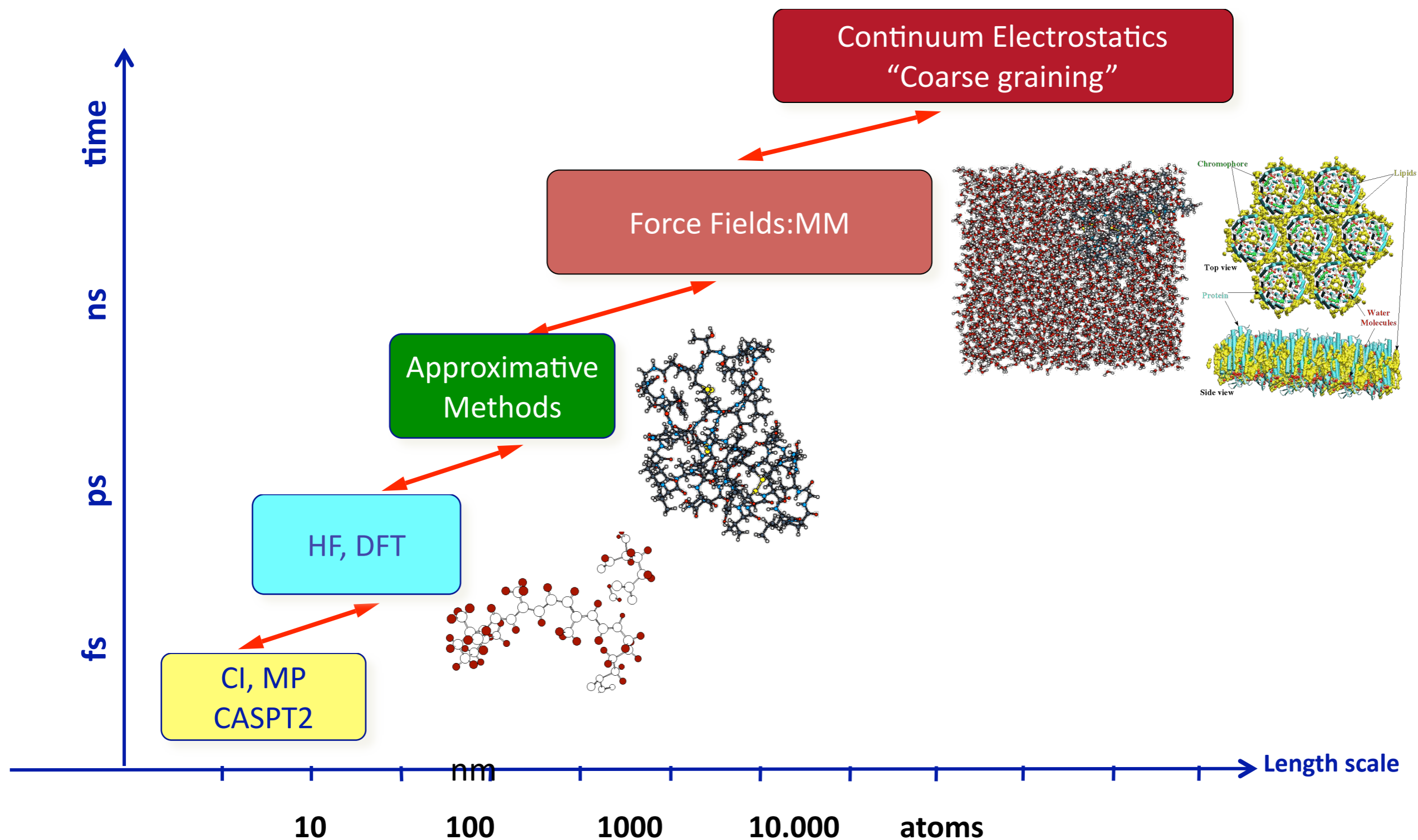# Multi-scale models in theoretical biophysics

Methodology: time scale

Molecular **D**ynamics: MD

$$V = \sum_{bonds} k_b \left(b - b_0\right)^2 + \sum_{angles} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{dihedrals} \sum_{n=1}^{N} k_\phi^{(n)} \left[1 + \cos\left((n\phi - \delta\right)\right] + \sum_{impropers} k_\omega \left(\omega - \omega_0\right)^2$$

$$+ \sum_{i,j} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^6\right] + \sum_{i,j} \left(\frac{q_i q_j}{Dr_{ij}}\right)$$

Molecular Dynamics (MD):

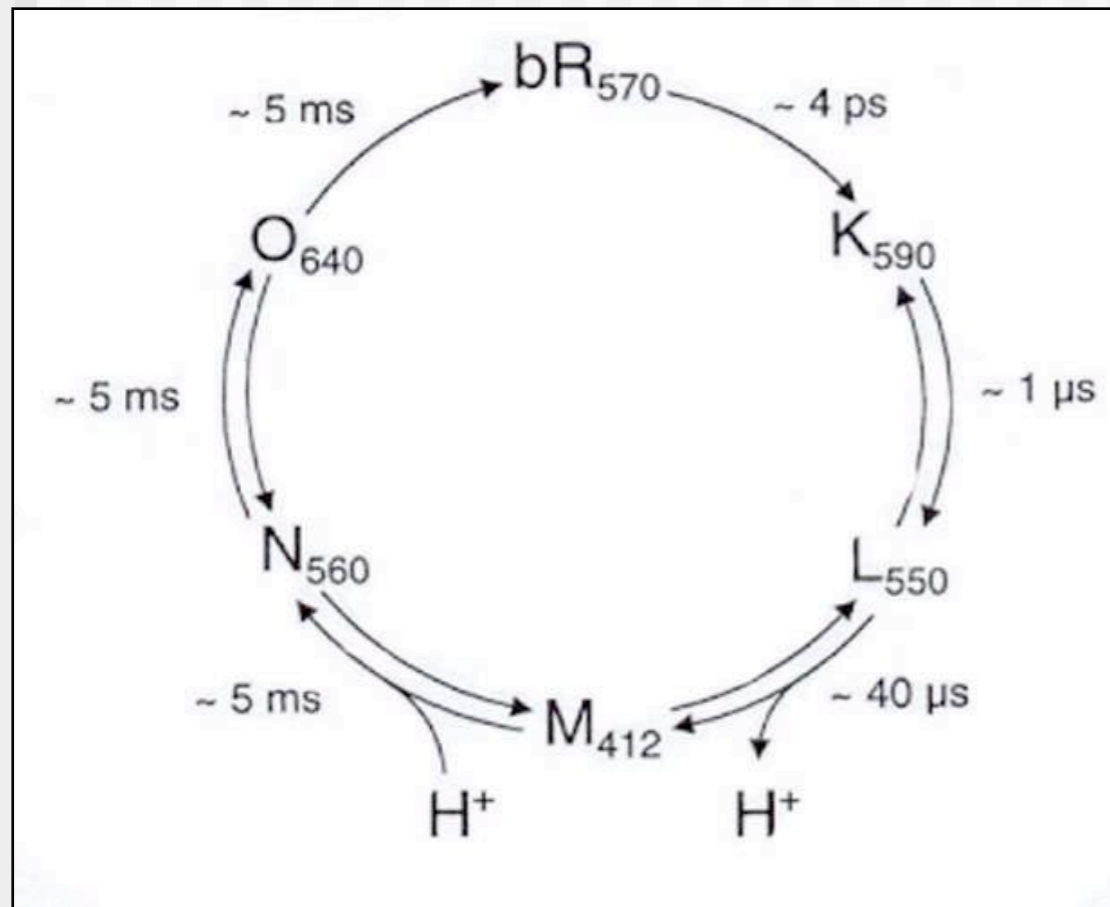Numerical integration of Newtons equation of motion F=m*a with

timestep: ~1fs
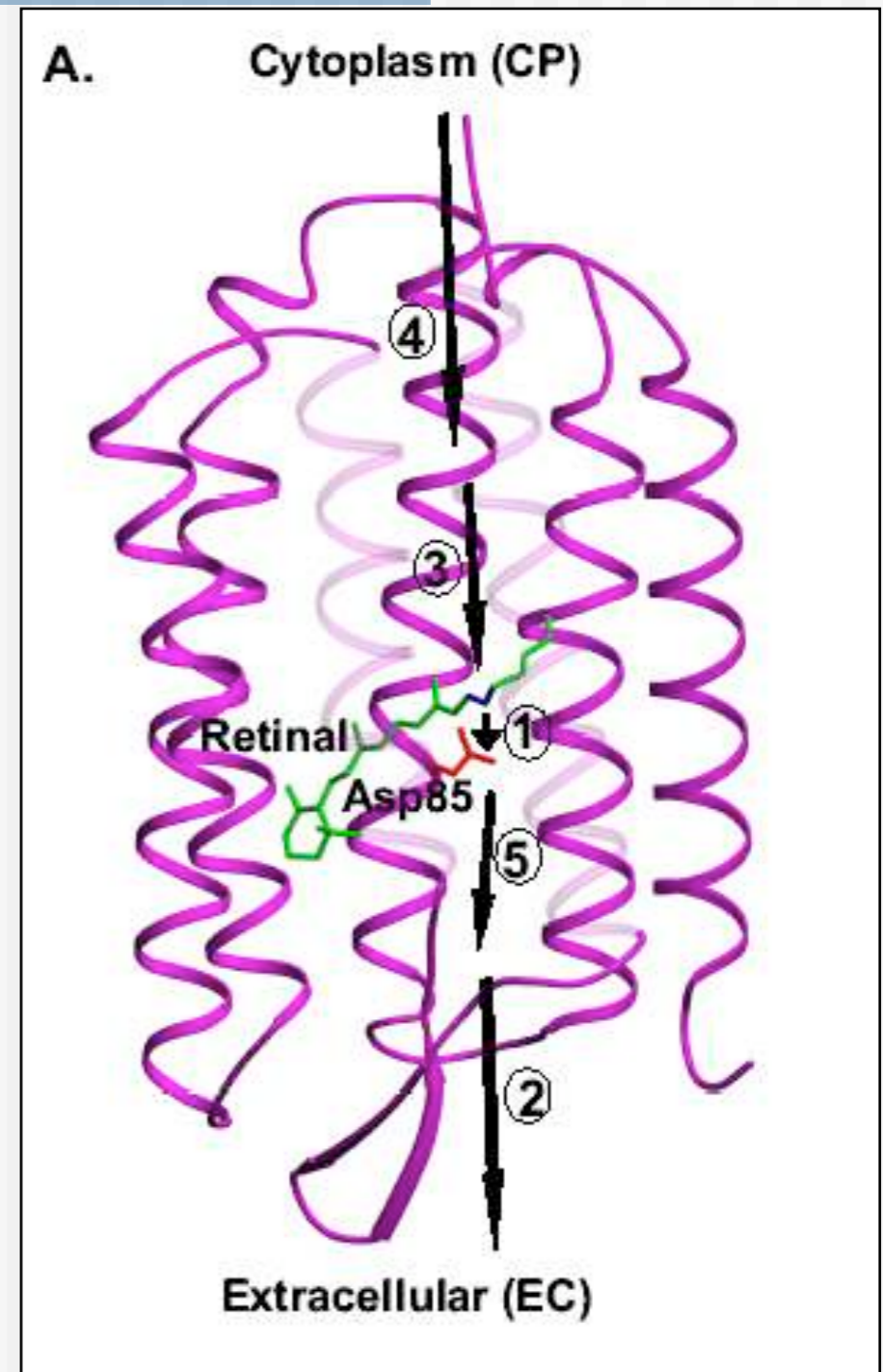
➔ trajectories

➔ 1ps dynamics: 1000 force evaluations

$$F_i = -\frac{\partial V}{\partial R_i}$$

Bacteriorhodopsin

2 Problems:
- size: 1000-100.000 atoms
- time scales: > ns

How to study reactions and (rare) dynamical events

- direct MD

- reaction path methods

  - NEB (nudged elastic band, Jonsson)

