

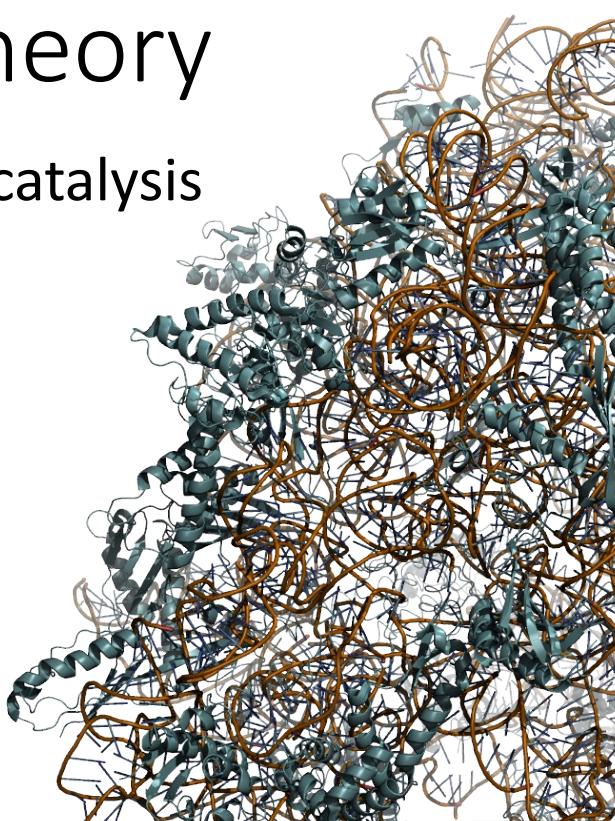


Density Functional Theory

Molecular modeling of enzymatic catalysis

Biocat C8

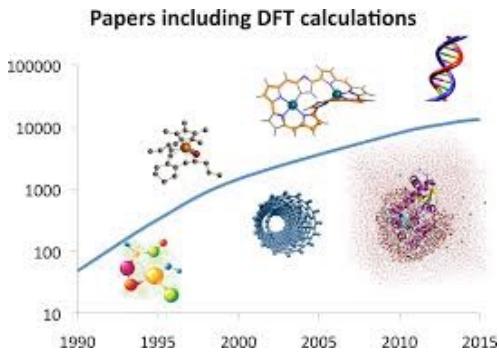
Geir Villy Isaksen



Geir Villy Isaksen

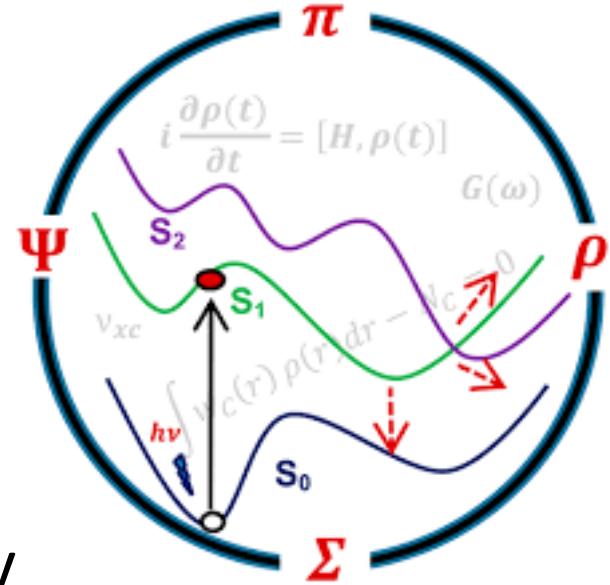
Why DFT?

- Reaction mechanisms
- Force field parameters
- Energies when experimental data is lacking
- ... Very easy to do!
 - The theory is a bit more complicated...

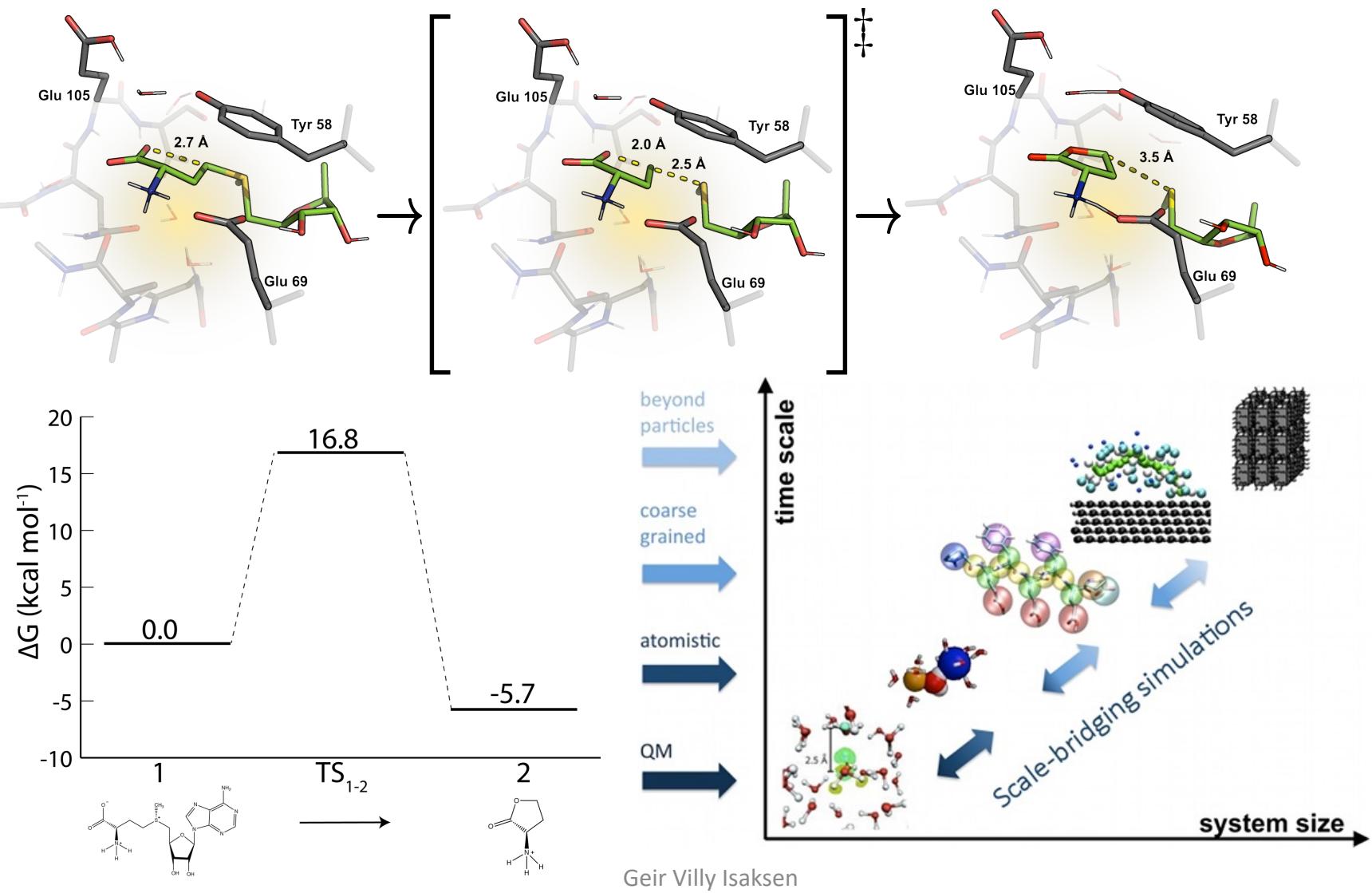


Outline

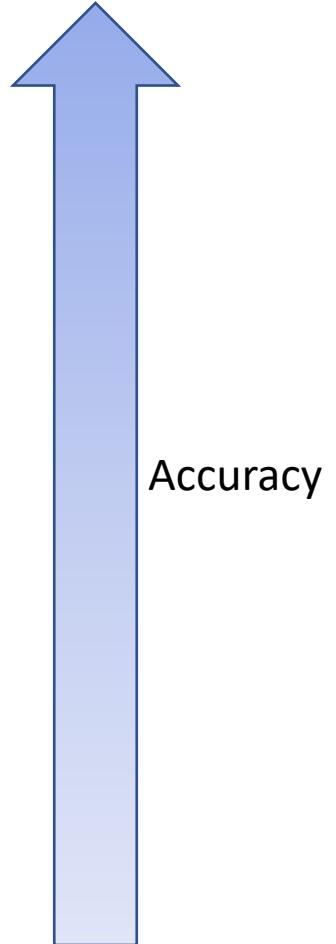
- Easy background
- Getting to know the terminology
- How to use DFT
 - Reaction energies



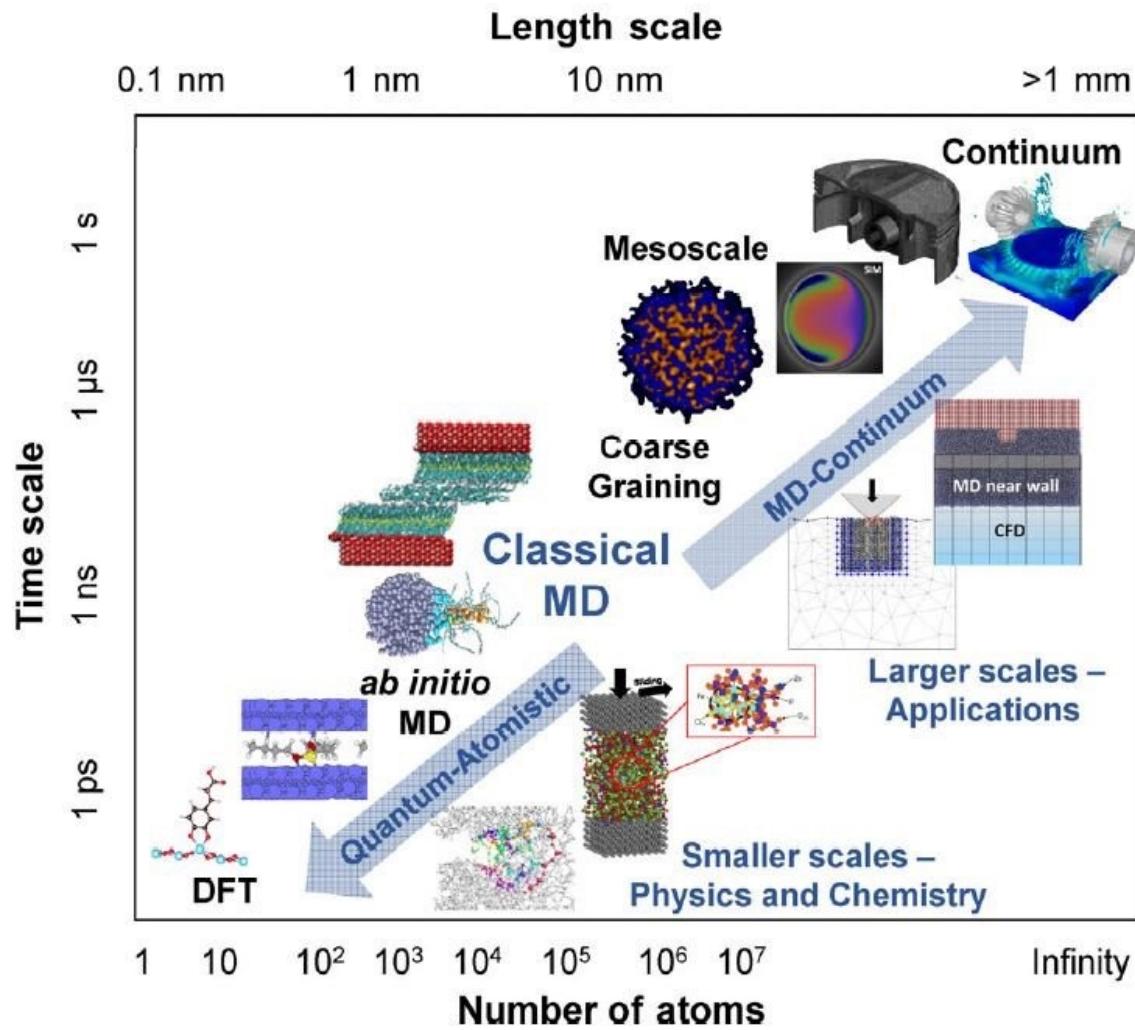
The main aim:



