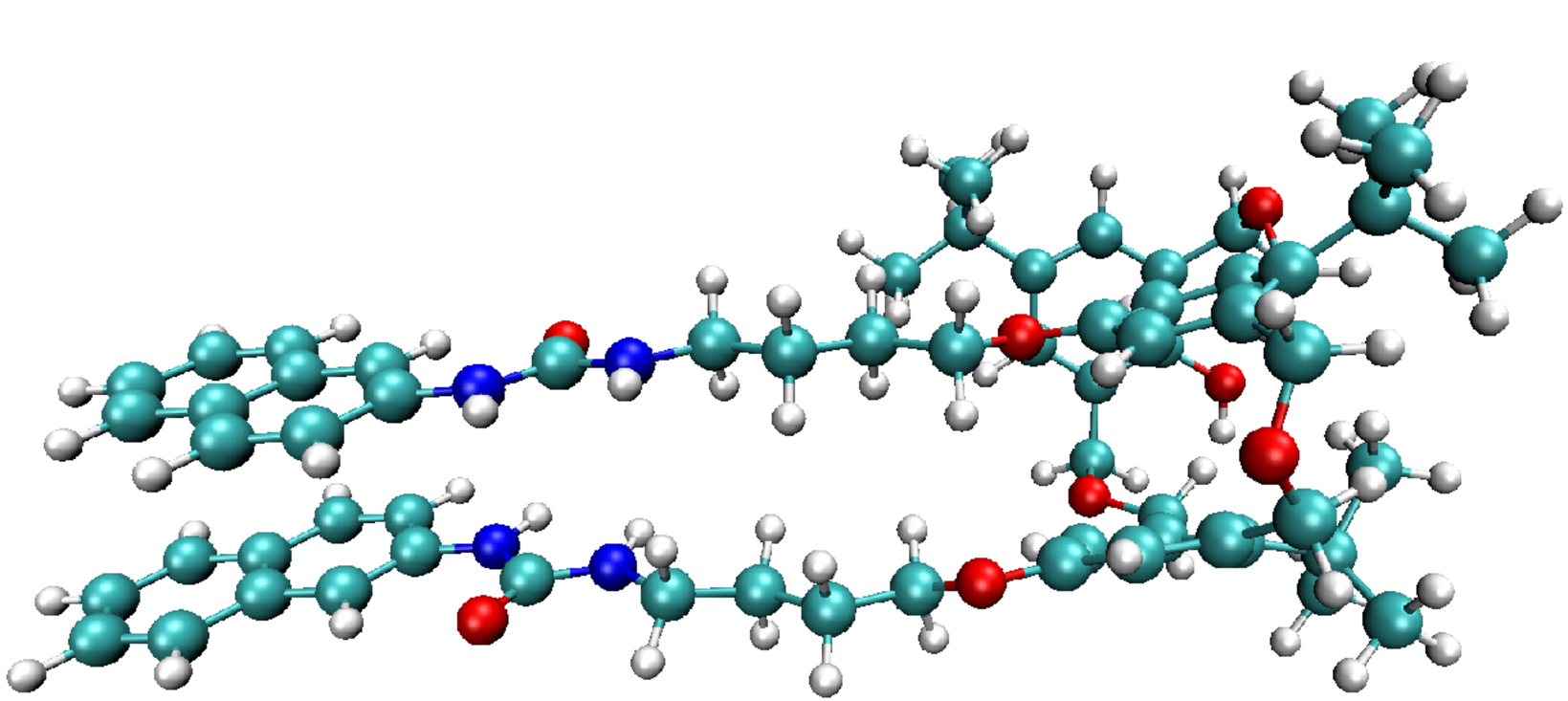


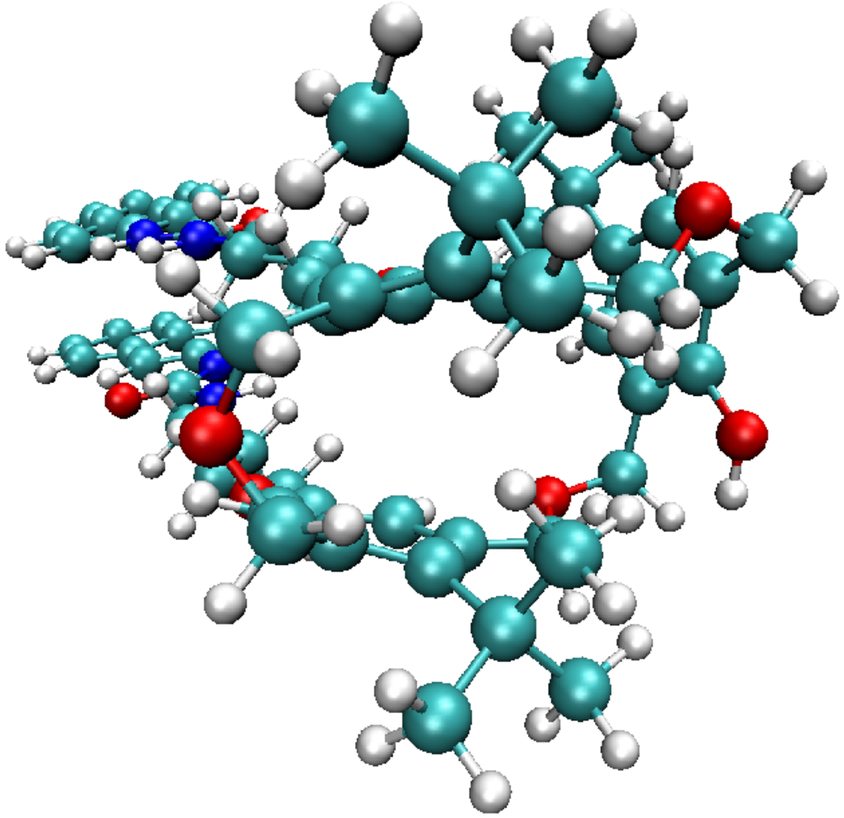
Problems of being Systematic



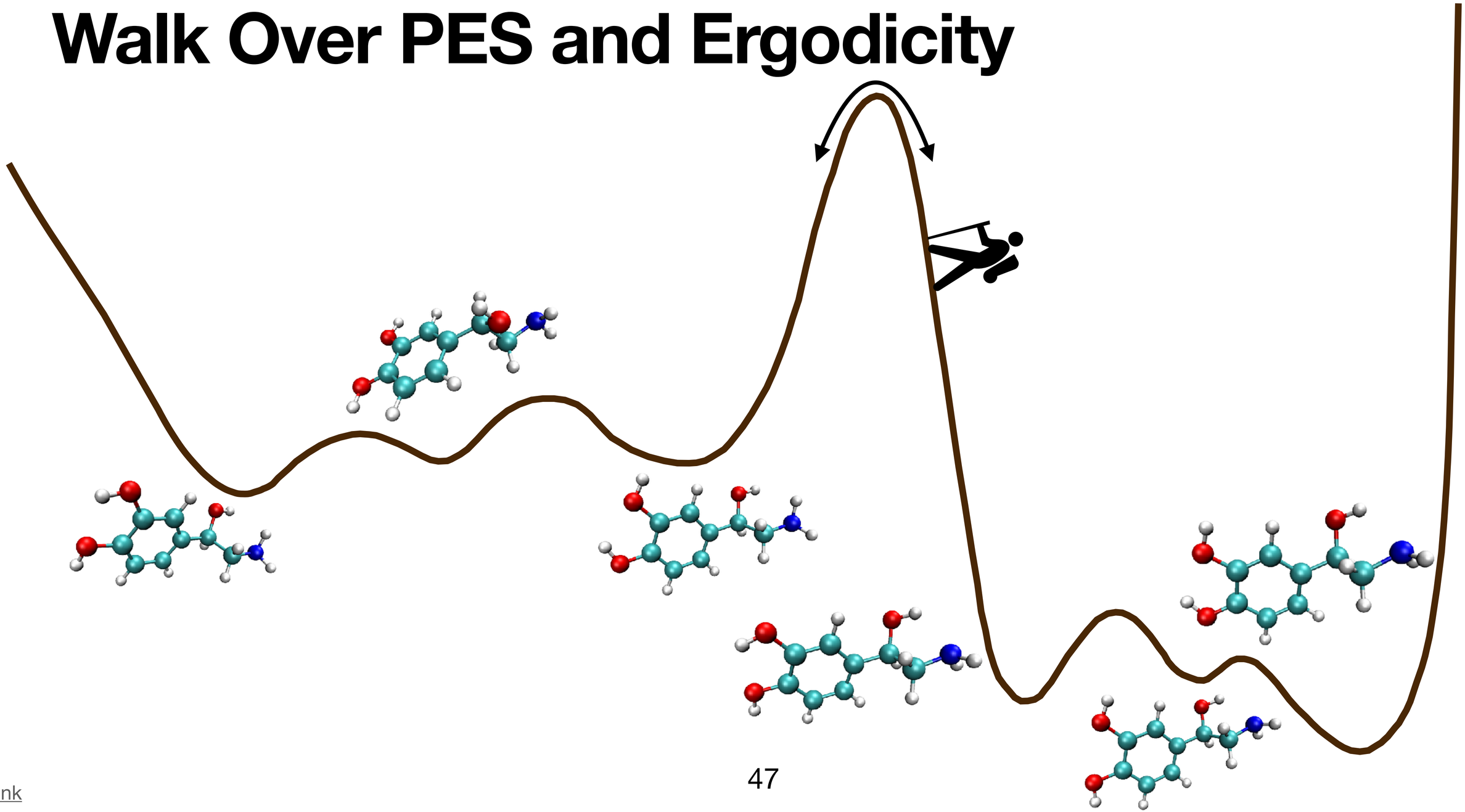