  - **CPR** (conjugate peak refinement, Fischer, Karplus)

  - dimer method (Jonsson)

- accelerated MD

  - hyperdynamics (Voter)

  - chemical flooding (Grubmüller)

  - metadynamics (Parinello)

  - replica exchange

- free energy sampling techniques

  - **umbrella sampling**

  - free energy perturbation

  - transition path sampling
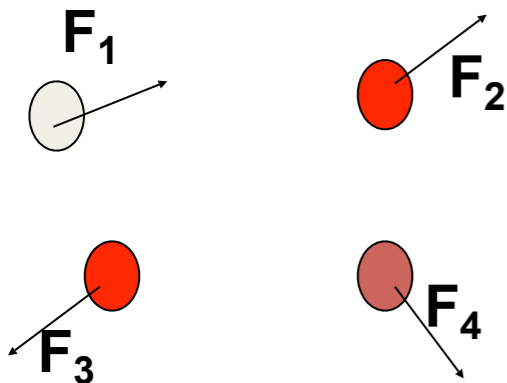
"**M**olecular **D**ynamics":   **MD**

Numerical Integration of Newtons equation of motion

F=m*a

start:

Calculate E and forces on atoms
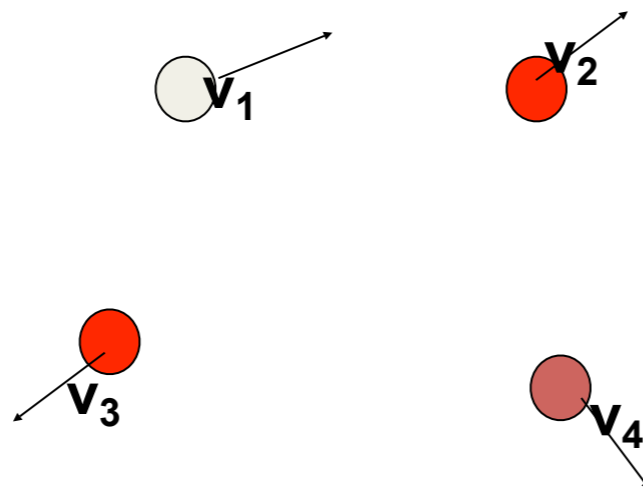
$$F_i = -\frac{\partial E}{\partial R_i}$$

F_1

F_2

F_3

F_4

step1:

atoms are accelerated in direction of forces
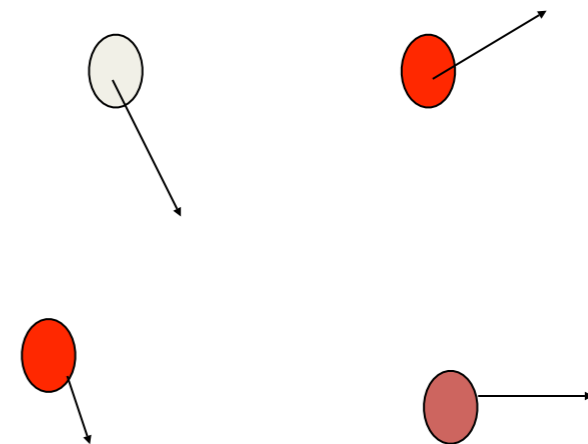
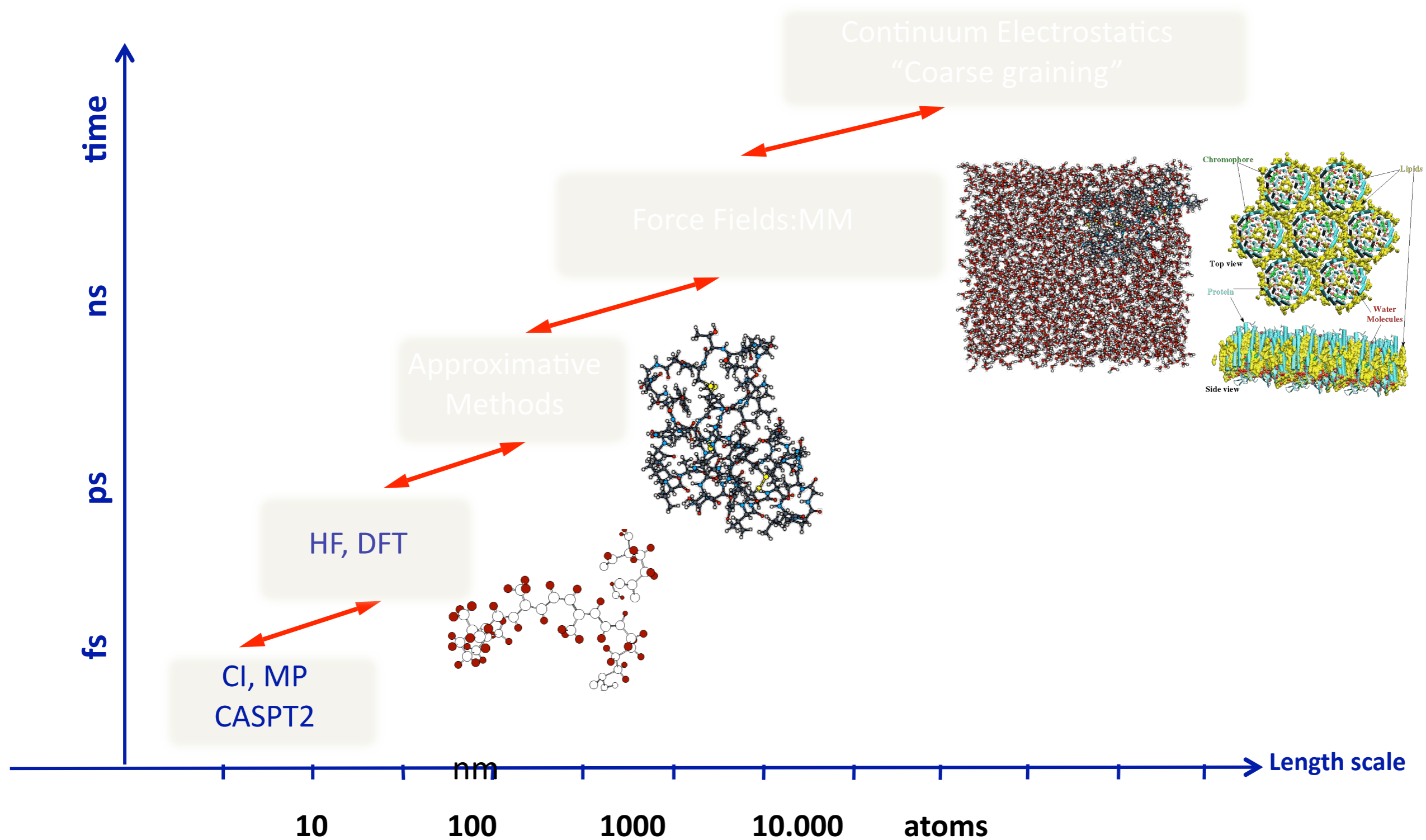**a= F/m**

Velocity v

v_1

v_2

v_3

v_4

Step 2:

Calculate E and forces for new structure

$$F_i = -\frac{\partial E}{\partial R_i}$$

Multi-scale models in theoretical biophysics
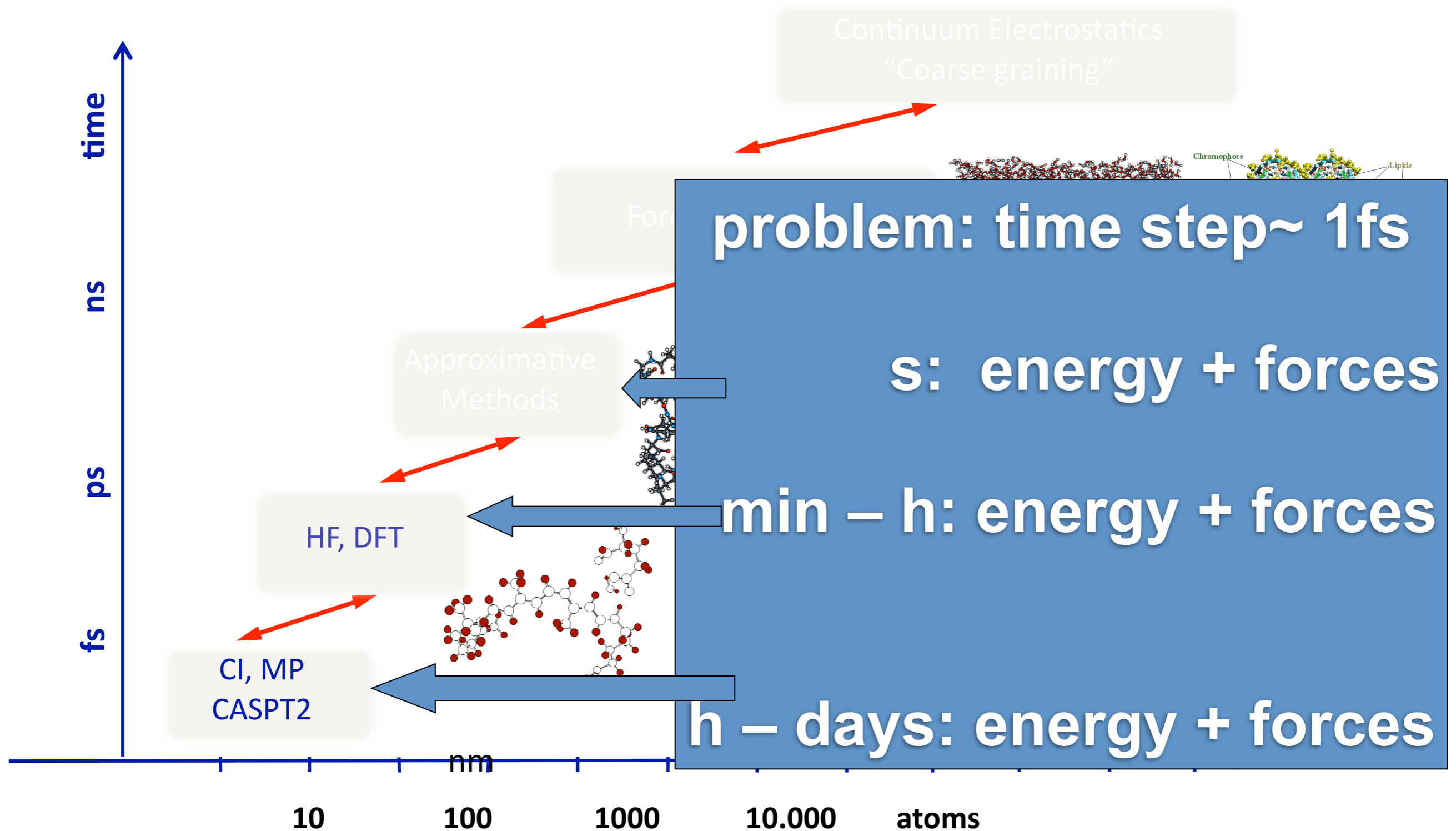
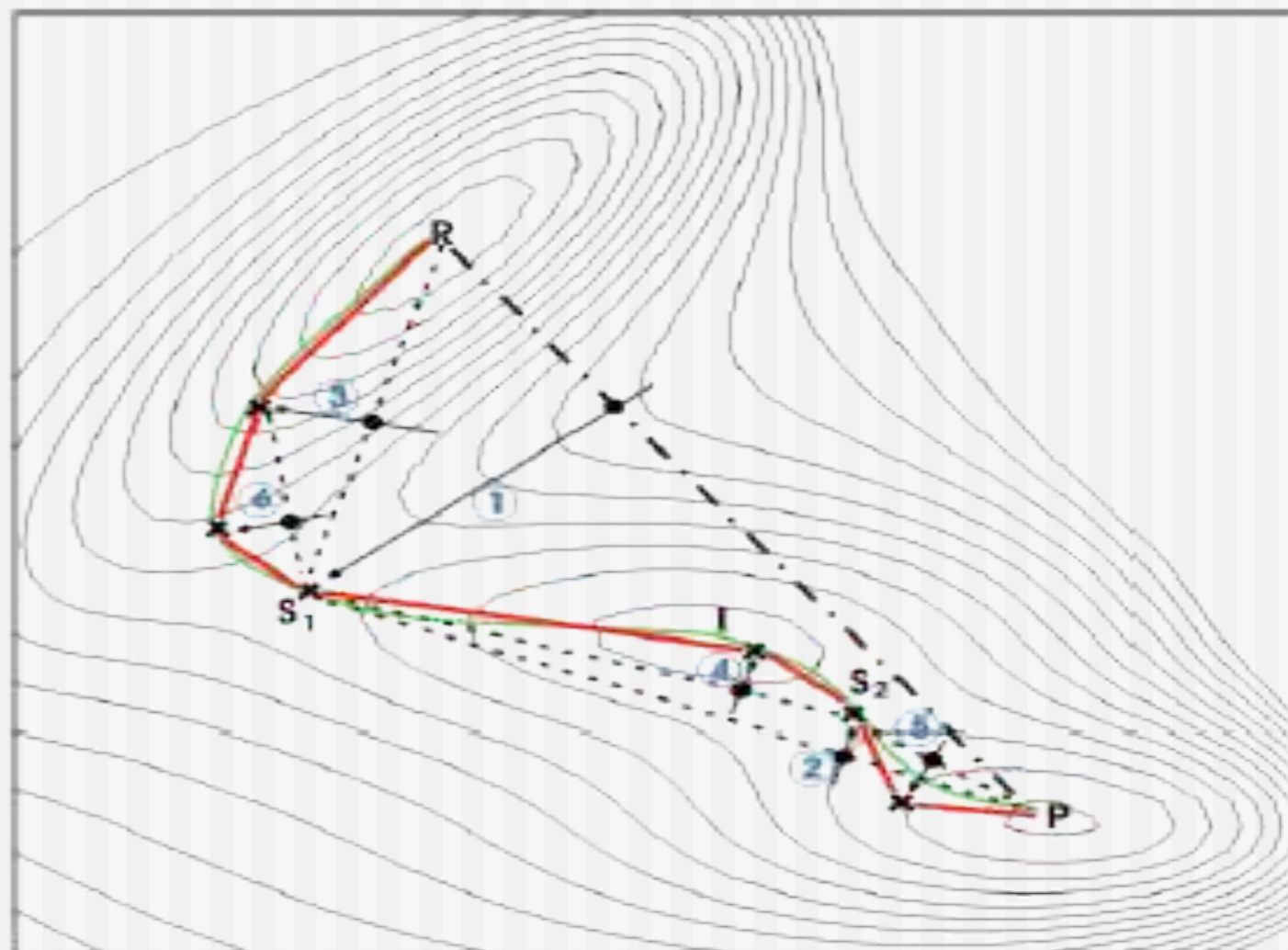# How to study reactions and (rare) dynamical events

- direct MD

How to study reactions and (rare) dynamical events

- direct MD

- reaction path methods

  - NEB (nudged elastic band, Jonsson)

  - CPR (conjugate peak refinement, Fischer, Karplus)

  - dimer method (Jonsson)

Reaction path methods: e.g. CPR

- **CPR: Conjugate peak refinement (Fischer & Karplus)**

- **NEB: nudged elastic band**

- **dimer method**

These methods are very 'costly', i.e.

they require the calculation of energy and

forces several 1.000-10.000 times

=> prohibitive for DFT/ab initio methods

CPR: fig. from S. Fischer

'Problem' of potential energy (MEP)

Different energy profiles for different
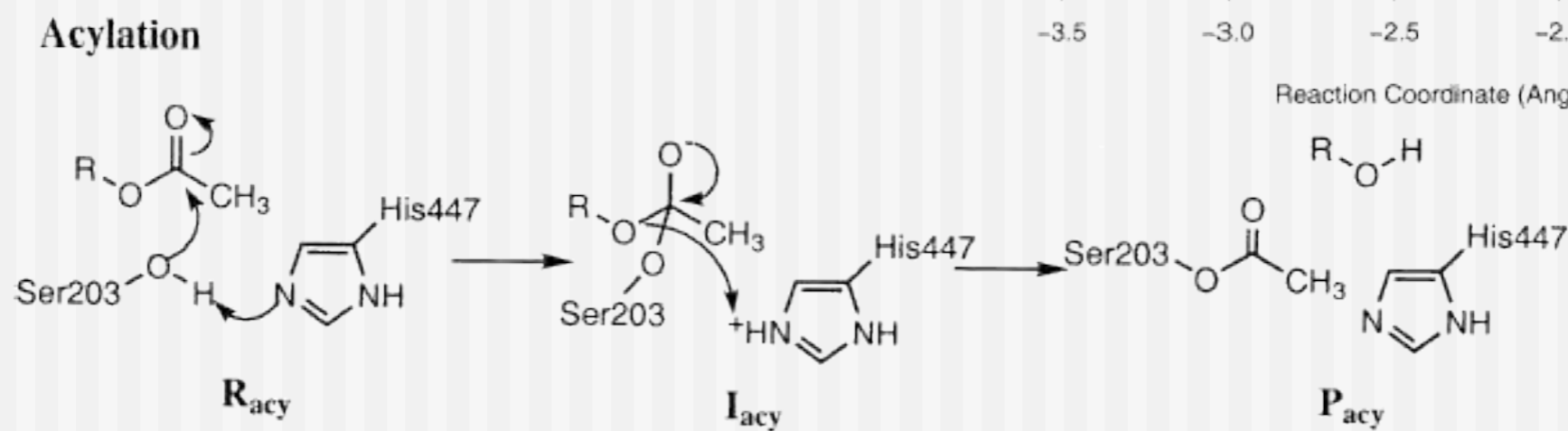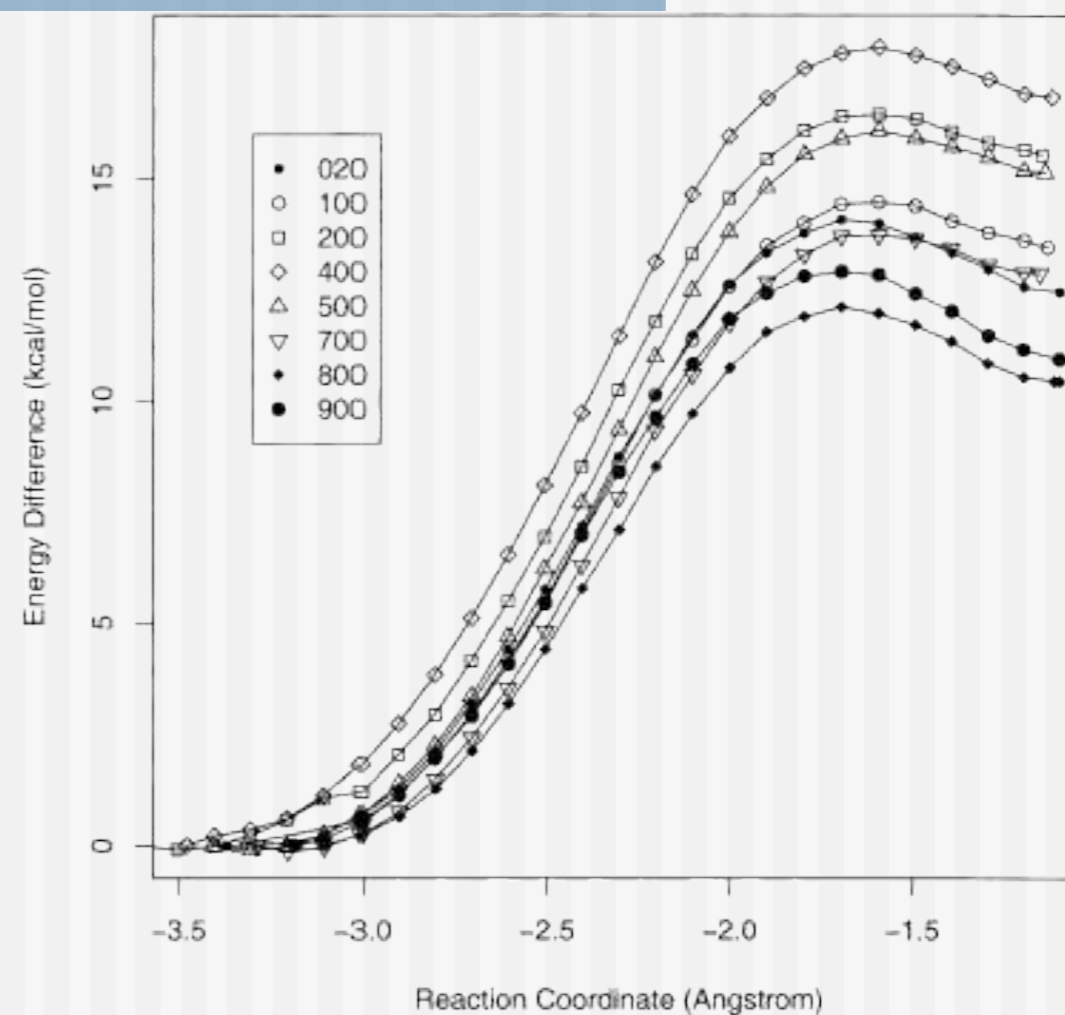
protein conformations

Figure 1. Acylation reaction mechanism of acetylcholine catalyzed by AChE.

Zhang et al JPCB 107 (2003) 44459

'Problem' of potential energy (MEP)

Different energy profiles for different

prot

A) One always has to average over the different conformations of the environment :

Total energy➜ inner energy

E➜ U

B) Entropy is often as important as accurate total energy :

U➜   F

Figure 1. Acylation reaction mechanism of acetylcholine catalyzed by AChE.