Computational models

- *Ab initio*
 - Uses Schrödinger's equation, but with approximations
 - Semi-empirical
 - Uses experimental parameters and extensive simplifications of Schrödinger's equation
 - Molecular mechanics
 - Does not use Schrödinger's equation
- 

Computational models



Ab Initio levels

- Hartree-Fock (HF)
 - The simplest calculation
 - Does not take electron correlation into consideration
- Møller-Plesset perturbation theory (MP)
- DFT
- Configuration Interaction (CI)

DFT in (very) short:

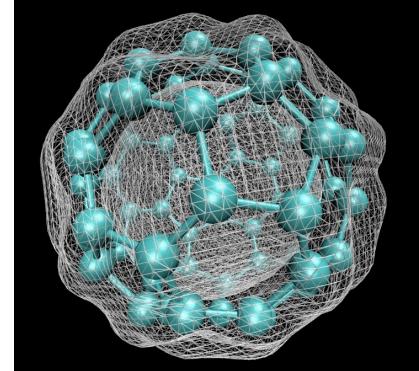
- Solving the Schrödinger equation is many-body problem ($3N$ to be exact)
 - The Schrödinger equation can't be solved for systems with more than two particles interacting ...
- DFT: Electrons are indistinguishable.
 - The quantity of interest is instead the probability that N electrons in any order have specific coordinates. Thus, the electron density, which is a function of only 3 coordinates (x,y,z), contains information that is physically observable from the full solution of the Schrodinger equation.

Density Functional Theory

- Uses electron density (ρ) instead of wave functions
- 2 fundamental mathematical theorems proven by Kohn and Hohenberg, and derivation of a set of equations by Kohn and Sham:
 - #1: There exists a 1-to-1 correspondance between the ground state wave function and the ground state electron density.
 - #2: The electron density that minimizes the energy of the above functional is the true electron density corresponding to the full solution to the Schrödinger equation.

$$\hat{H}\Psi = E\Psi$$

Density Functional Methods

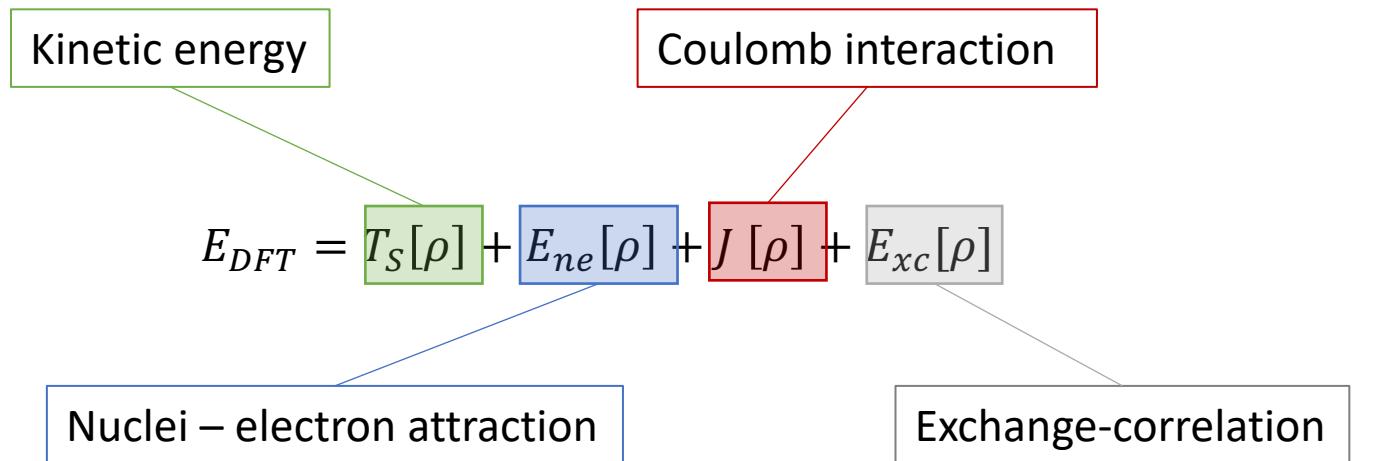


- Hohenberg and Kohn:
 - “*The ground state electronic energy is determined completely by the electronic density ρ* ”
- Kohn and Sham:
 - “*The electron kinetic energy should be calculated from an auxiliary set of orbitals used for representing the electron density*”
- n one-electron Schrödinger-like equations (Born-Oppenheimer)

$$E_{DFT} = T_S[\rho] + T_{ne}[\rho] + J[\rho] + E_{xc}$$

$$\rho(\vec{r}) = \sum_{i=1}^N |\phi_i(\vec{r})|^2$$

Kohn-Sham theory



$$T_S[\rho] = \sum_{i=1}^N \langle \phi_i \left| -\frac{1}{2} \nabla^2 \right| \phi_i \rangle$$

$$E_{ne}[\rho] = \sum_A \int \frac{Z_A \rho(\vec{r})}{|\vec{R}_A - \vec{r}|} d\vec{r}$$

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}'$$

- Can compute every energy term exactly, except for the **exchange-correlation**:
- $$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho])$$
- Hence, we use various approximate exchange-correlation functionals (B3LYP, MO62X...)

Exchange-correlation

$$E_{DFT} = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho]$$

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho])$$

Kinetic correlation

Potential correlation and exchange energy

- Local density approach (LDA)
 - S-VWN
- Gradient-corrected approximation (GGA)
 - PW91, LYP, B88
- Meta-GGA
 - M06-L
- Hybrid DFT (HF exchange)
 - B3LYP (hybrid GGA), M06-2X (hybrid meta-GGA)

Exchange-correlation

$$E_{DFT} = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho]$$

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho])$$

Kinetic correlation

Potential correlation and exchange energy

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho]$$

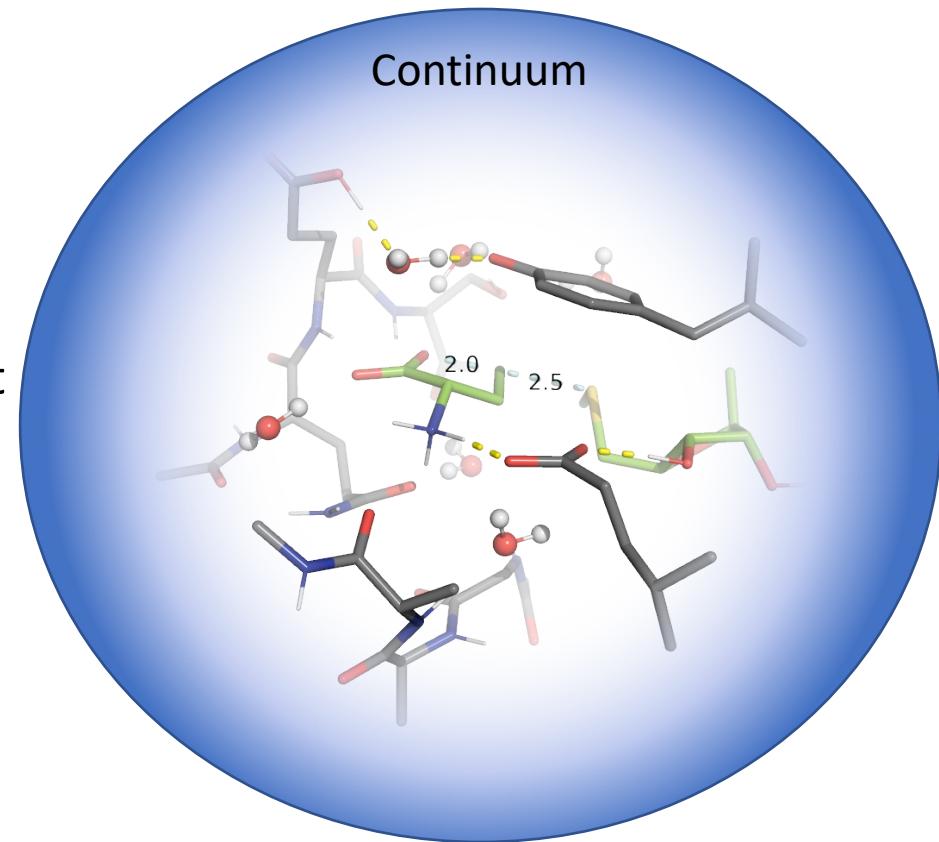
- Exchange functional: S, B, B3, etc.
- Correlation functional: LYP, PW91, etc

Nice to know...

- DFT improves upon Hartree-Fock by including an approximate treatment of the correlated motions of electrons (these are treated in Hartree-Fock in only an averaged sense)
- Treatment of electron correlation is much cheaper than in correlated wavefunction methods like MP2, CCSD, CCSD(T)
- Although there is a loose ordering of density functionals from “less sophisticated” to “more sophisticated” (e.g., LDA to GGA to hybrid, etc.), in practice there is no reliable way to improve your computation by going to the “next better” functional. By contrast, this is possible with wavefunction methods: one almost always has CCSD(T)>CCSD>MP2>HF.