Link

Problems of being Systematic



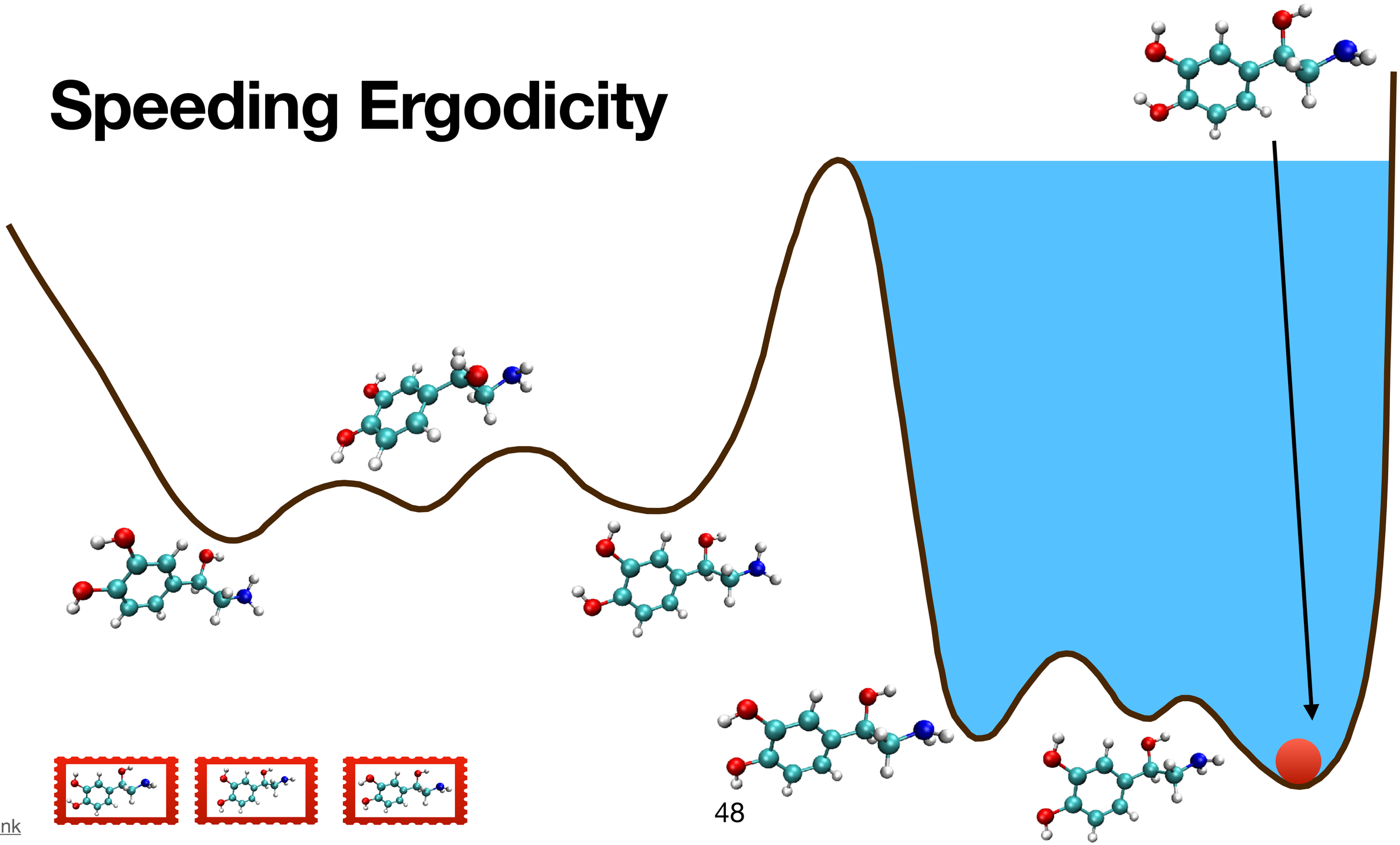


Walk Over PES and Ergodicity



Link

Speeding Ergodicity



Link

Metadynamics Conformer Search

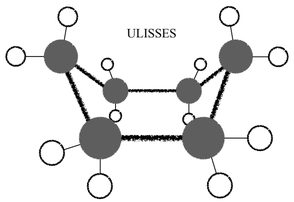
- Start with fast, but reasonable
  Only for small molecules, different methods yield similar minima.

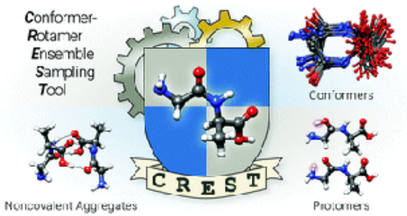
- After running low-level model chemistries, increase the level as much as possible for your resources.