Zhang et al JPCB 107 (2003) 44459

Problems with the PES: complex energy landscape

- differences in protein conformations

    (starting the reaction path calculation)

- problems along the reaction pathway

        * flipping of water molecules

        * size of movable MM region

         different H-bonding pattern

→ average over these effects:

    potential of mean force/free energy
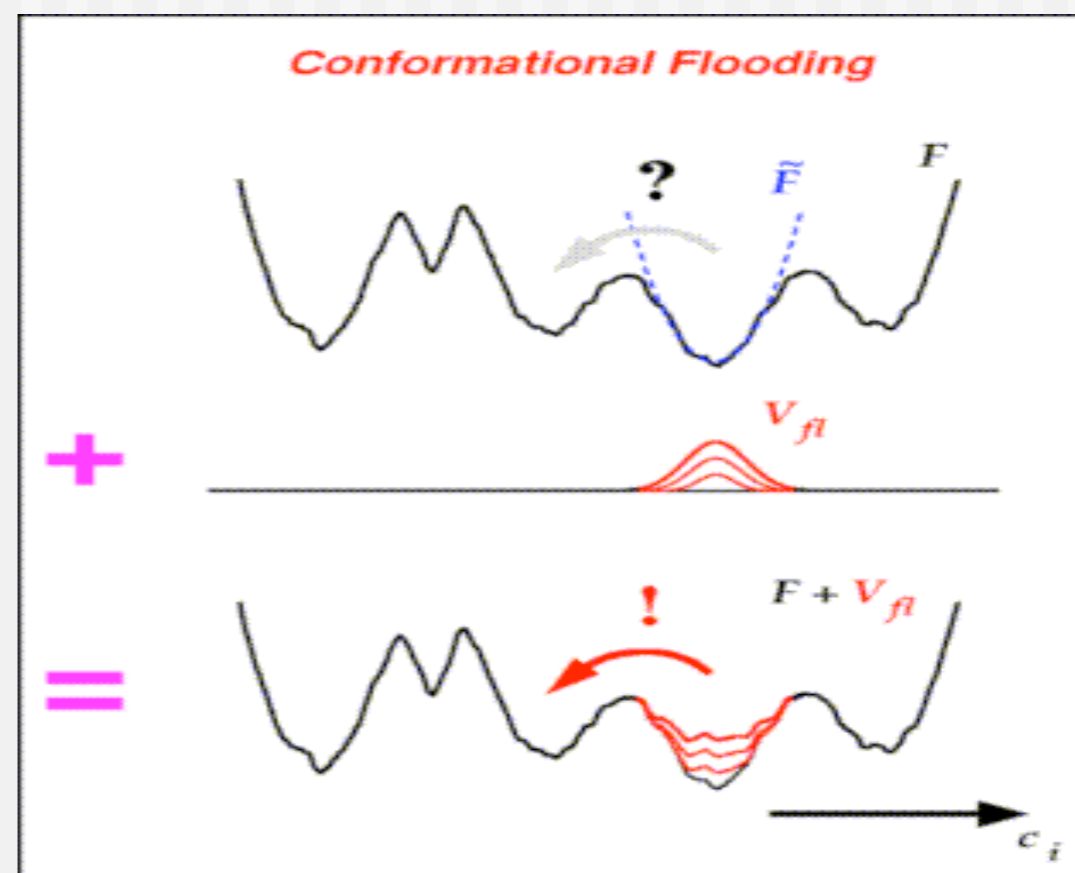
How to study reactions and (rare) dynamical events

- direct MD

- reaction path methods

  - **NEB (nudged elastic band, Jonsson)**

  - **CPR (conjugate peak refinement, Fischer, Karplus)**

  - **dimer method (Jonsson)**

- accelerated MD

  - **hyperdynamics (Voter)**

  - **chemical flooding (Grubmüller)**

  - **metadynamics (Parinello)**

  - **replica exchange**

Fig. from H. Grubmüller

How to study reactions and (rare) dynamical events

- direct MD

- reaction path methods

    - **NEB (nudged elastic band, Jonsson)**

    - **CPR (conjugate peak refinement, Fischer, Karplus)**

    - **dimer method (Jonsson)**

- accelerated MD

    - **hyperdynamics (Voter)**

    - **chemical flooding (Grubmüller)**

    - **metadynamics (Parinello)**

    - **replica exchange**

- free energy sampling techniques

    - **umbrella sampling**

    - **free energy perturbation**

    - **transition path sampling**

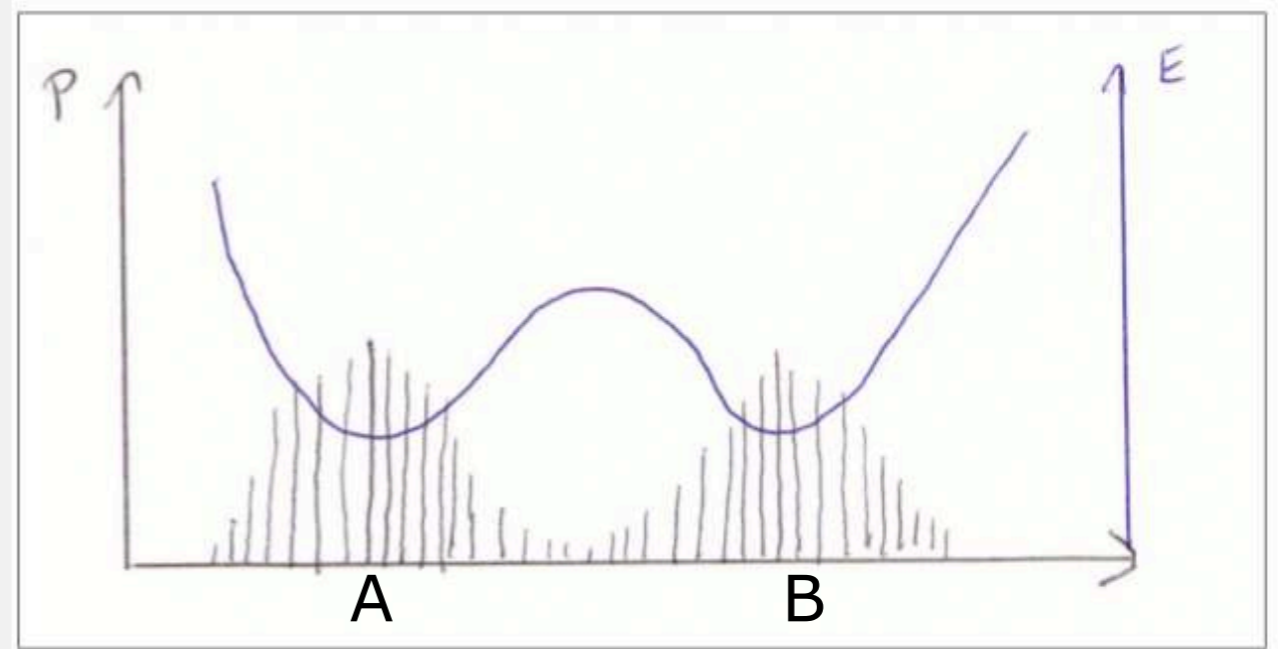# Calculate potential of mean force

Sample the states A and B in MD:

Free energy is calculated from probabilities:

$$F_B - F_A = -kTln\frac{P(q_B)}{P(q_A)}$$

When the barrier is to high, force the system to cross it with additional potentials:

Subtract these afterwords:

'Umbrella sampling'

How to study reactions and (rare) dynamical events

- direct MD

- reaction path methods

  'no entropy'

  - NEB (nudged elastic band, Jonsson)

  - CPR (conjugate peak refinement, Fischer, Karplus)

  - dimer method (Jonsson)

- accelerated MD

  'which mode to flood?'

  - hyperdynamics (Voter)

  - chemical flooding (Grubmüller)

  - metadynamics (Parinello)

  - replica exchange

- free energy sampling techniques

  need reaction coordinate!

  - umbrella sampling

  - free energy perturbation

  - transition path sampling

How to study reactions and (rare) dynamical events

- direct MD

- reaction path methods   'no entropy'

    - **NEB (nudged elastic band, Jonsson)**

    - **CPR (conjugate peak refinement, Fischer, Karplus)**

    - **dimer method (Jonsson)**

- accelerated   node to flood?'

    - **hyperd**

    - **chemic**

    - **metady**   Very time consuming

    - **replica**

- free energy sampling techniques   need reaction coordinate!

    - **umbrella sampling**

    - **free energy perturbation**

    - **transition path sampling**

Gliederung

Physics, we have considered a **discrete** system, i.e. a quantum mechanical system with discrete energy levels. For that, we found:

$$Z = \sum_i exp(-\beta E_i)$$

and the (discrete) distribution function

$$p_i = \frac{1}{Z} exp(-\beta E_i)$$
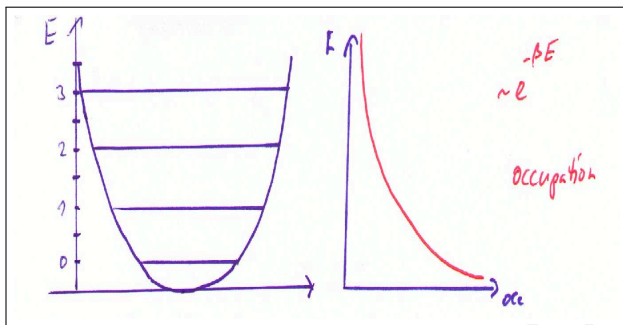
On the other hand, we disussed the dynamics of molecules, where temperature allows to sample a certain part of the protein conformations.

$$p(x, p) = \rho(x, p) = \frac{1}{Z} exp[-E(x, p)/kT]$$

To get the partition function Z, we no have to substitute the summation in the discrete case by the integral

$$Z = \int exp[-E(x, p)/kT]dxdp$$

Relation to the partition function

Consider the canonical distribution ($\beta^{-1} = kT$):

$$p_i = \frac{1}{Z} exp(-\beta E_i)$$

The expectation value of the energy is:

$$< E > = \frac{1}{Z} \sum_i E_i exp(-\beta E_i)$$

A nice mathematical trick is:

$$-\frac{\partial}{\partial \beta} Z = -\sum_i \frac{\partial}{\partial \beta} exp(-\beta E_i) = \sum_i E_i exp(-\beta E_i).$$

Relation to the partition function

Therefore,

$$< E >= -\frac{1}{Z}\frac{\partial}{\partial \beta}Z = -\frac{\partial lnZ}{\partial \beta} \tag{1}$$

To relate the free energy to Z, the easiest way is to use the thermodynamic relation:

$$F = U - TS.$$

Multiplying with β,

$$\beta F = \beta U - S/k$$

Taking the derivative

$$\frac{\partial(\beta F)}{\partial \beta} = U =< E >$$

and comparing with eq. 1 gives the expression for the free energy:

Relation to the partition function

$$F = -kT \ln Z \tag{2}$$

and for the entropy $S = -F/T - U/T$:

$$S = k \ln Z + k\beta < E > \tag{3}$$

Relation to the partition function

This is a remarkable result, since the only thing we have to do in our simulations it to get the partition function Z, i.e. we have to get the phase space distribution ρ, i.e. the density of points in phase space from the simulation and then integrate it over phase space to get Z. Everything follows from there.

Relation to the partition function

$$U = <E> = -\frac{\partial lnZ}{\partial \beta}$$

$$H = U + pV$$

$$S = klnZ + k\beta U$$

$$F = -kTlnZ$$

$$G = F + pV$$

Therefore, the computational problem is to determine Z.

Umbrella sampling

Umbrella sampling is the method of choice, if we want to have the change of free energy along a certain reaction coordinate q for the transition between state A and B. The reaction coordinate can be the distance between to atoms, the position of a proton when moving between donor and acceptor or something more complicated.