DFT & Enzymatic catalysis

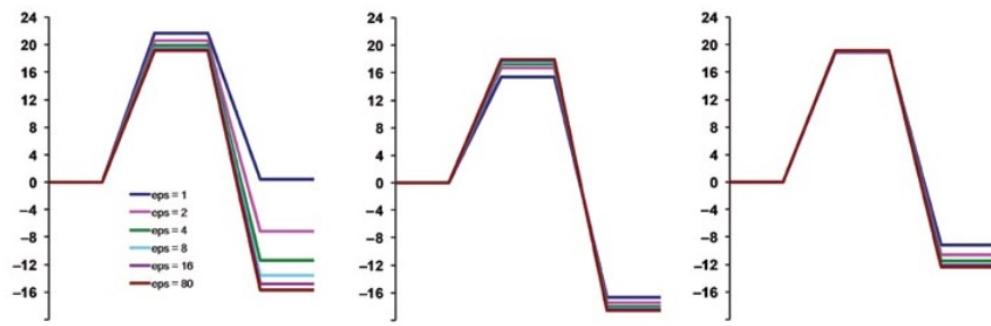
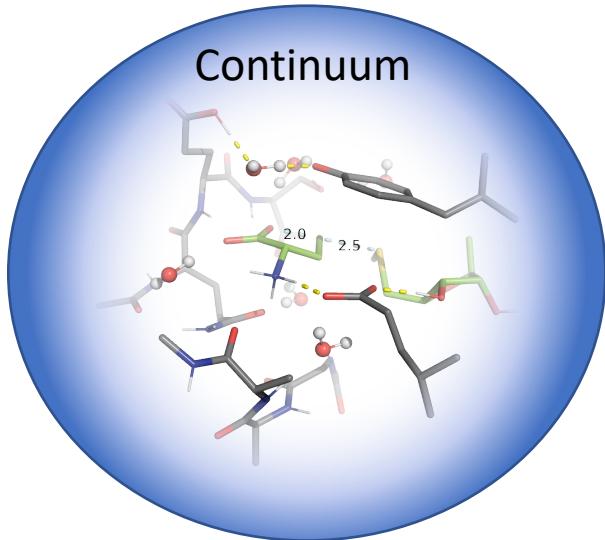
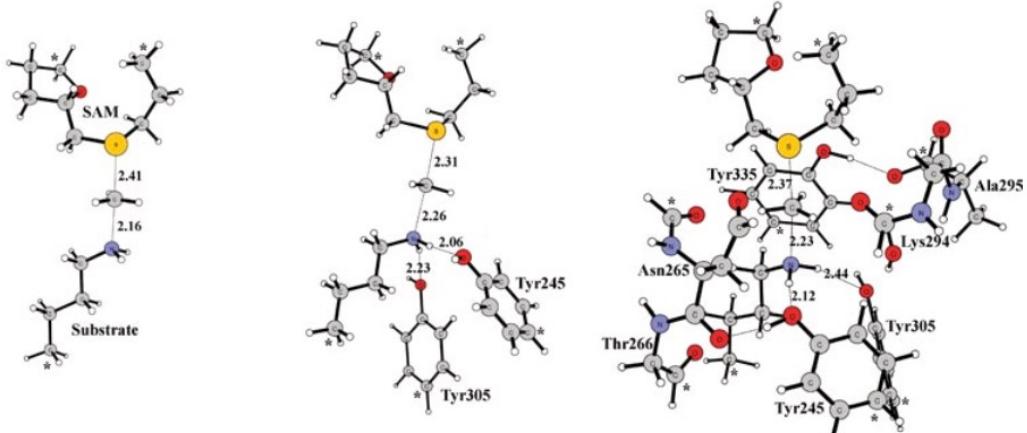
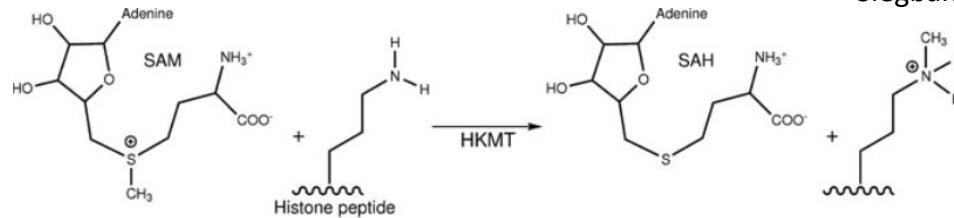
- Cluster models
 - Active site of enzyme
 - >200 atoms no problem
 - Solvation?
 - Depends on system size
 - Important waters – explicit
 - Continuum
 - Dielectric constant?
 - Solvation model?
 - Functional?
 - Basis set?



DFT & Enzymatic catalysis

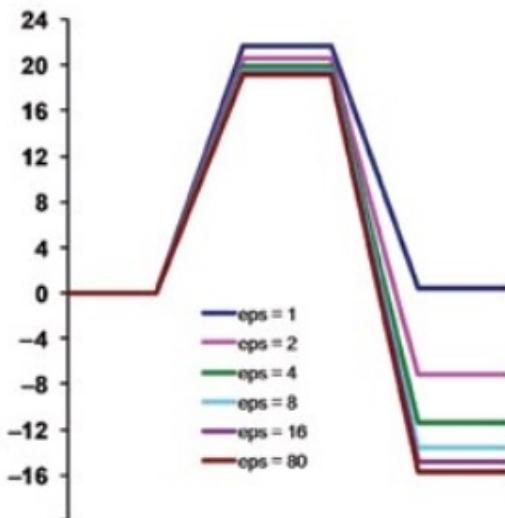
Siegbahn, 2011 #66

- Solvation effects
 - 150-200 atoms
 - $\epsilon = 4$

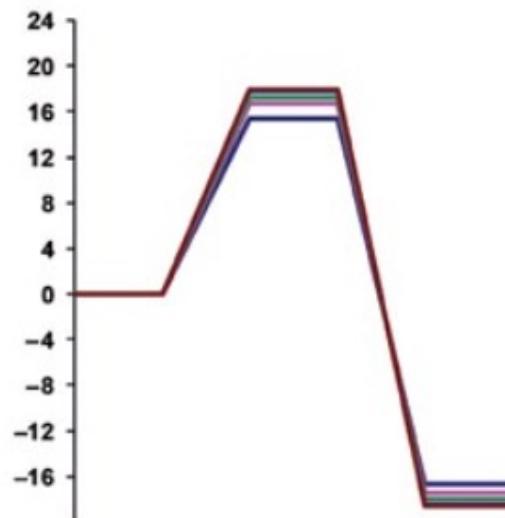


DFT & Enzymatic catalysis

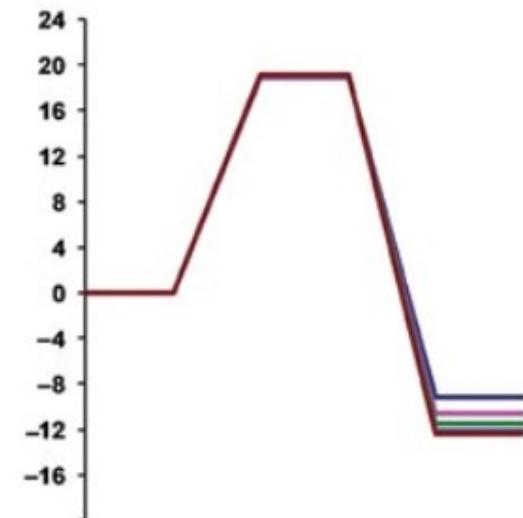
- Solvation effects
 - 150-200 atoms
 - $\epsilon = 4$



Model I



Model II



Model III

DFT & Enzymatic catalysis

- Functionals
 - Hybrid DFT methods most popular
- B3LYP
 - $F^{B3LYP} = (1 - A) \cdot F_x^{Slater} + A \cdot F_x^{HF} + B \cdot F_x^{Becke} + C \cdot F_c^{LYP} + (1 - C) \cdot F_c^{VWN}$
- The most sensitive parameter is the exact exchange in the functional
 - 20% in B3LYP
- Lacking attractive long-range dispersion
 - Grimme dispersion correction (GD3)

DFT & Enzymatic catalysis

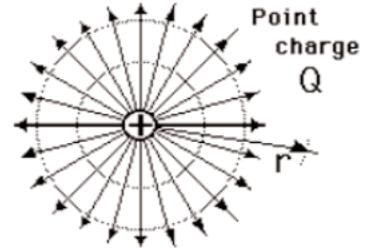
- Whatever functional you chose...

Stick with it!

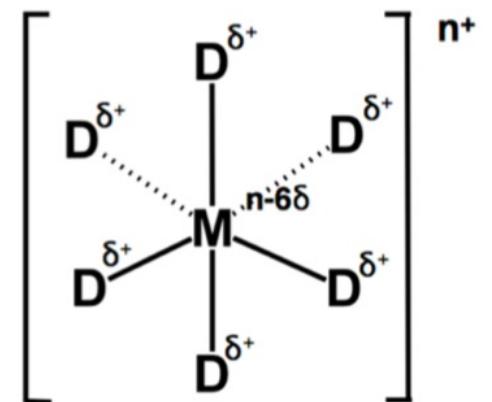
How can we take advantage of DFT?

- Improve force field parameters
- Find correct reaction mechanism
- Energies for reactions (when experimental data is missing)

Point charges (FF)

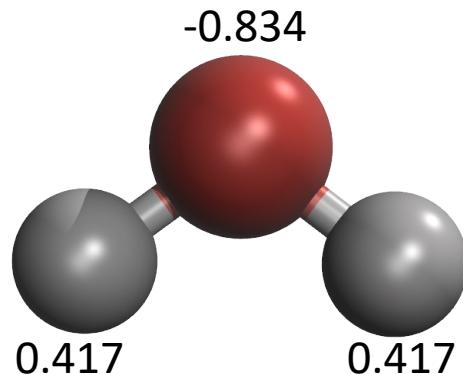


- “*An idealized model of a particle which has and electrical charge*”
- “*An Electrical charge at a mathematical point with no dimensions*”
- Are point charges good enough?

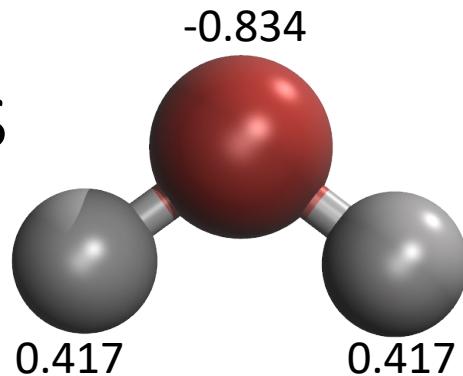


Force Fields

- $E_{FF} = E_{bonded} + E_{non-bonded}$
- $E_{bonded} = E_{bond} + E_{angle} + E_{torsion} + E_{improper}$
- $E_{non-bonded} = E_{vdW} + E_{El}$



Force Fields & atomic Charges



- $E_{FF} = E_{bonded} + E_{non-bonded}$
 - $E_{bonded} = E_{bond} + E_{angle} + E_{torsion} + E_{improper}$
 - $E_{non-bonded} = E_{vdW} + E_{EL}$
- Partial charges (net atomic charge) matters!
 - H-bonds for example:
 - Electron-deficient hydrogen
 - Electronegative atoms such as oxygen or Nitrogen
- Long-range electrostatics dominating
 - Not in non-polar systems ($E_{vdW} + E_{tors}$)