$$FF < SEQM < DFT \leq \Psi$$

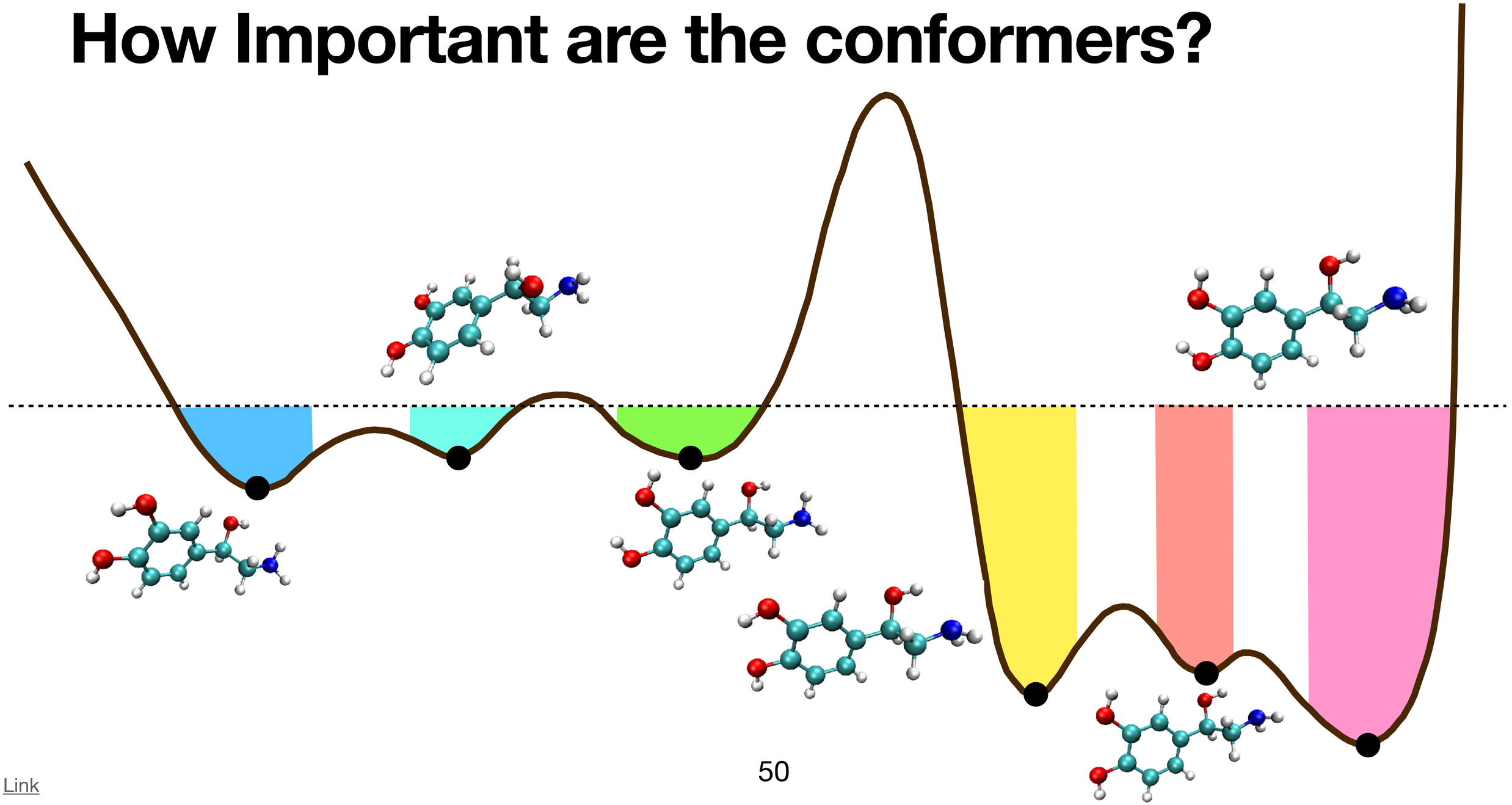
- In between runs, optimise all structures.

- Use previous ensemble as starting point for next run.





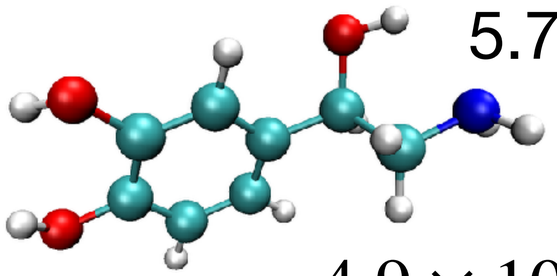
How Important are the conformers?