Umbrella sampling

Now, we make a coordinate transformation from $\overrightarrow{x} = (x_1 ... x_{3N})$ to a set $(u_1 ... u_{3N-1}, q)$: q is the degree of freedom represented by the reaction coordinate, and $\overrightarrow{u}$ the remainder, that we can write:

$$d\overrightarrow{x} = d\overrightarrow{u}\,dq$$

Umbrella sampling

When we look for the free energy for a certain value of q, we average over all remaining degrees of freedom, i.e. we perform and MD and sample all degrees of freedom except for q. An example would be the free energy for the formation of an ion pair in solution, as shown in Fig. 3. For every value of q, a MD is performed to calculate the free energy for that value of the reaction coordinate.
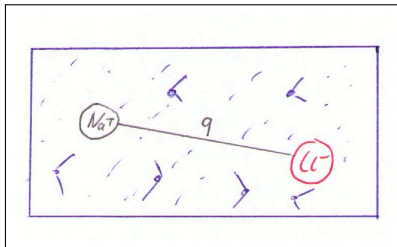
Abbildung: Na^+ and Cl^- in water solution: the distance between the ions is the reaction coordinate q, all other degrees of freedom (water), which are represented by u, are allowed to vary.

Umbrella sampling

The free energy is given by:

$$F = -kTln \int \int e^{-\beta E(\overrightarrow{x}, \overrightarrow{p})} dxdp$$

If we want to have one coordinate fixed at a certain value, e.g. q_0, it is convenient to use the delta function:

$$\delta(q - q_0).$$

With that, we can write the free energy for the fixed reaction coordinate:

Umbrella sampling

$$F(q_0) = -kTln \int \int \delta(q - q_0)e^{-\beta E(\overrightarrow{x}, \overrightarrow{p})} d\overrightarrow{p} \, d\overrightarrow{u} \, dq$$

$$= -kTln \left(Z \int \int \delta(q - q_0)\frac{e^{-\beta E(\overrightarrow{x}, \overrightarrow{p})}}{Z} d\overrightarrow{p} \, d\overrightarrow{u} \, dq \right)$$

$$= -kTln \left(Z \int \int \delta(q - q_0)\rho(\overrightarrow{x}, \overrightarrow{p})d\overrightarrow{p} \, d\overrightarrow{u} \, dq \right)$$

$$= -kTln \left(Z < \delta(q - q_0) > \right)$$

$$= -kTlnZ - kTln < \delta(q - q_0) >$$

Umbrella sampling

How can we interpret this? $\rho(\overrightarrow{x}, \overrightarrow{p})$ is the probability, that the system is at the point $(\overrightarrow{x}, \overrightarrow{p})$. Then,

$$P(q_0) = \int \int \delta(q - q_0)\rho(\overrightarrow{x}, \overrightarrow{p})d` =< \delta(q - q_0) >$$

is the probability $P(q_0)$, that the system is at the value of the reaction coordinate q_0, since the integral 'sums' over the whole phase space and the delta function 'cancels out' all points, where the reaction coordinate is NOT at q_0! I.e., the integration collects all points in phase space, where the reaction coordinate has this specific value!

Umbrella sampling

In the example of the ion pair: we perform an MD for the system, and then only count, how often the reaction coordinate is found at a specific value, i.e. calculate the probability $P(q_0)$ to find the system at q_0.

Therefore, the difference of free energy for two states A and B is:

Umbrella sampling

$$F_B - F_A = -kTlnZ - kTln < \delta(q - q_B) > +kTlnZ + kTln < \delta(q - q_A)) >$$

$$= -kTln\frac{< \delta(q - q_B) >}{< \delta(q - q_A) >}$$

$$= -kTln\frac{P(q_B)}{P(q_A)}$$

Umbrella sampling

So, the task is clear: perform a MD, specify a coordinate, and then just count, how often the system is at special values of the reaction coordinate: the difference of these numbers gives the free energy difference!
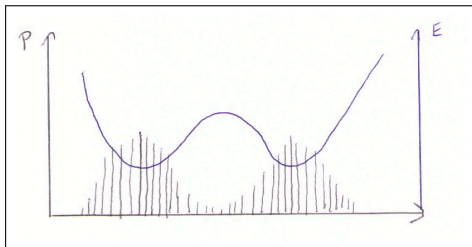
Abbildung: Schematic energy profile along a reaction coordinate and the probability distribution. The barrier region is sampled poorly.

Umbrella sampling

This is nice, but we also know the problem: If we have to cross a high barrier along the reaction coordinate to come from A to B, a pure MD will never do it. Therefore, we have to **drive** the system 'somehow': This can be done, by applying an additional potential V!

Lets try something:

Umbrella sampling

$$F(q_0) = -kTln\left(\frac{\int\int \delta(q - q_0)e^{-\beta E}d`}{\int\int e^{-\beta E}d`}\right)$$

$$= -kTln\left(\frac{\int\int \delta(q - q_0)e^{\beta V}e^{-\beta(E+V)}d`}{\int\int e^{-\beta(E+V)}d`}\frac{\int\int e^{-\beta(E+V)}d`}{\int\int e^{-\beta(E)}d`}\right)$$

$$= -kTln\left(< \delta(q - q_0)e^{\beta V} >_{E+V}\frac{\int\int e^{-\beta(E+V)}d`}{\int\int e^{\beta(V)}e^{-\beta(E+V)}d`}\right)$$

$$= -kTln\left(< \delta(q - q_0)e^{\beta V} >_{E+V}\frac{1}{< e^{\beta V} >_{E+V}}\right)$$

$$= -kTln\left(e^{\beta V(q_0)}< \delta(q - q_0) >_{E+V}\frac{1}{< e^{\beta V} >_{E+V}}\right)$$

$$= -kTln< \delta(q - q_0) >_{E+V} - V(q_0) + kTln< e^{\beta V} >_{E+V}$$

$$= -kTlnP^*(q_0) - V(q_0) + kTln< e^{\beta V} >_{E+V}$$

Umbrella sampling

The last equation has the form:

$$F(q) = -kTln P^*(q) - V(q) + K \qquad (4)$$

This result is very interesting: we have added an arbitrary potential V to our system: Now we have to calculate the averages using the ensembles with the energy E+V, that is indicated by $<>_{E+V}$. Now, we get the probability $P^*(\vec{q_0})$ to find the system at the value of the reaction coordinate for the ensemble E+V, which can be very different from that of the ensemble E, $P(\vec{q_0})$! But we still get the right free energy $F(\vec{q_0})$, if we use the probability $P^*(\vec{q_0})$ and subtract the potential $V(\vec{q_0})$ at the value of the reaction coordinate and the red term K.

Umbrella sampling

We can use this scheme efficiently, when we move harmonic potentials along the reaction coordinate as shown in Fig. 5.
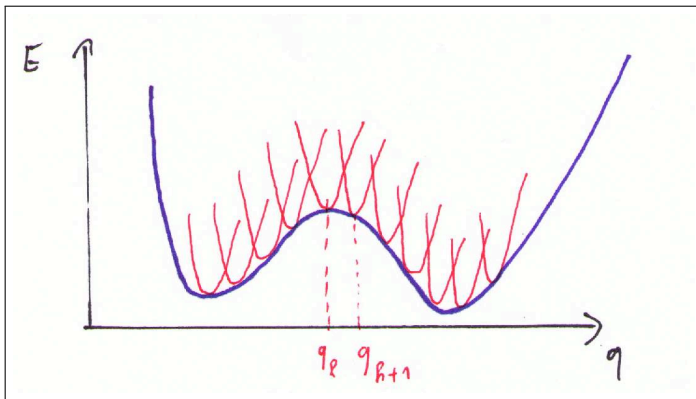
Abbildung: Additional harmonic potentials to keep the system in the region of the desired value of the reaction coordinate
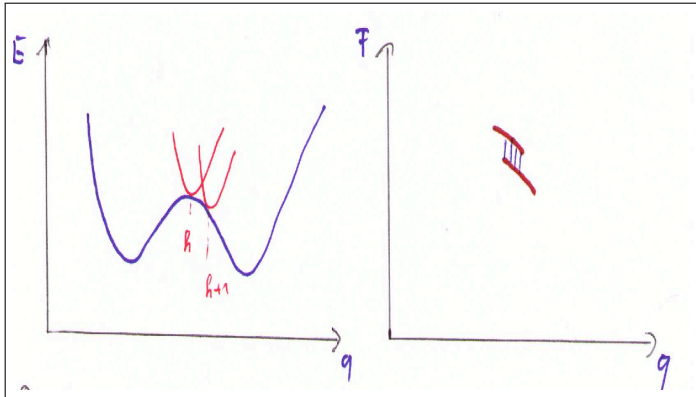
Umbrella sampling

Abbildung: Offset in free energy between two simulations k and (k+1). The offset is given by K_k-K_{k+1}

Umbrella sampling
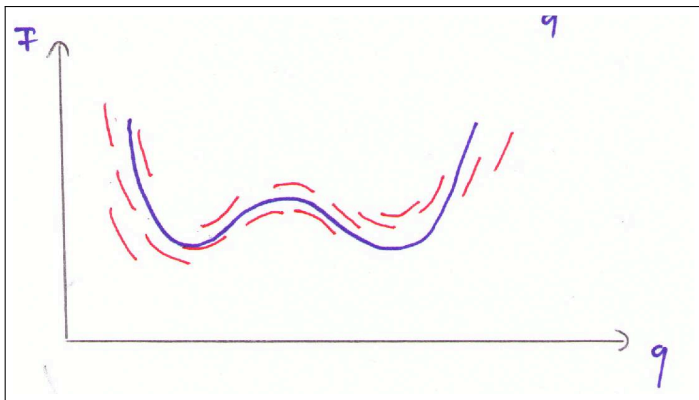
Abbildung: Matching of histograms from different simulations

'Multi-scale' methods
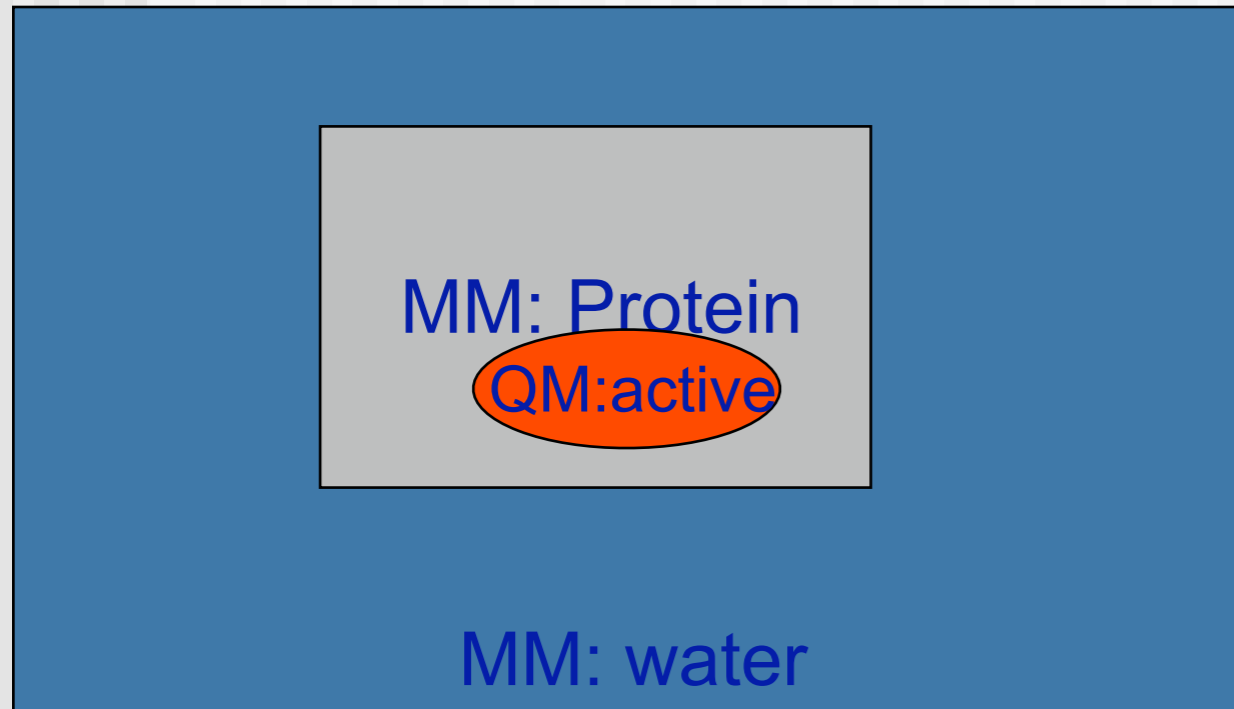
- periodic boundary conditions