The electrostatic energy

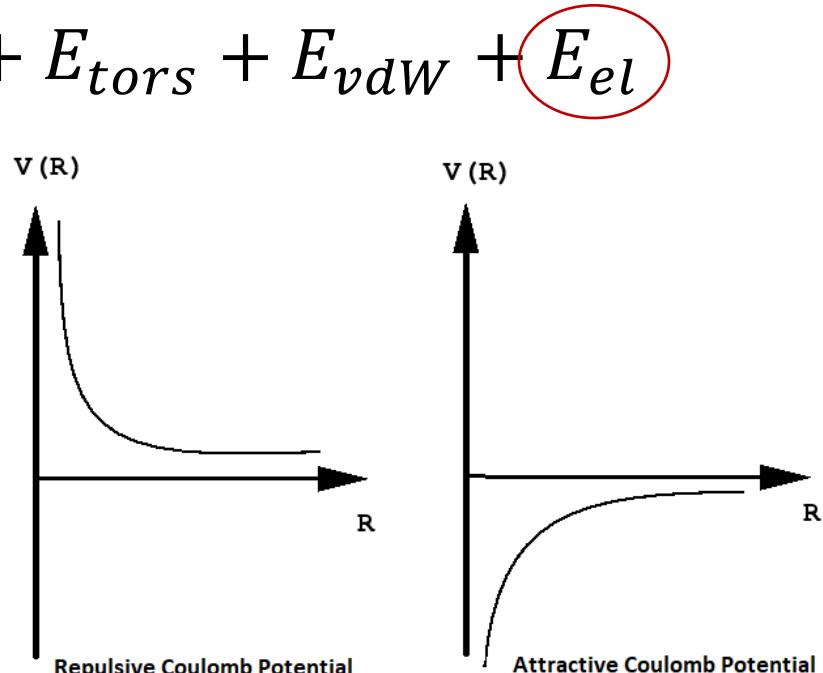
- $E_{FF} = E_{bond} + E_{angle} + E_{tors} + E_{vdW} + E_{el}$

- Coulomb potential:

- $$E_{el}(r_{i,j}) = \frac{q_i q_j}{\epsilon_0 r_{i,j}}$$

- Can be assigned by

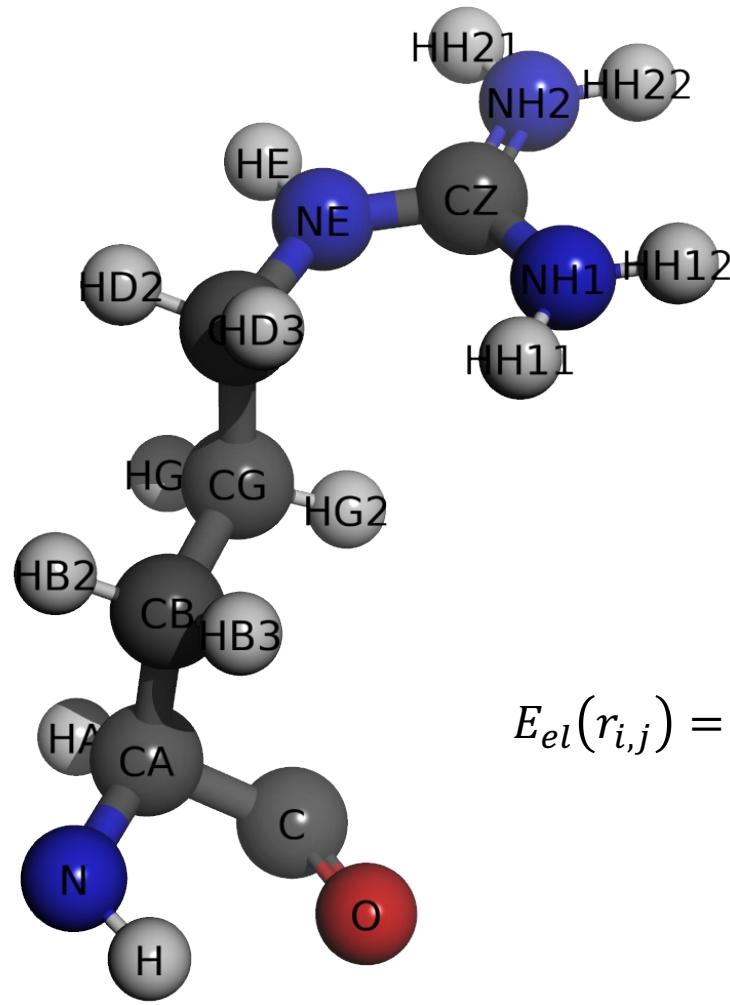
- Empirical rules
- Fitting of calculated electrostatic potential



Example: Arg OPLS charges

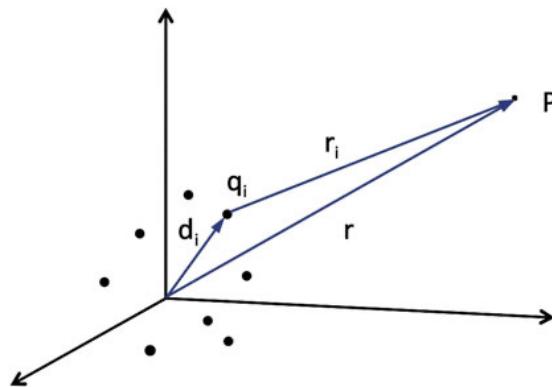
atomtype

		q_i
1	N	-0.5000
2	H	0.3000
3	CA	0.1400
4	HA	0.0600
5	CB	-0.1200
6	HB2	0.0600
7	HB3	0.0600
8	CG	-0.0500
9	HG2	0.0600
10	HG3	0.0600
11	CD	0.1900
12	HD2	0.0600
13	HD3	0.0600
14	NE	-0.7000
15	HE	0.4400
16	CZ	0.6400
17	NH1	-0.8000
18	HH11	0.4600
19	HH12	0.4600
20	NH2	-0.8000
21	HH21	0.4600
22	HH22	0.4600
23	C	0.5000
24	O	-0.5000



$$E_{el}(r_{i,j}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{i,j}}$$

Potentials from point charges



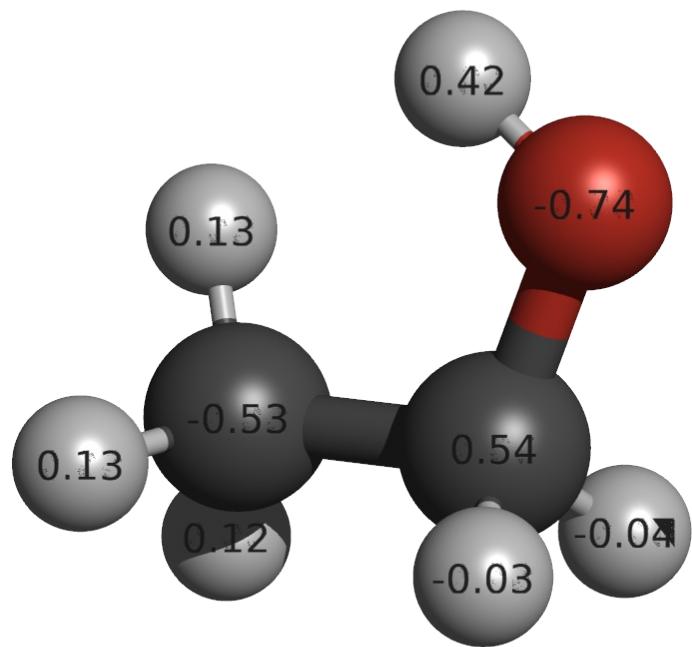
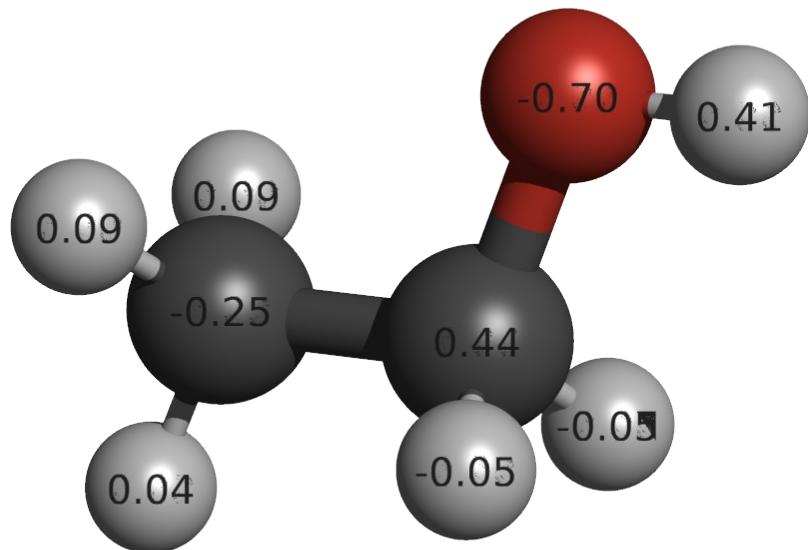
The potential $U_0(P)$ at the point $P(x,y,z)$ due to a distribution of N point charges q_i is equal to the sum of the potentials at P from each charge of the system.

The principle of superposition of the electric potentials:

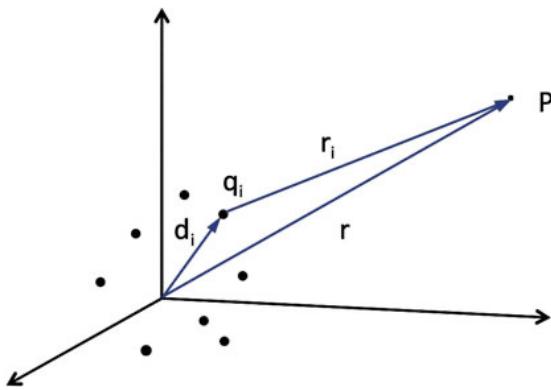
$$U_0(P) = \sum_{i=1}^{i=N} U_{0i}(P) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{i=N} \frac{q_i}{r_i}$$

Geometry affects charges ...

- But charges are stationary in FF methods ...
- Polarizable force fields



Multipole expansions of point charges



Multipole expansion truncated to the second order:

$$U_2(P) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{i=N} \frac{q_i}{r} + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{i=N} \frac{q_i \mathbf{d}_i \cdot \hat{\mathbf{r}}}{r^2} + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{i=N} \frac{q_i}{r^3} \left[\frac{3}{2} (\mathbf{d}_i \cdot \hat{\mathbf{r}})^2 - \frac{1}{2} d_i^2 \right]$$

monopole

Q

dipole

μ

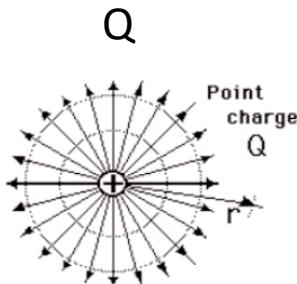
quadrupole

θ

Multipole expansions of point charges

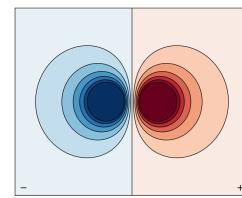
$$U_0(P) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{i=N} \frac{q_i}{r} + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{i=N} \frac{q_i \mathbf{d}_i \cdot \hat{\mathbf{r}}}{r^2} + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{i=N} \frac{q_i}{r^3} \left[\frac{3}{2} (\mathbf{d}_i \cdot \hat{\mathbf{r}})^2 - \frac{1}{2} d_i^2 \right]$$

monopole



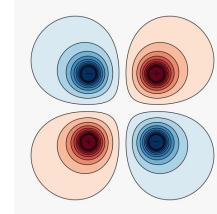
dipole

μ



quadrupole

θ



Distance dependence of multipole interactions

	Q	μ	θ
Q	r^{-1}	r^{-2}	r^{-3}
μ	r^{-2}	r^{-3}	r^{-4}
θ	r^{-3}	r^{-4}	r^{-5}

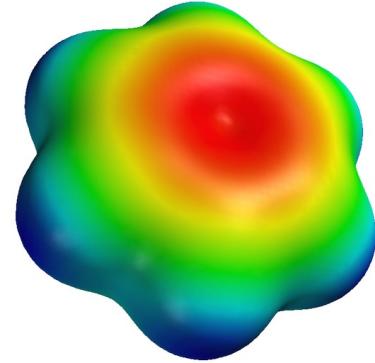
Status Quo

- Few polarizable force fields able to include higher order terms
 - Computationally expensive
- Implicit inclusion of polarizability by scaling of partial charges in many FF methods
 - The partial charges are stationary ...
- Electrostatic energies → Realistic modelling of biological systems

Good parameters, good results!