Link

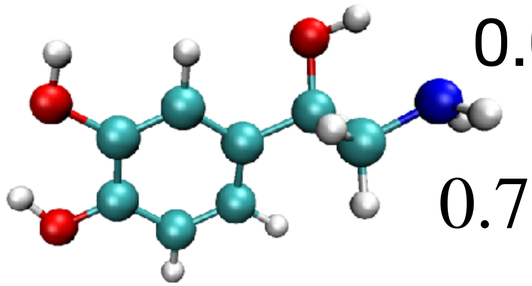
Which Conformers are relevant?

Ludwig Boltzmann



5.73

4.9×10^{-5}

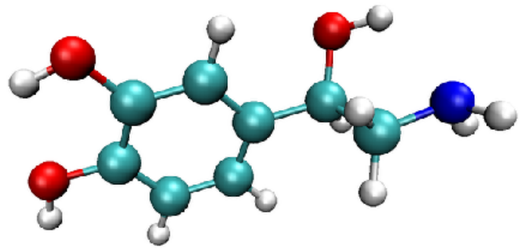


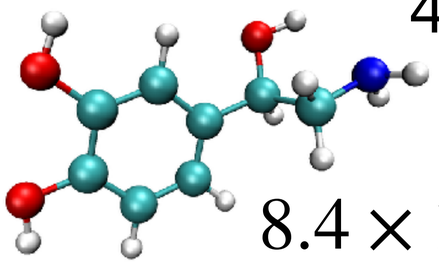
0.00

0.78

0.84

0.19

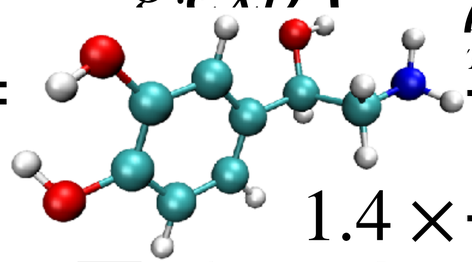




4.05

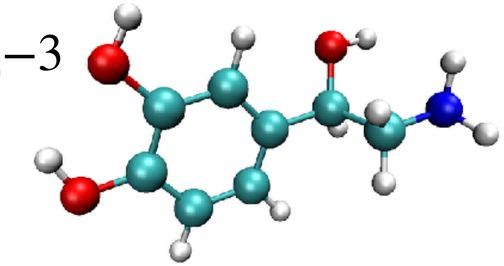
8.4×10^{-4}

$$P_i = \frac{g_i \exp\left(-\frac{\Delta E_i}{k_B T}\right)}{\sum_j g_j \exp\left(-\frac{\Delta E_j}{k_B T}\right)}$$



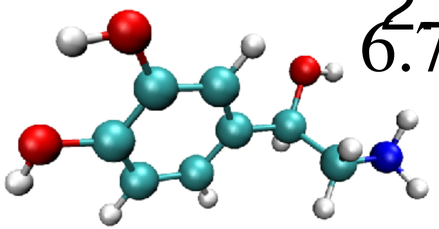
2.15

8.8×10^{-3}



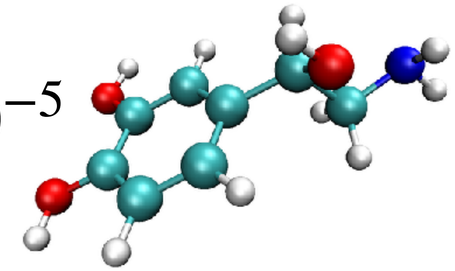
2.82

6.7×10^{-3}



5.39

8.8×10^{-5}



This is a semi-classical approach, valid when quantum effects are not relevant for statistics.

Which Conformers are relevant?