- Ewald summation

'Multi-scale' methods

MM: Protein

QM:active

MM: water
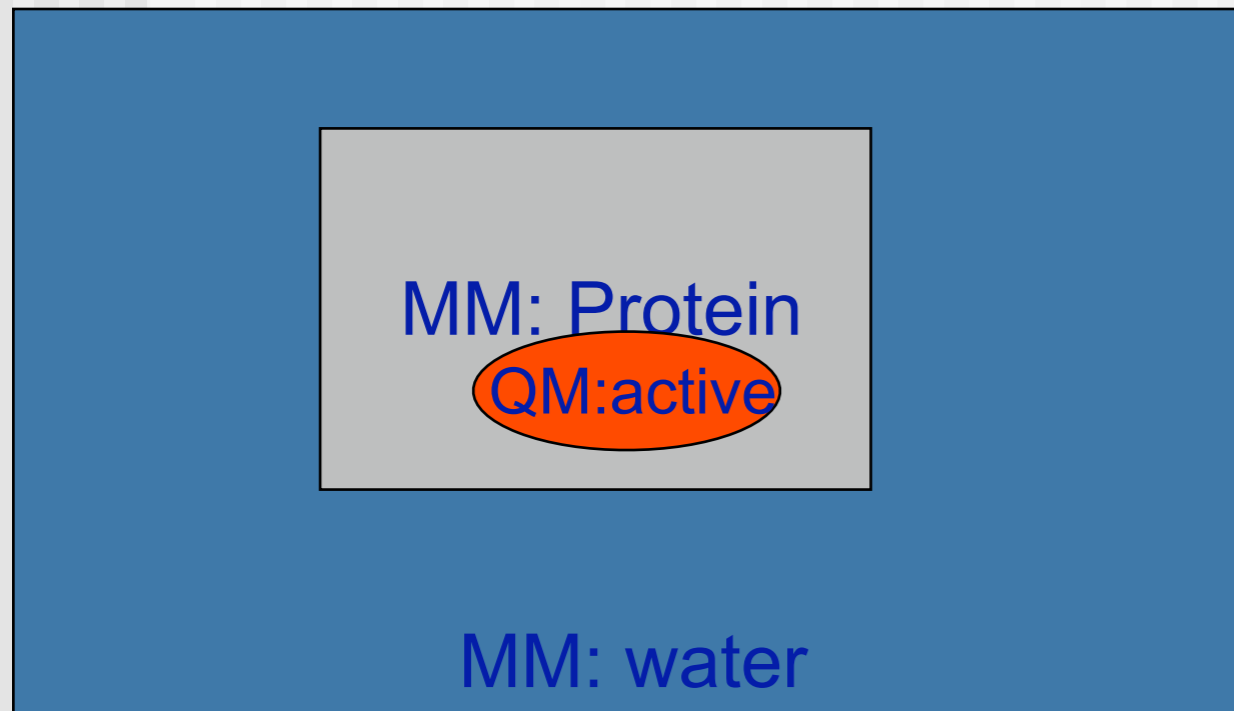
- periodic boundary conditions

- Ewald summation

Problems:

- large proteins

- setup (membrane) tedious

- solvent relaxation time

Multi-scale models in theoretical biophysics

time

Continuum Electrostatics "Coarse graining"