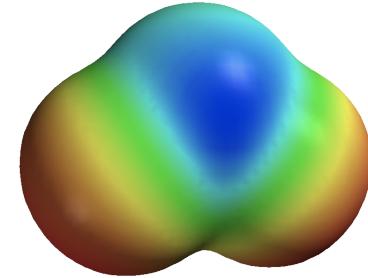
- FF methods are fast!
 - Large systems = no problem
- Results depends on parameter quality
- Electrostatics dominates
 - → we can come a long way with proper charges!
 - Can be assigned by
 - Empirical rules
 - Fitting of calculated electrostatic potential

Charges from electrostatic potential fitting



- Fit a set of point charges to best reproduce the molecular ESP from QM calculations
 - Atomic centers or/and bond midpoints.
- Deeply buried atoms can be problematic
 - ESP charges can fluctuate widely
 - → The RESP method
- ESP charges are superior compared to other schemes → the norm for MD simulations.

Fitting methodology



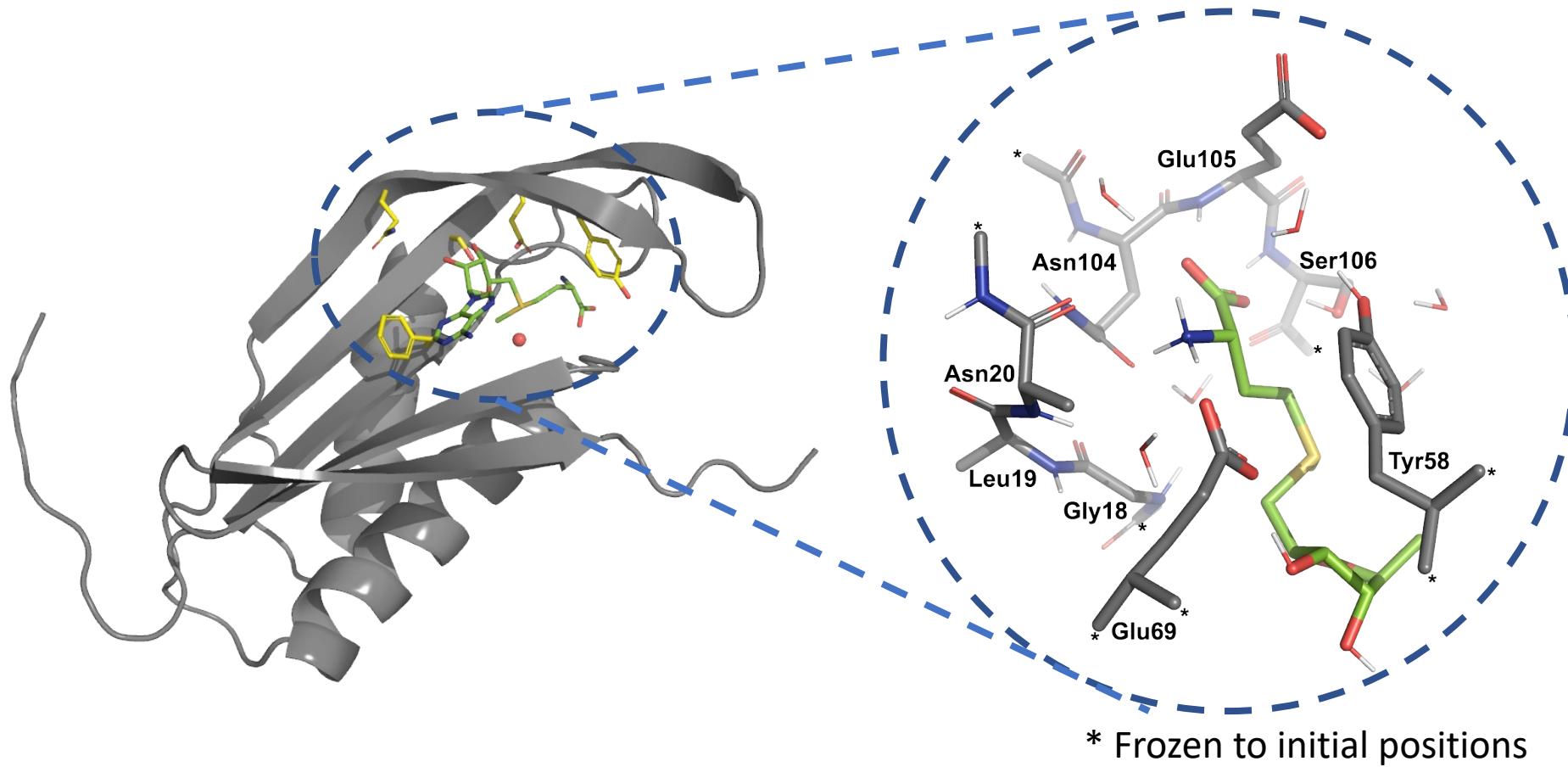
- The electrostatic potential (EP) at \vec{r} :
$$\phi_q(\vec{r}) = \sum_j^{N_q} \frac{q_j}{|\vec{r} - \vec{r}_j|}$$
- QM generates reference EP: $\phi_{QM}(\vec{r})$
- Set of point charges $\{q_j\}$ is then adjusted to minimize the difference between the “QM” EP and the EP due to the atomic charges on grid points surrounding the molecule:

$$F(\{q_j\}) \sum_{grid} \left(\phi_{QM}(\vec{r}_{grid}) - \phi_q(\vec{r}_{grid}) \right)^2 + \lambda \left(\sum_j q_j - q_{tot} \right)$$

DFT reaction energies

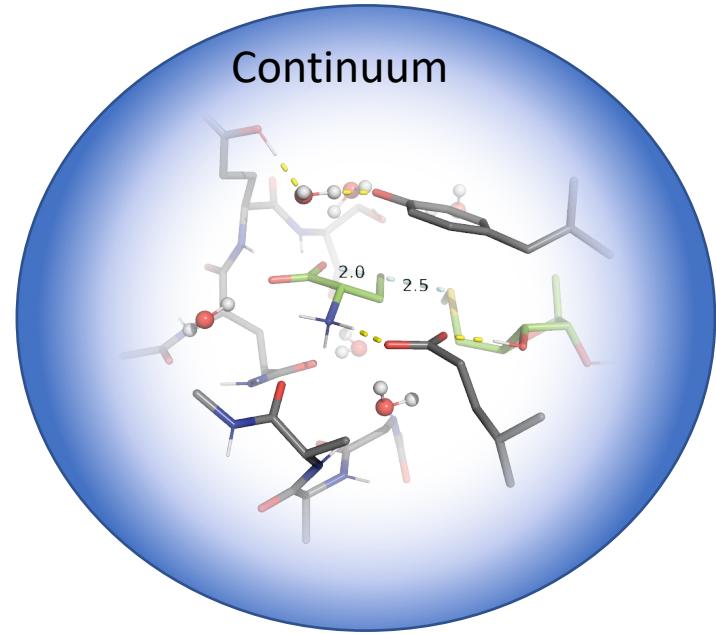
- Case:
- Experimental data does not exist for a new enzyme
 - You need to know
 - Reaction mechanism
 - Reaction free energies
- With the above knowledge, you can use the Empirical Valence bond methods to simulate the whole enzyme and calculate properties like ΔS^\ddagger and ΔH^\ddagger , which is beyond the capacity of QM/MM methods.

DFT cluster model



DFT cluster model

- Start with a small model
 - Investigate mechanisms
- Increase size ~200 atoms
- Geometry optimizations
 - gas phase
 - Dispersion correction included (grimme)
 - Small basis set (6-31g(d,p))
 - Correct for big basis set, solvation ZPE on optimized structures

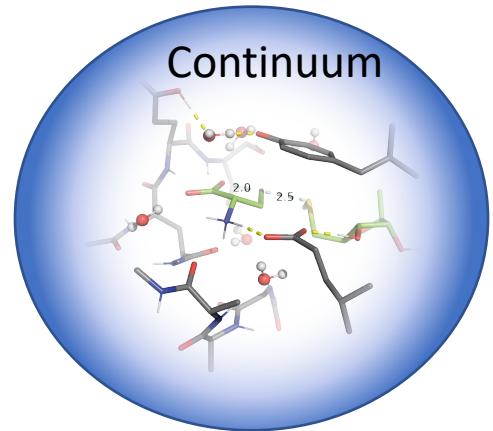


Gaussian input file

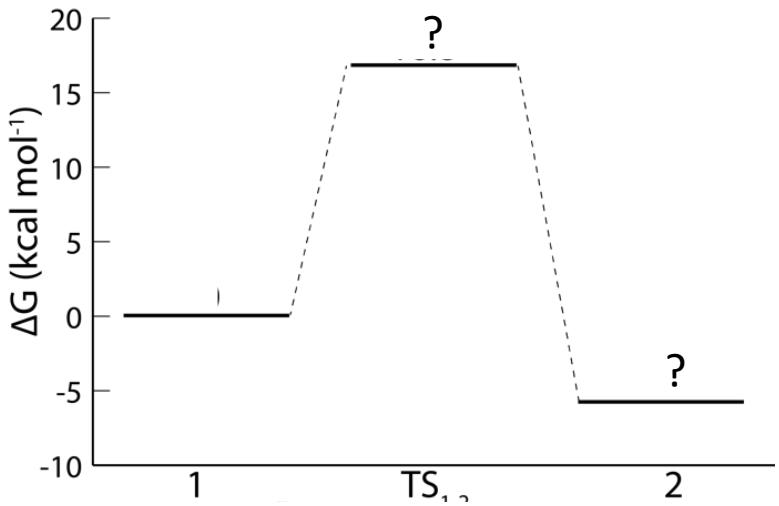
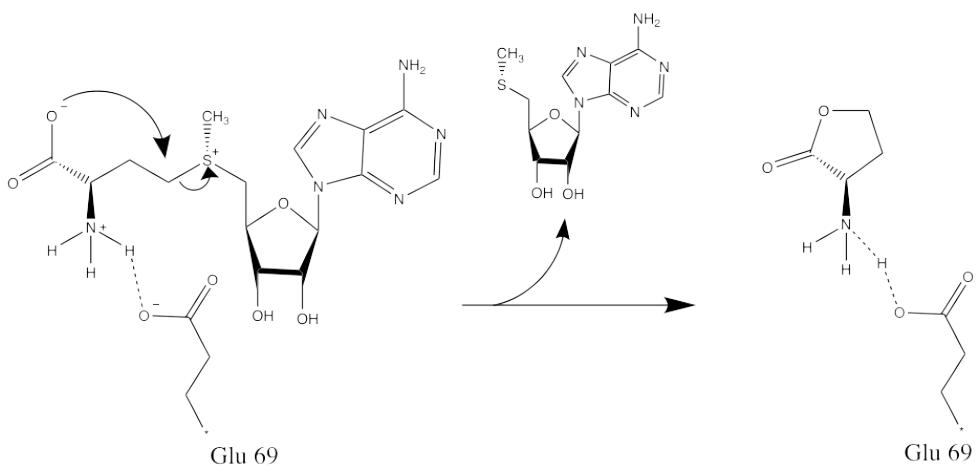
```
# opt=(calcfc,ts,noeigentest) rb3lyp/6-31g(d,p) empiricaldispersion=gd3
Title Card Required

-2 1
C          -1.37744800   0.01329800  -0.40699500
C          -0.47307000   0.20100500  -1.55931400
C          -1.38343000   1.00671000   0.56888900
C          0.49252700   1.12437800  -1.54929400
H          -0.64986700  -0.45108900  -2.40829800
C          -2.63290000  -0.86027000  -0.57342300
C          -0.18553600   1.68355200   0.79941300
H          -2.15994100   0.99811700   1.32502800
C          0.77868300   1.96547500  -0.34217000
H          1.14999700   1.26988400  -2.40155500
O          -2.84789900  -1.30692700  -1.71747100
O          -3.32339500  -1.02728900   0.47166500
H          -0.14355000   2.41470200   1.60541400
O          0.68752500   3.35602600  -0.74021100
H          1.179670600  rev 1.75341000   TS 0.00481700
```