Ludwig Boltzmann

$$P_i = \frac{g_i exp\left(-\frac{\Delta E_i}{k_B T}\right)}{\sum g_j exp\left(-\frac{\Delta E_j}{k_B T}\right)}$$

VALIDITY:

- Gas phase
- All structures similar (Similar vibrations)

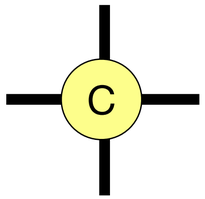
$$P_i = \frac{g_i exp\left(-\frac{\Delta G_i}{k_B T}\right)}{\sum g_j exp\left(-\frac{\Delta G_j}{k_B T}\right)}$$

VALIDITY:

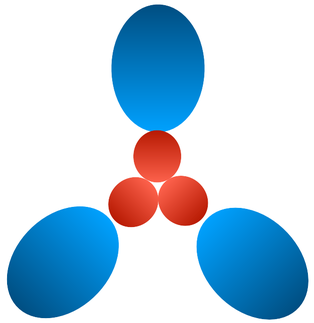
- Any phase
- Entropy is considered

This is a semi-classical approach, valid when quantum effects are not relevant for statistics.

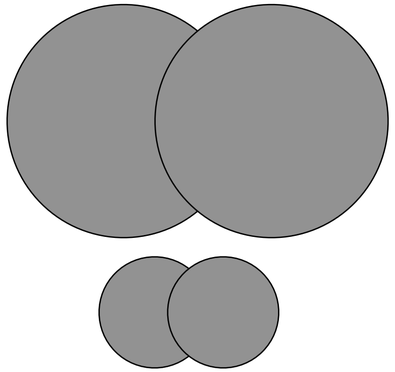
Summary



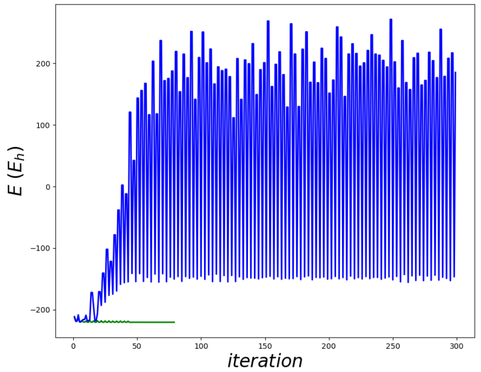
Octet Rule



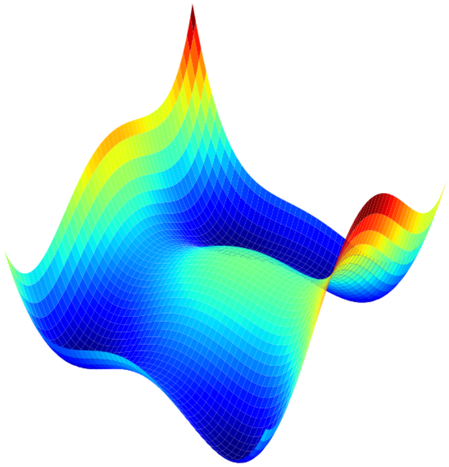
Hybridization



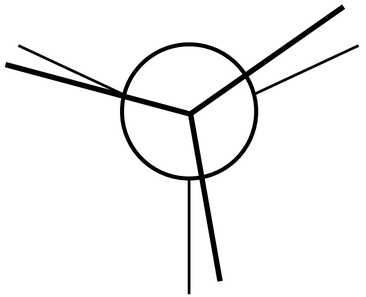
Hard-Soft



Convergence Analysis



Optg, MD



Conformers

Acknowledgments

- Grzegorz Popowicz
- Michael Sattler
- Igor Tetko
- Pavel Karpov
- BMWi ZIM. KK 5197901TS0
- BMBF, SUPREME, 031L0268