Force Fields:MM

Approximative Methods

HF, DFT

CI, MP
CASPT2

Chromophore
Lipids
Top view
Protein
Water Molecules
Side view

nm

Length scale

10   100   1000   10.000   atoms
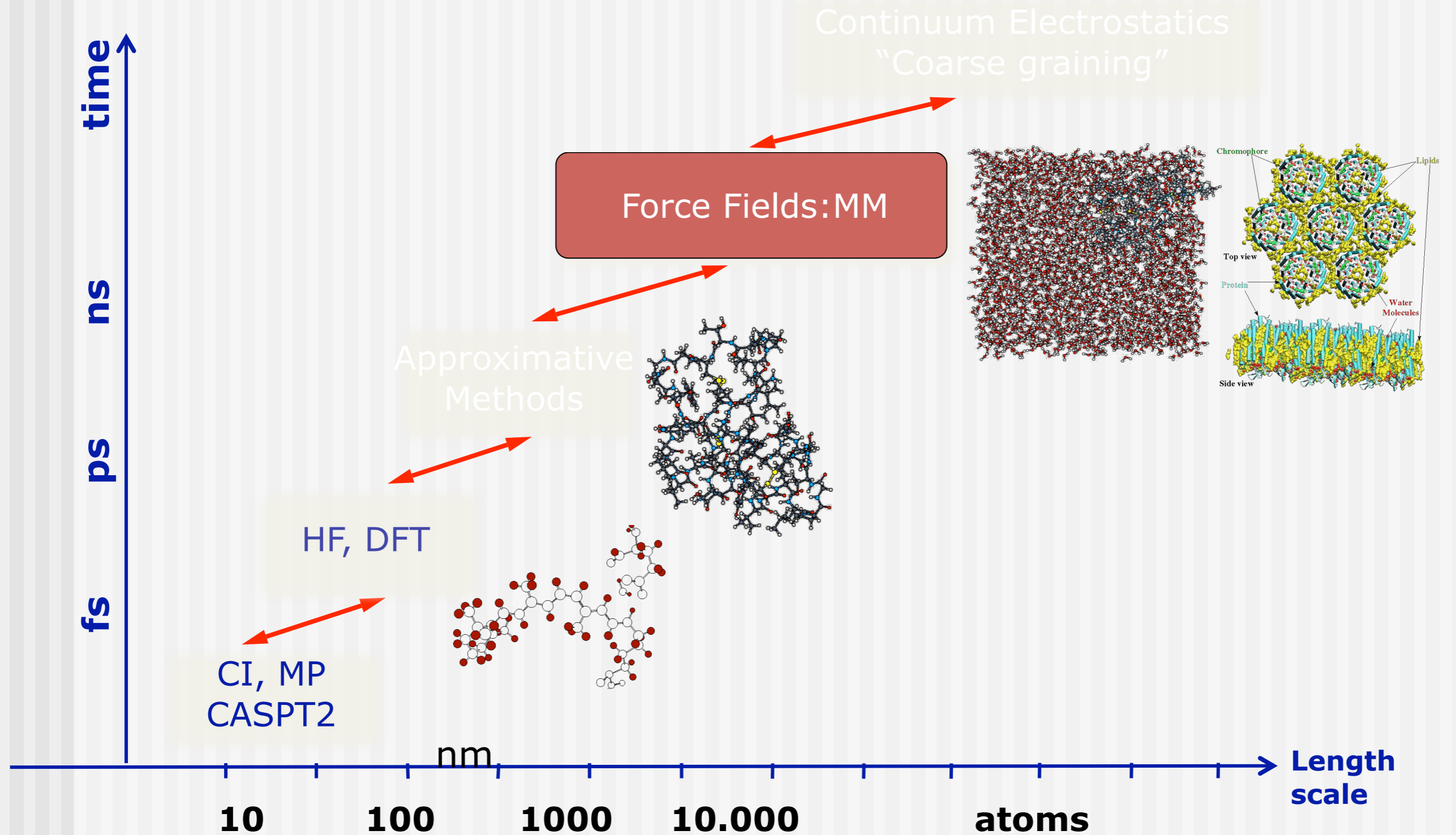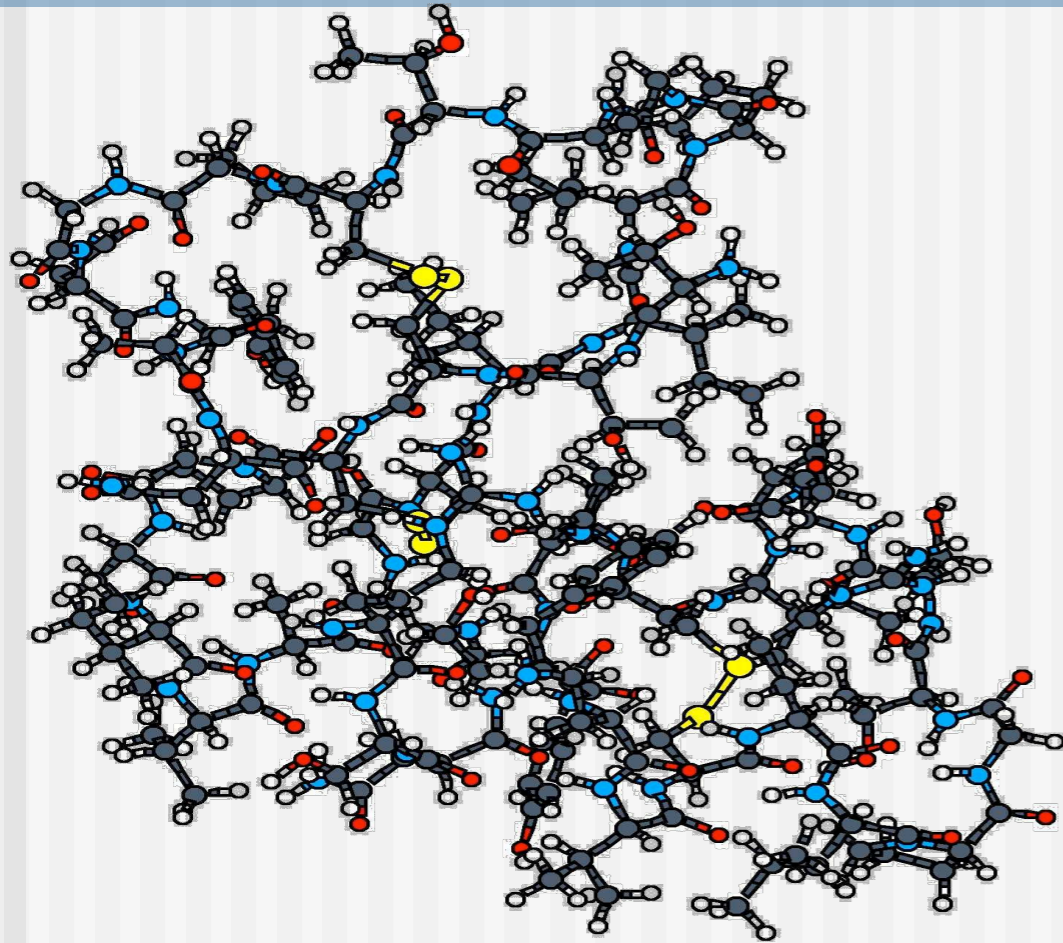
Coarse Graining (CG)
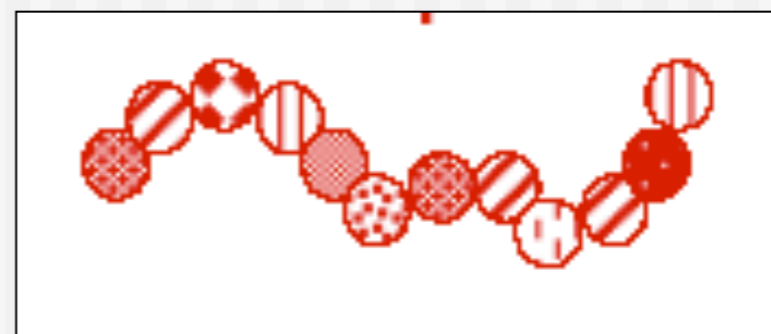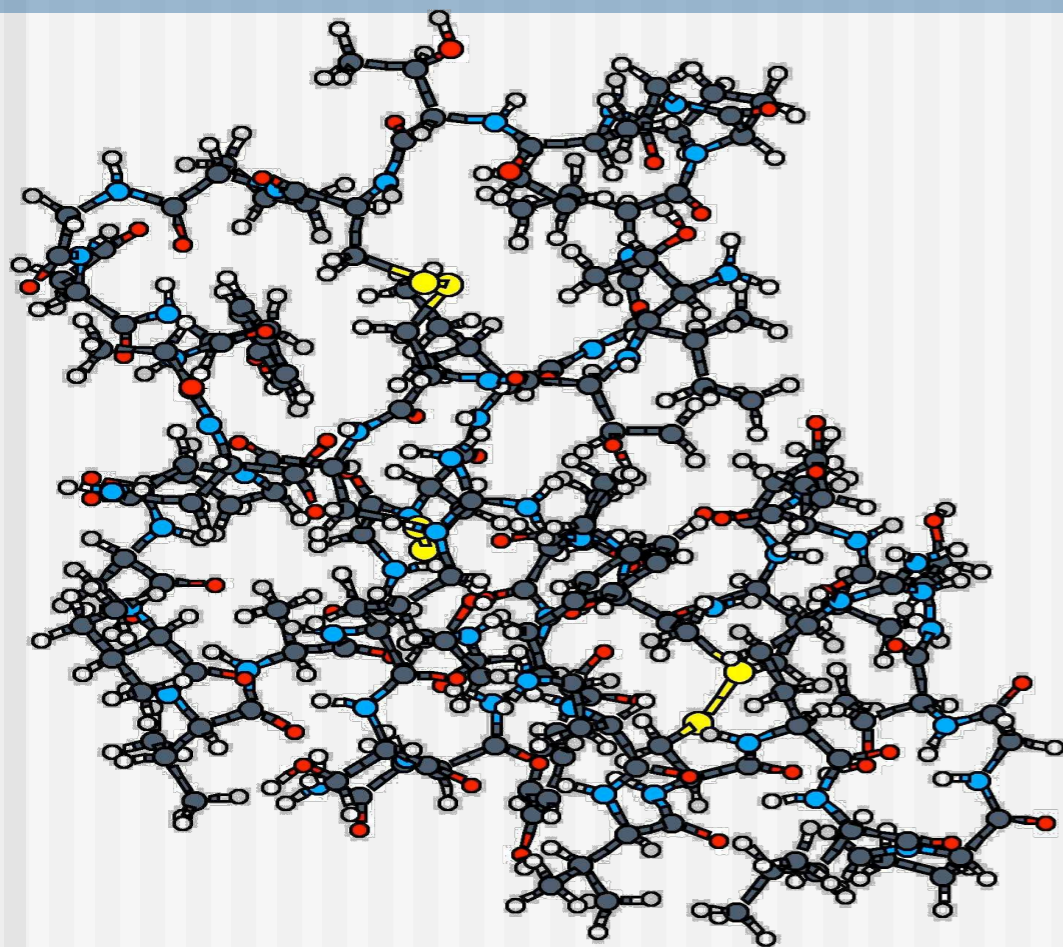
Coarse Graining (CG)
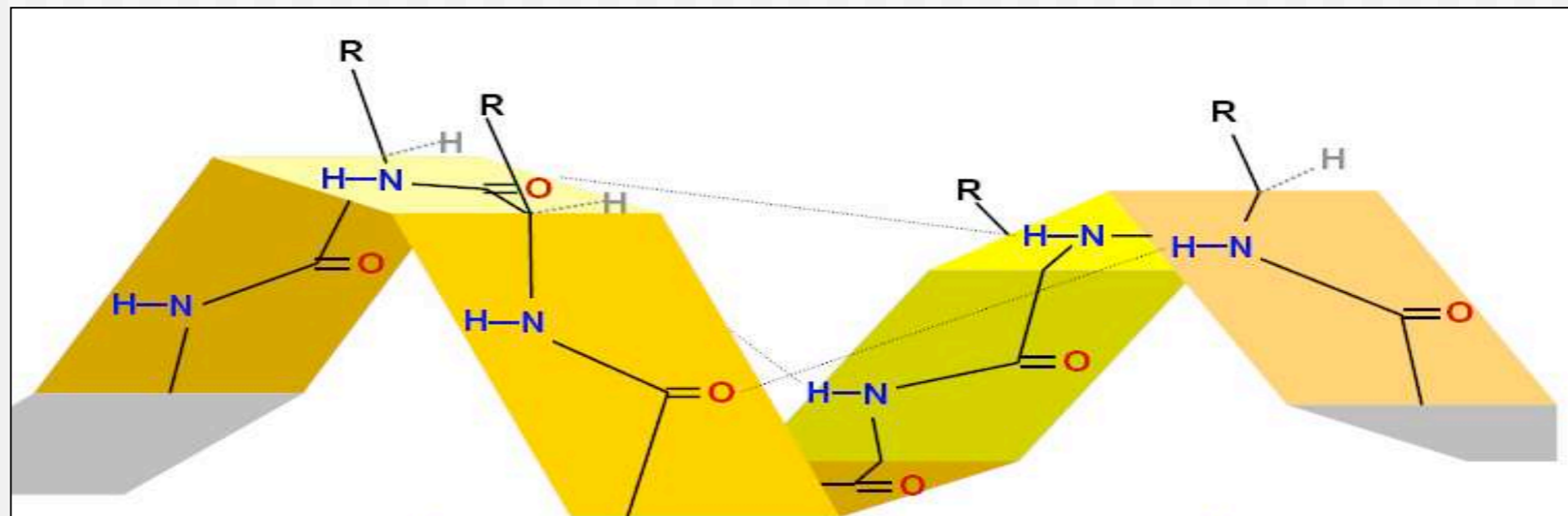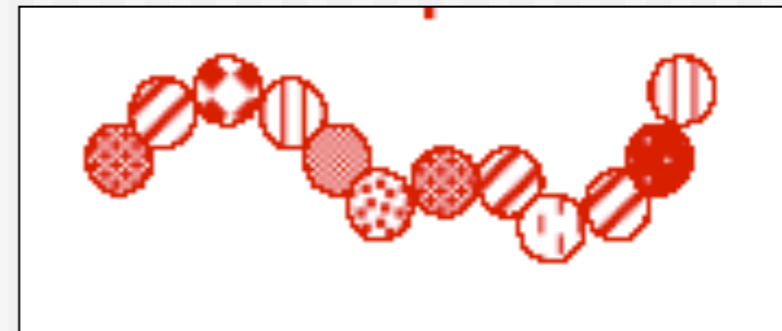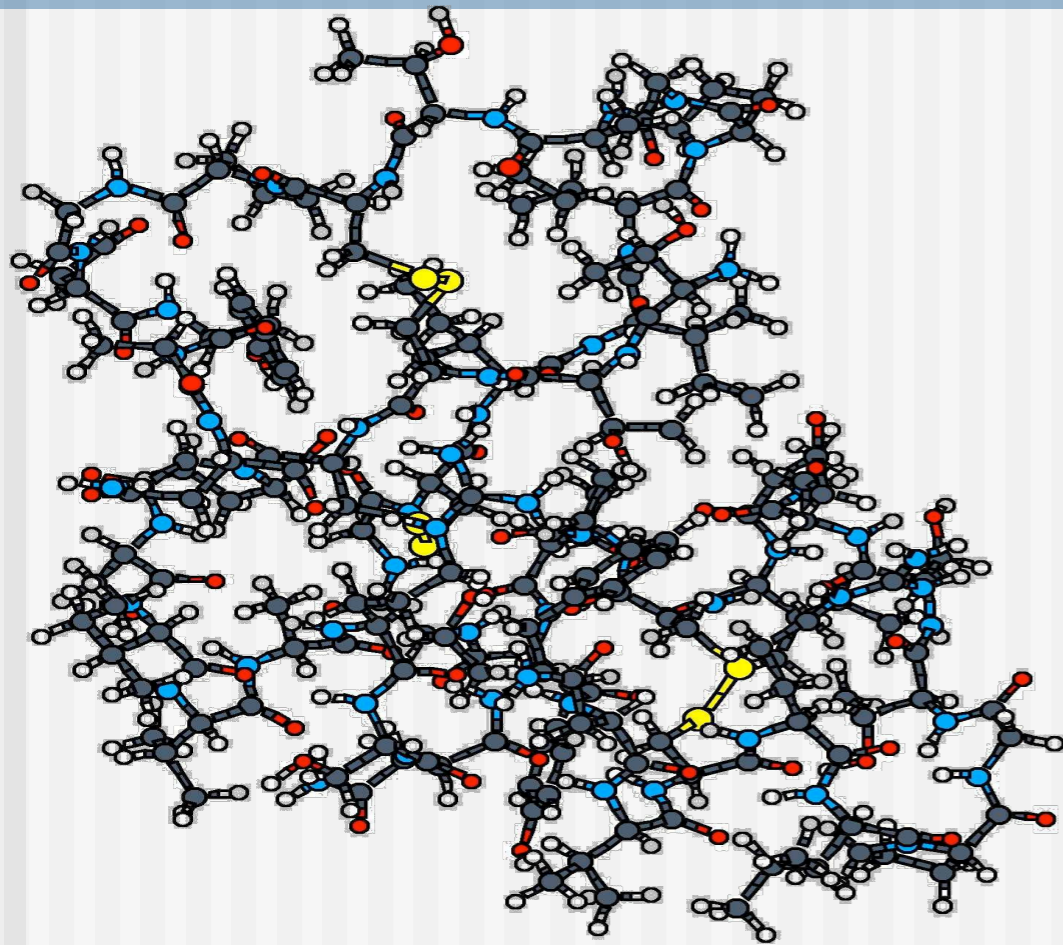
Coarse Graining (CG)
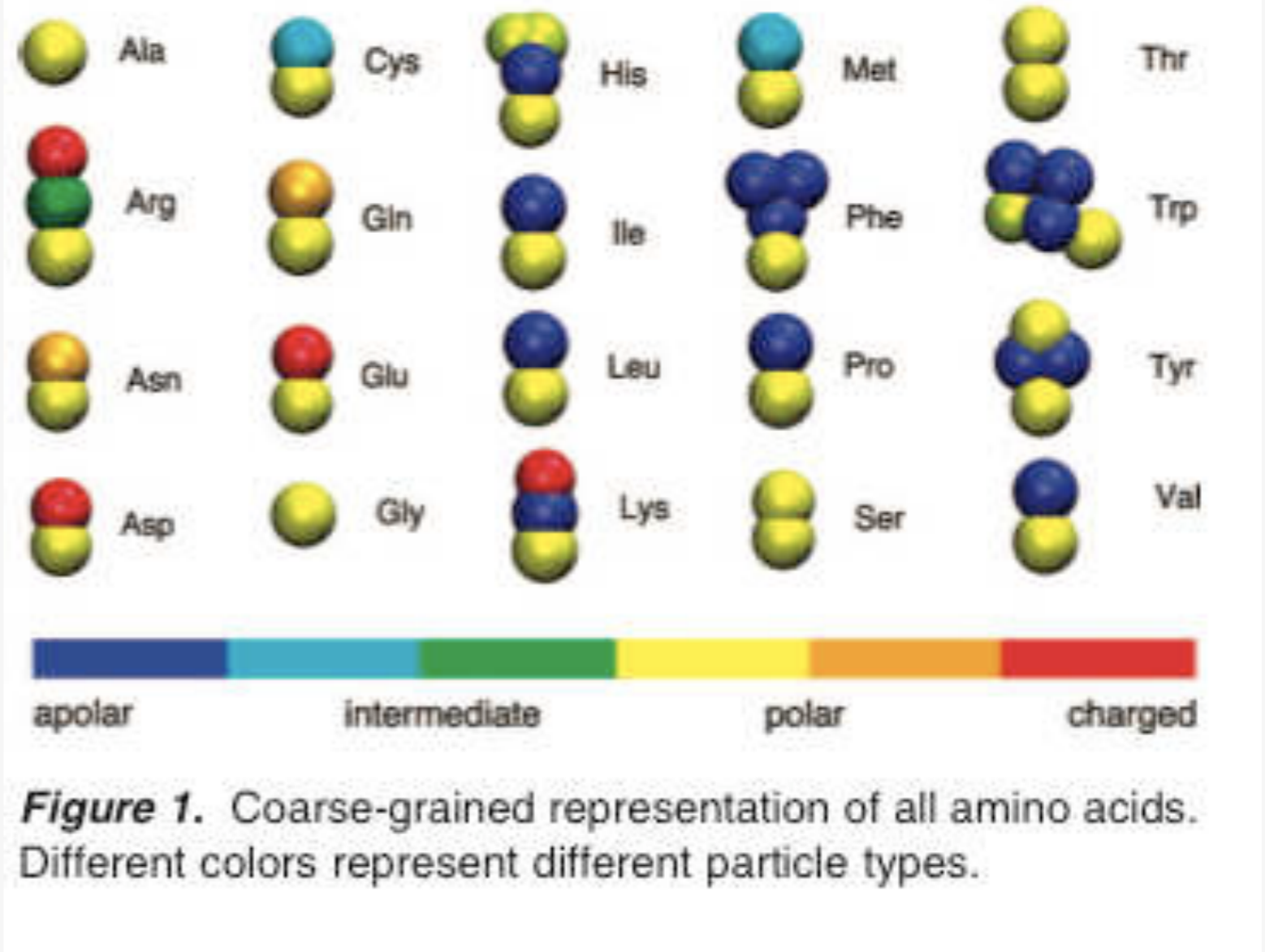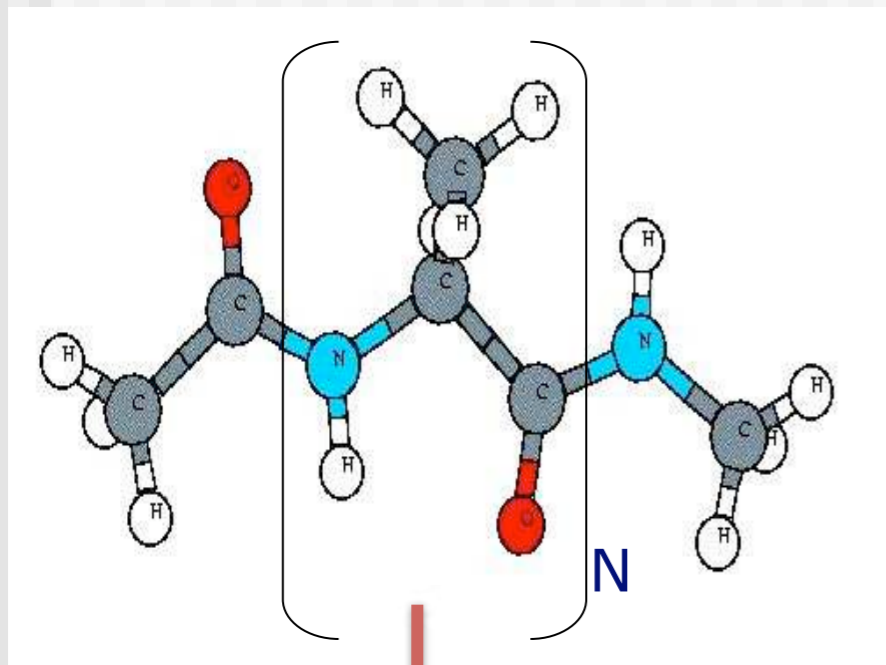
Coarse Graining (CG)