Find the transition state(s)

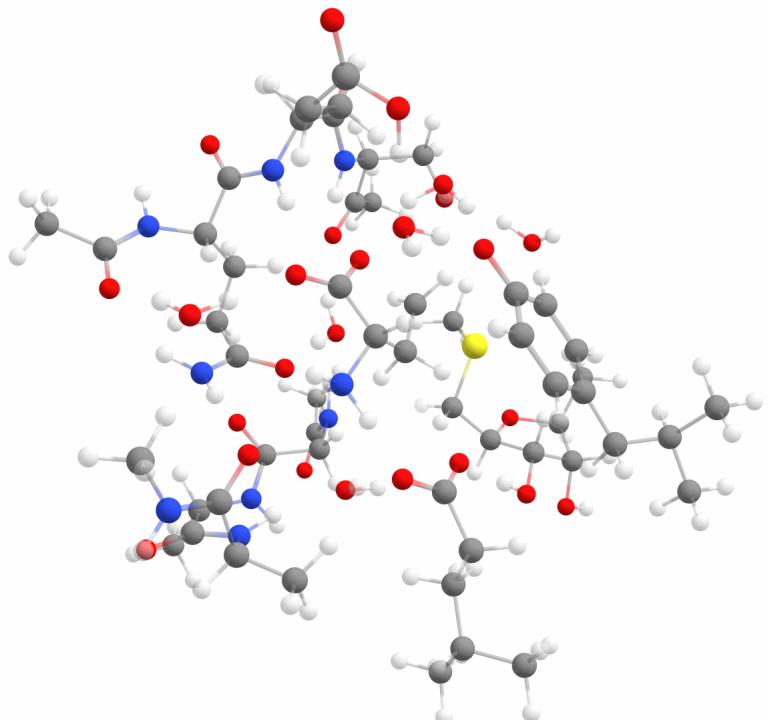
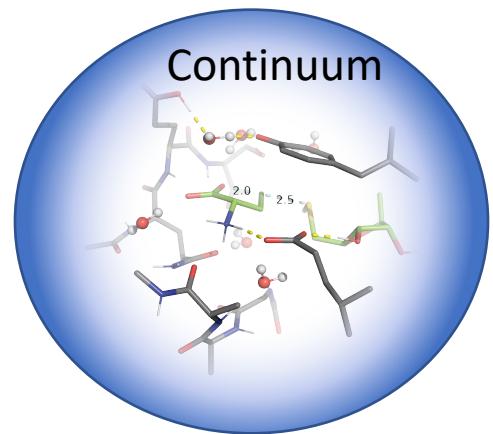


- Use your chemical intuition and guess
- #p opt(ts, calcfc, noeigentest) b3lyp/6-31g(d,p)



Find the transition state(s)

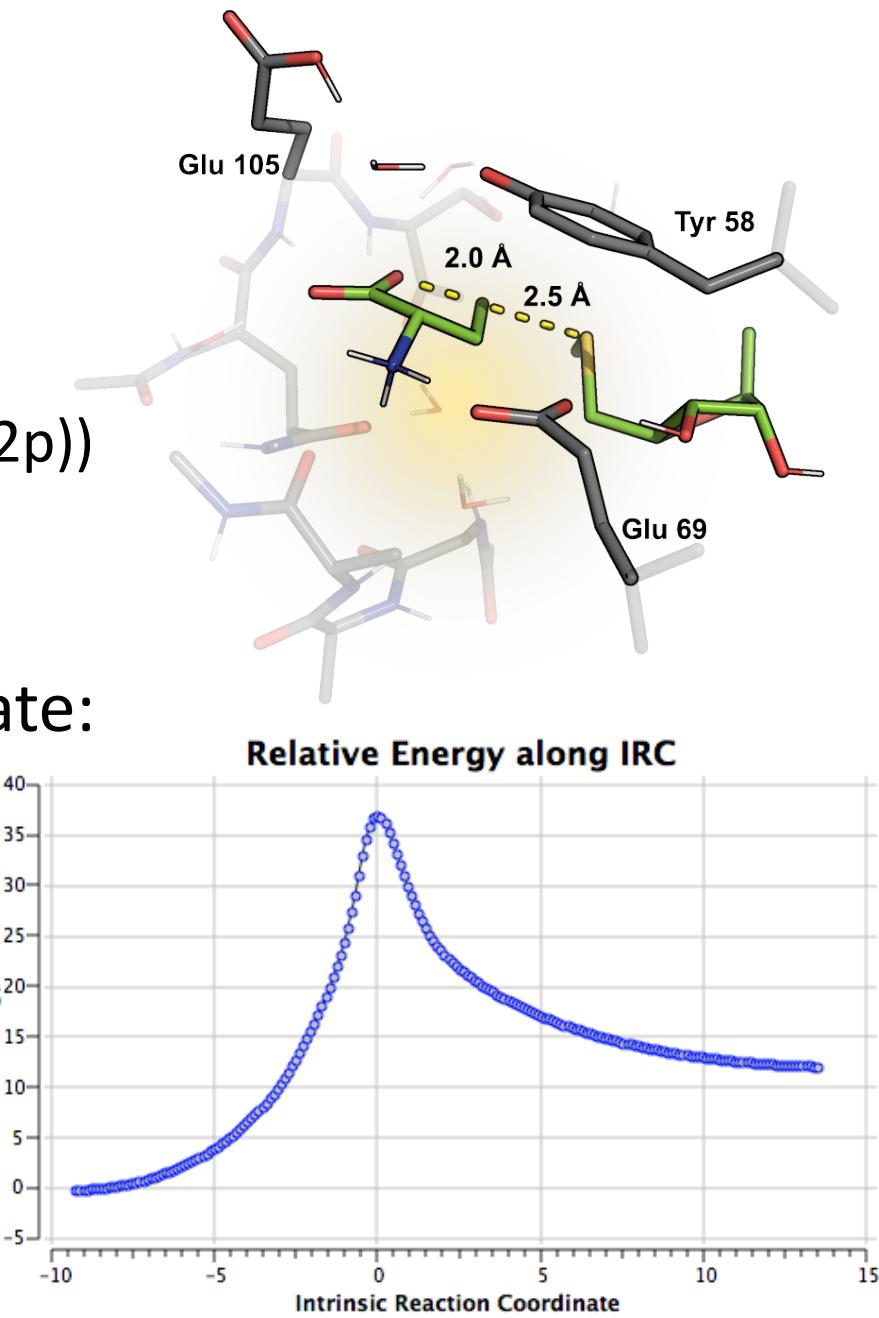
- Converged? Do frequency calculation
- TS – imaginary frequency



	f	int	
1	-344.840	759.43	A
2	-17.710	3.47	A
3	-6.190	0.26	A
4	-4.560	1.04	A
5	10.350	0.06	A
6	13.120	0.72	A
7	17.620	0.93	A
8	19.560	0.47	A
9	20.980	0.87	A
10	23.710	1.10	A
11	26.780	0.00	A
12	28.310	0.34	A
13	30.550	1.09	A

Optimized TS

- Single point calculations
 - Big basis set (6-311G+(2d,2p))
 - Solvation (SMD f.ex)
 - ZPE (from frequency)
- Intrinsic reaction coordinate:

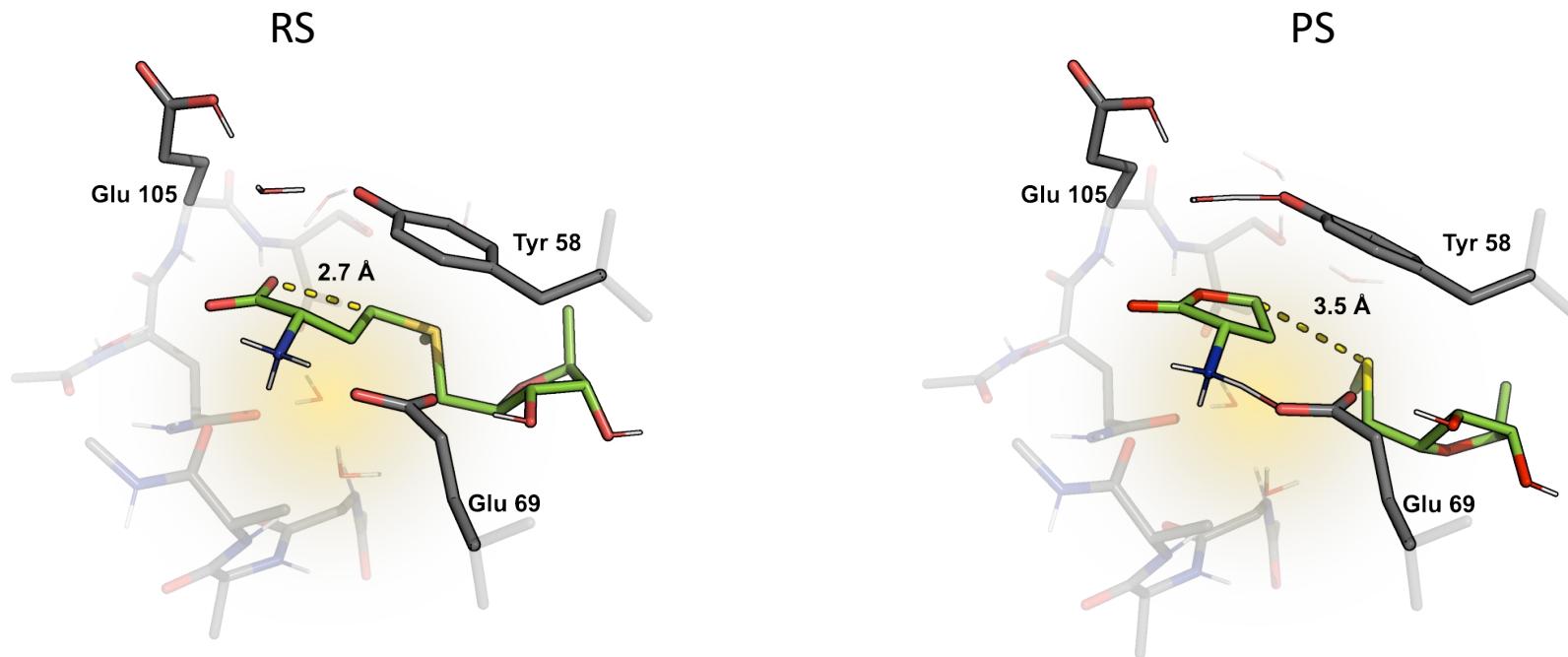


Intrinsic Reaction Coordinate (IRC)

- Useful to confirm that the located saddlepoint (TS) lies on the minimum energy path between 2 assumed minima (instead of being a saddle point for something else).
- If TS is correct, the IRC should bring you towards the reactant and product states of the reaction.