Advanced machine learning for Innovative Drug Discovery

HELMHOLTZ MUNICH

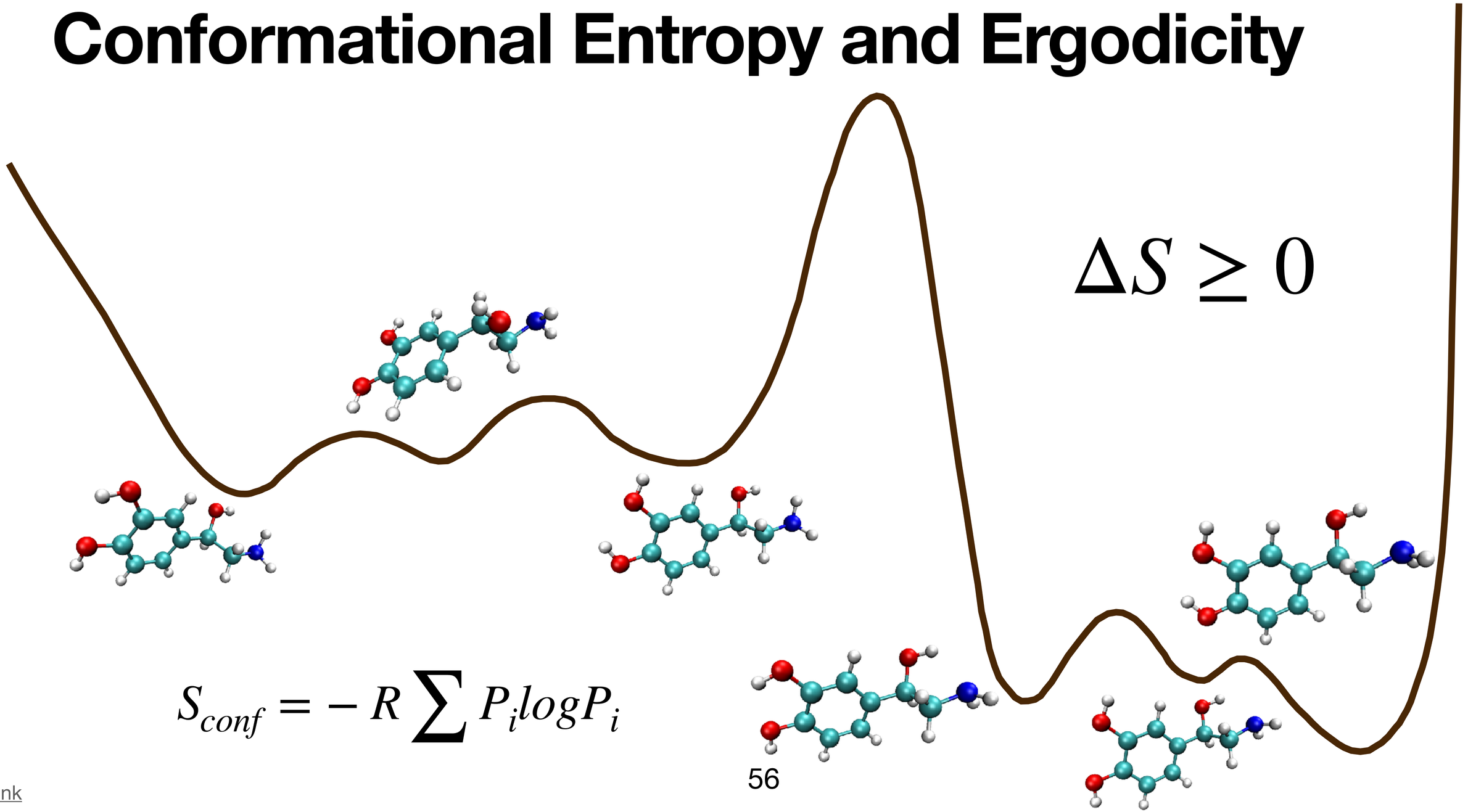
Institute of Structural Biology

Bayerisches NMR Zentrum

- Khumbu AI
- Sattler Group
- All Participants

The End

Conformational Entropy and Ergodicity



$$\Delta S \geq 0$$

$$S_{conf} = -R \sum P_i log P_i$$

Link

Conformational Entropy and Ergodicity