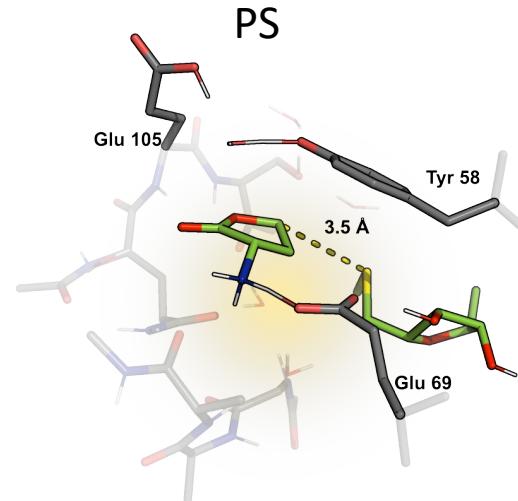
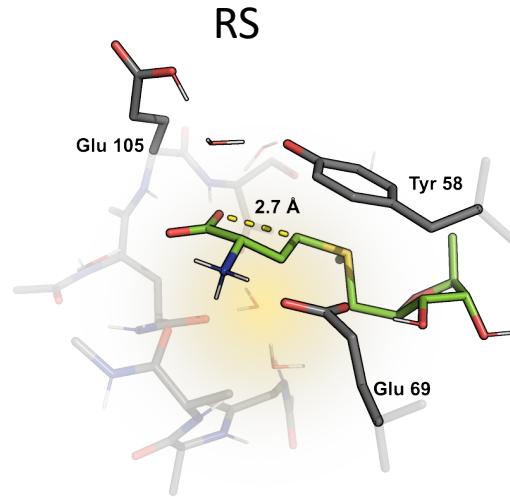
Optimize IRC endpoints

- Reactant and product states
 - #p opt b3lyp/6-31g(d,p) empiricaldispersion=gd3

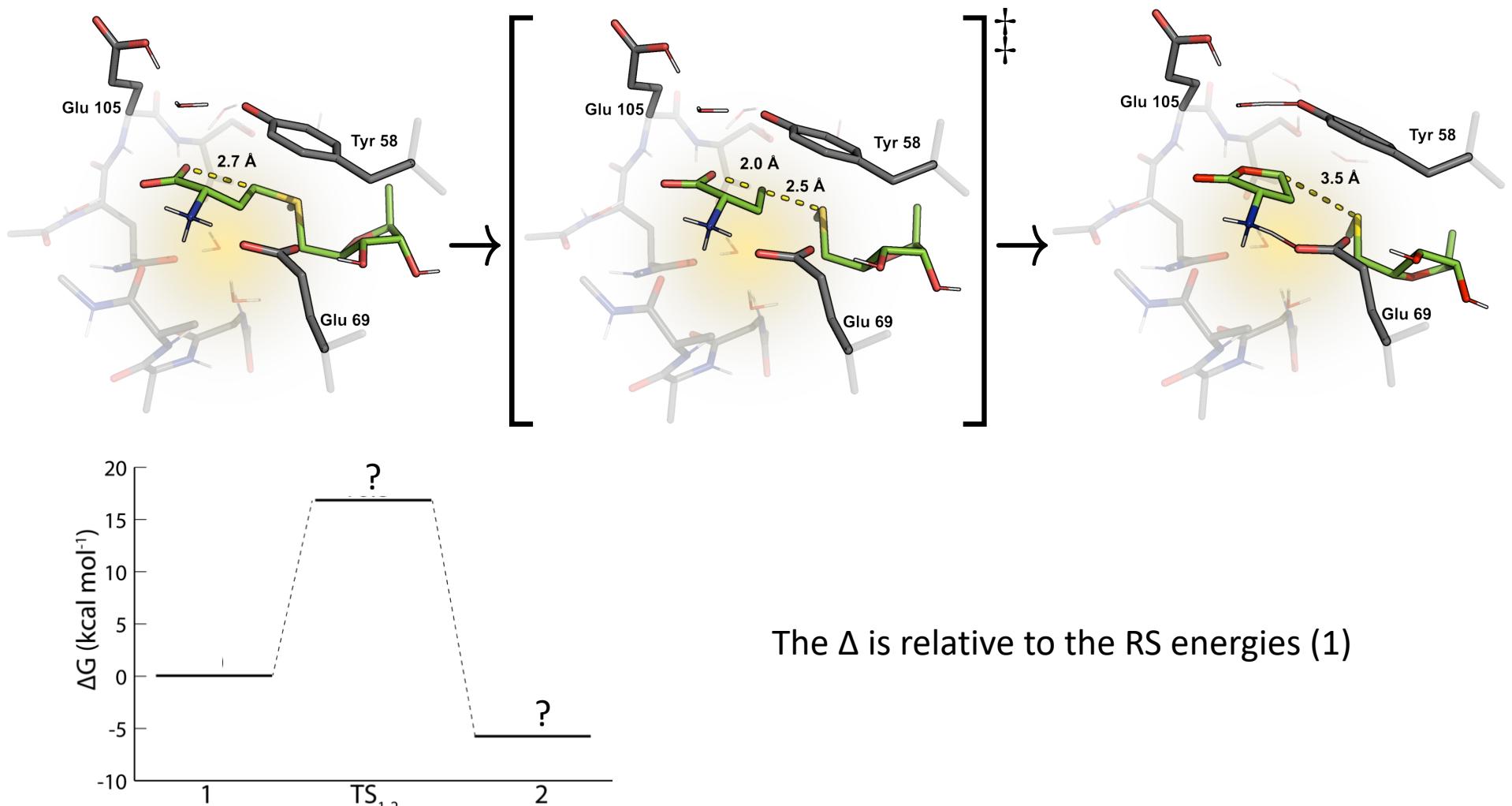


Optimized RS and PS

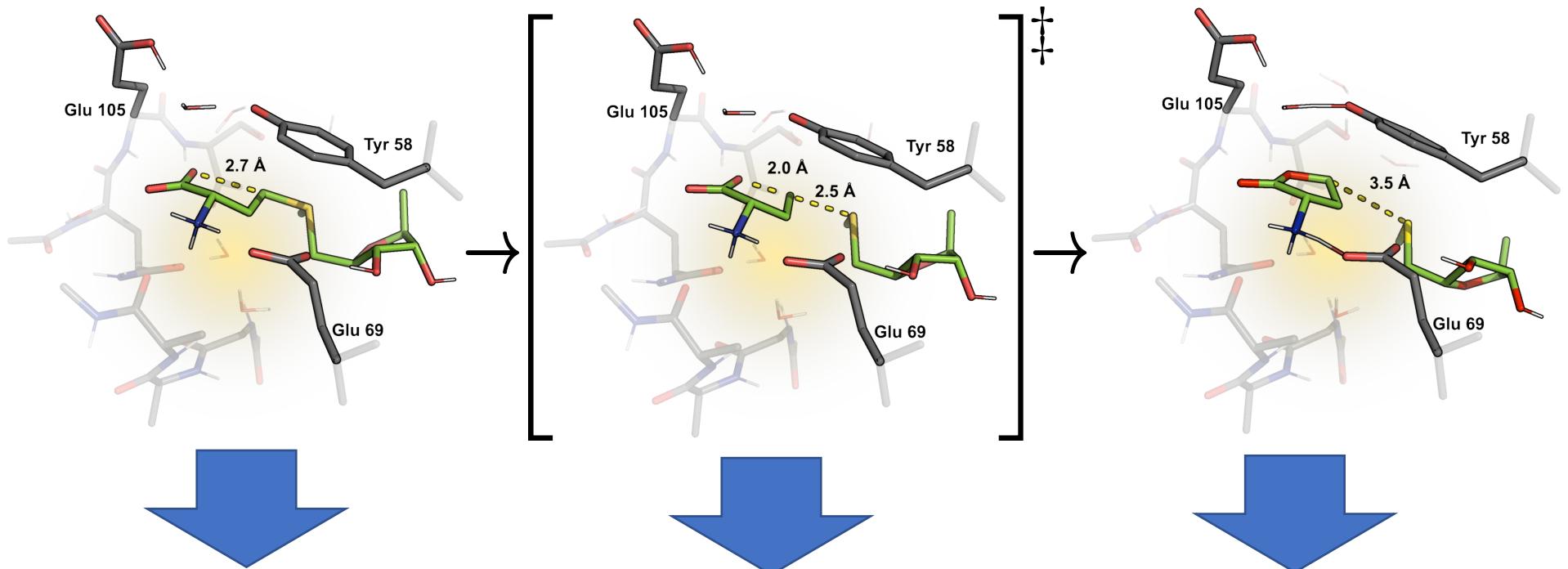
- Single point calculations
 - Big basis set (6-311G+(2d,2p))
 - Solvation (SMD f.ex)
 - Frequency calculation (ZPE)



Putting it all together

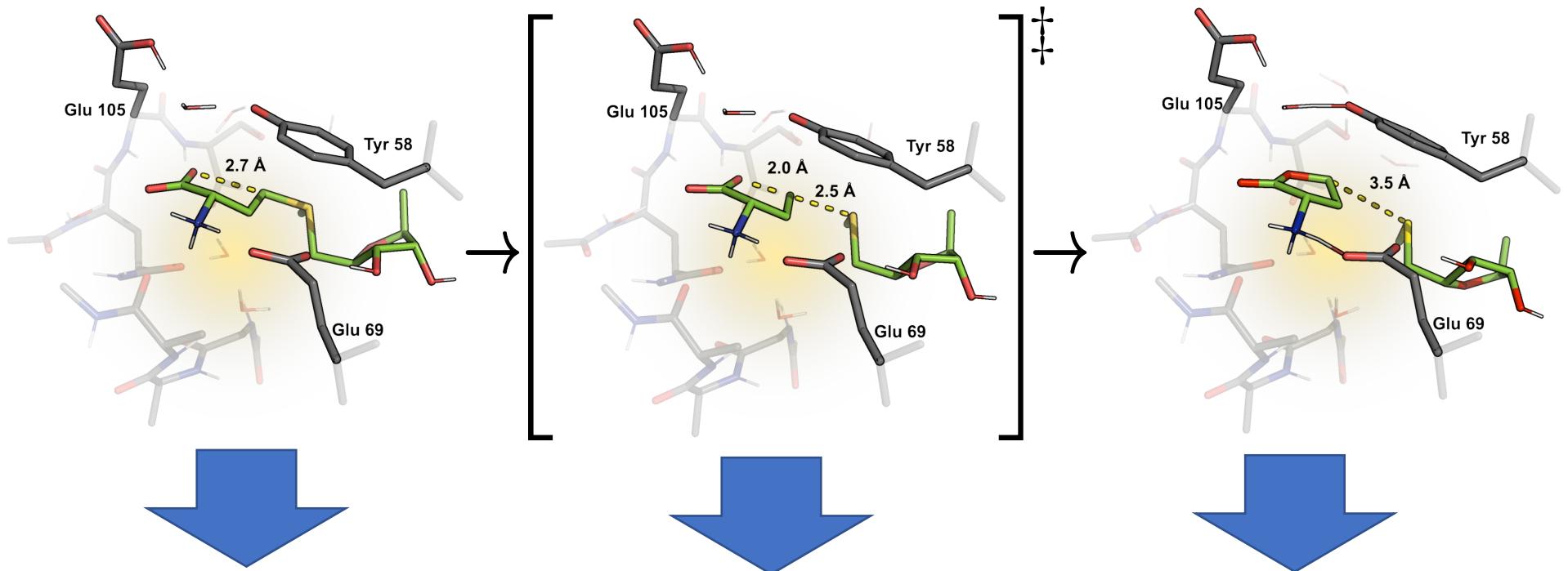


Putting it all together



- Optimized (6-31g d(d,p))
 - Frequency (ZPE)
 - SP with solvation
 - SP big basis (6-311g+(2d,2p))
- Optimized (6-31g d(d,p))
 - Frequency (ZPE)
 - SP with solvation
 - SP big basis (6-311g+(2d,2p))
- Optimized (6-31g d(d,p))
 - Frequency (ZPE)
 - SP with solvation
 - SP big basis (6-311g+(2d,2p))

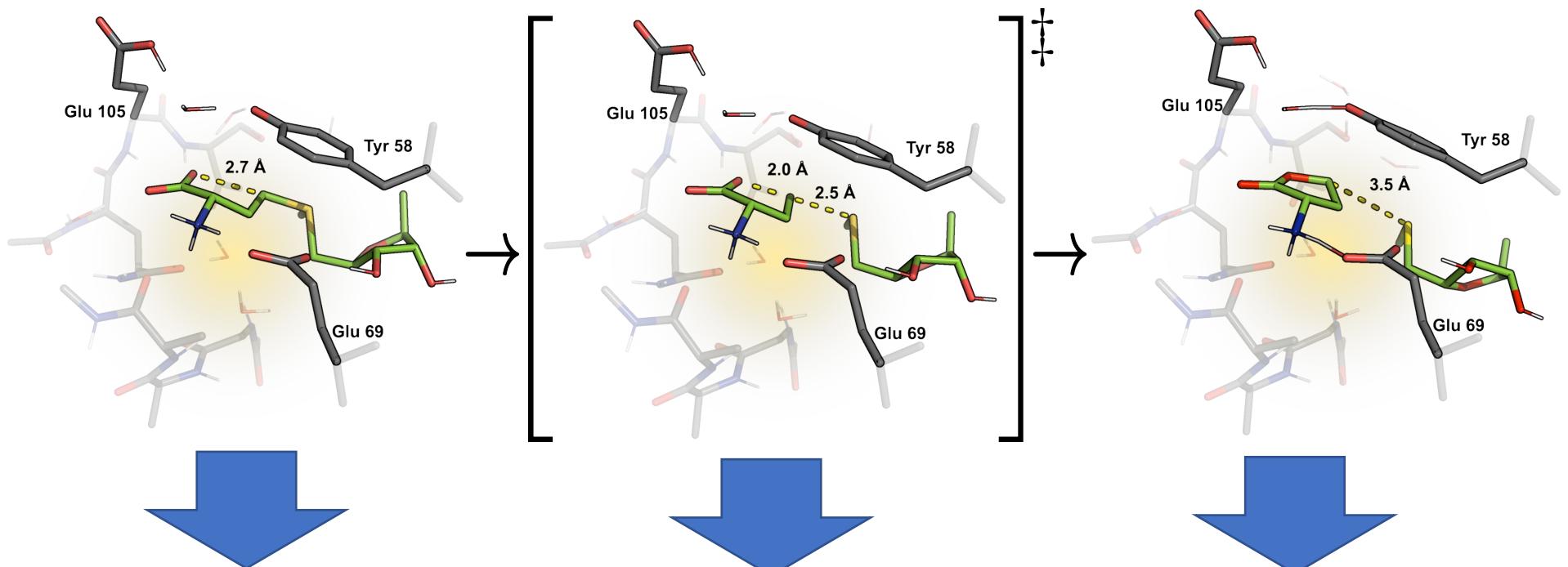
Putting it all together



- E_{gas}
- δG (correction from ZPE)
- E_{solv}
- E_{big}
- $\Delta E_{\text{solv}} = E_{\text{solv}} - E_{\text{gas}}$
- E_{gas}
- δG (correction from ZPE)
- E_{solv}
- E_{big}
- $\Delta E_{\text{solv}} = E_{\text{solv}} - E_{\text{gas}}$
- E_{gas}
- δG (correction from ZPE)
- E_{solv}
- E_{big}
- $\Delta E_{\text{solv}} = E_{\text{solv}} - E_{\text{gas}}$

Geir Villy Isaksen

Putting it all together



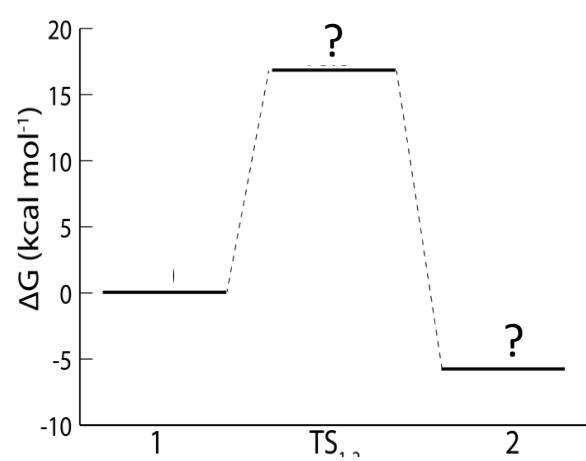
- δG (correction from ZPE)
 - E_{big}
 - $\Delta E_{\text{solv}} = E_{\text{solv}} - E_{\text{gas}}$
- δG (correction from ZPE)
 - E_{big}
 - $\Delta E_{\text{solv}} = E_{\text{solv}} - E_{\text{gas}}$
- δG (correction from ZPE)
 - E_{big}
 - $\Delta E_{\text{solv}} = E_{\text{solv}} - E_{\text{gas}}$

Putting it all together

- δG (RS)
- E_{big} (RS)
- ΔE_{solv} (RS)
- δG (TS)
- E_{big} (TS)
- ΔE_{solv} (TS)
- δG (PS)
- E_{big} (PS)
- ΔE_{solv} (PS)

$$\begin{aligned}\Delta G^\ddagger &= (E_{\text{big}}^{\text{TS}} - E_{\text{big}}^{\text{RS}}) + (\partial G^{\text{TS}} - \partial G^{\text{RS}}) + (\Delta E_{\text{solv}}^{\text{TS}} - \Delta E_{\text{solv}}^{\text{RS}}) \\ &= \Delta E_{\text{big}}^\ddagger + \Delta \partial G^\ddagger + \Delta \Delta E_{\text{solv}}^\ddagger\end{aligned}$$

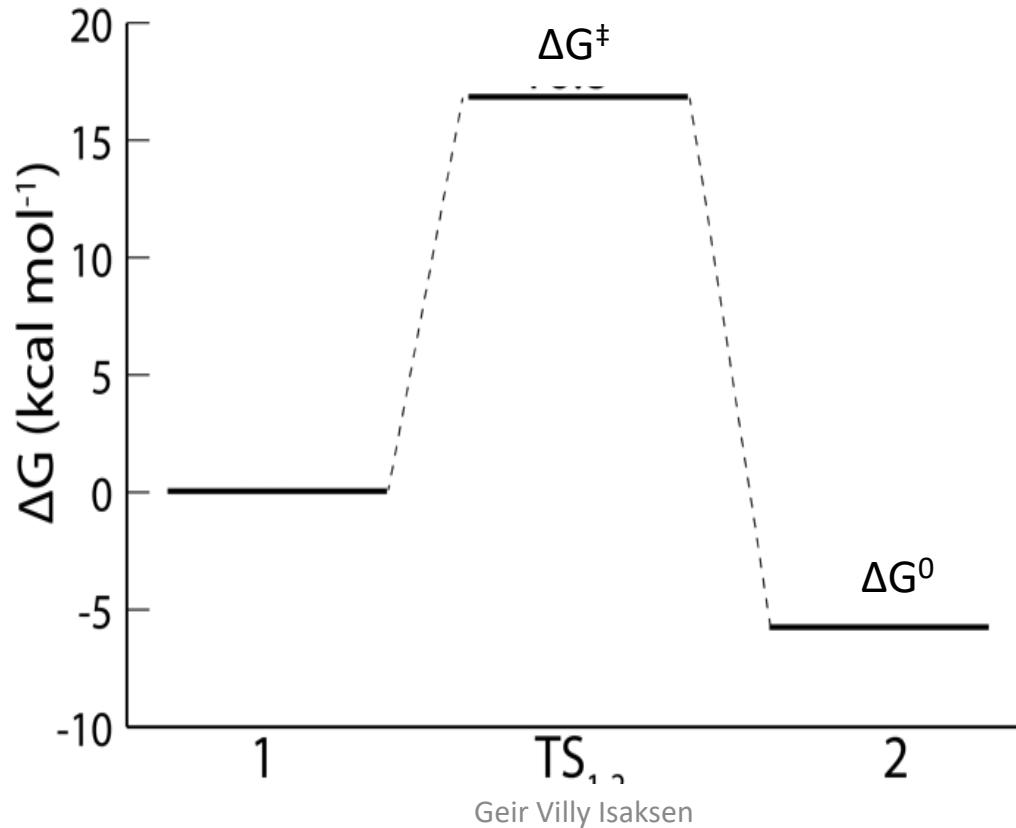
$$\begin{aligned}\Delta G^0 &= (E_{\text{big}}^{\text{PS}} - E_{\text{big}}^{\text{RS}}) + (\partial G^{\text{PS}} - \partial G^{\text{RS}}) + (\Delta E_{\text{solv}}^{\text{PS}} - \Delta E_{\text{solv}}^{\text{RS}}) \\ &= \Delta E_{\text{big}}^0 + \Delta \partial G^0 + \Delta \Delta E_{\text{solv}}^0\end{aligned}$$



Putting it all together

$$\Delta G^\ddagger = \Delta E_{\text{big}}^\ddagger + \Delta \partial G^\ddagger = \Delta \Delta E_{\text{solv}}^\ddagger$$

$$\Delta G^0 = \Delta E_{\text{big}}^0 + \Delta \partial G^0 = \Delta \Delta E_{\text{solv}}^0$$



- <https://www.sciencedirect.com/topics/medicine-and-dentistry/density-functional-theory>