Figure 1. Coarse-grained representation of all amino acids. Different colors represent different particle types.

Martini force field, Monticelli et al. JCTC 4 (2008) 819

Multi-scale models in theoretical biophysics
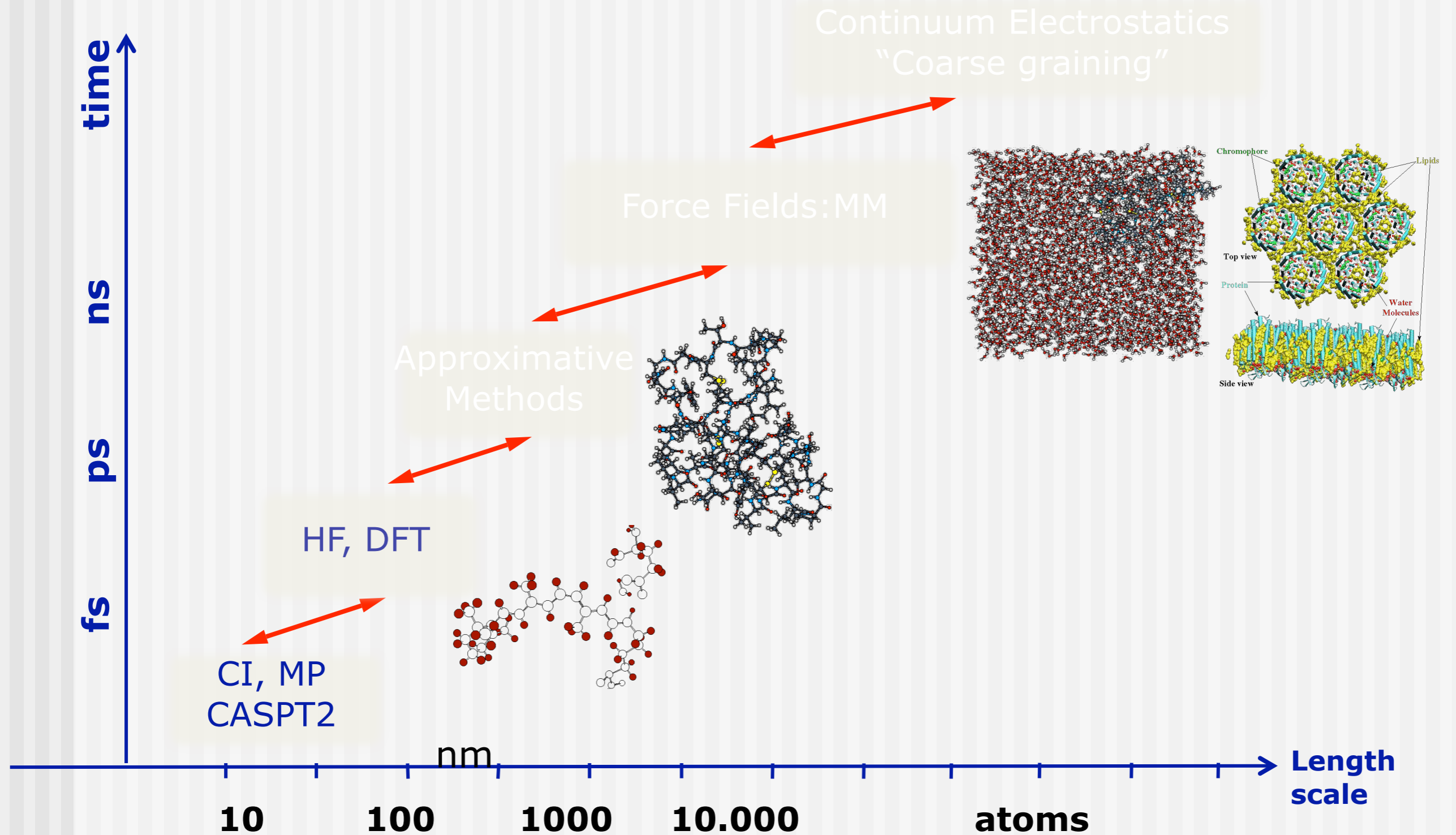
Continuum electrostatics

Describe the solvent around a molecule by an effective dielectric

- quantum chemistry: continuum solvation models
    solvent changes molecules properties

- biophysics: get electrostatic properies of proteins/DNA

Continuum electrostatics
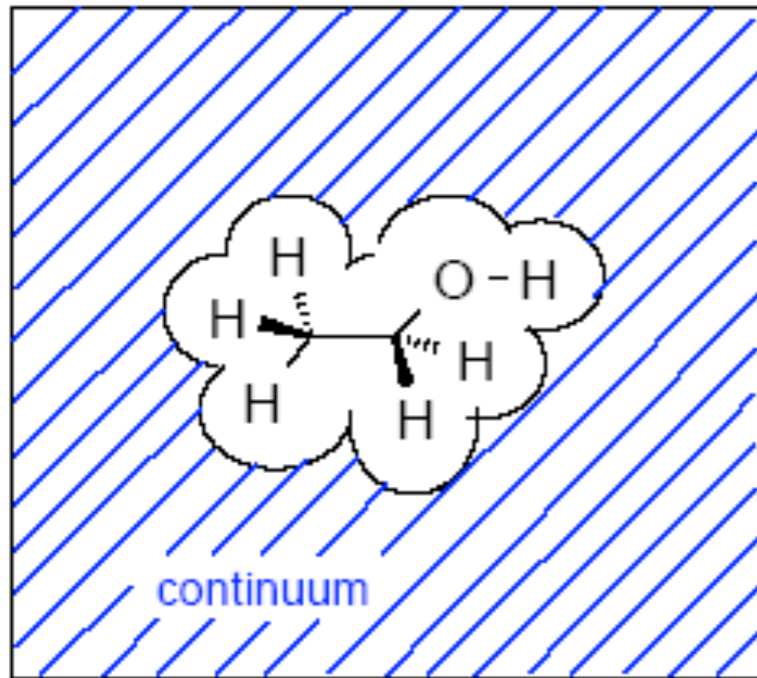
continuum

The solvation energy contains three terms:

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{VdW} + \Delta G_{ele}$$

This contribution changes all properties, like energies, forces (i.e. geometry), vib. frequencies …

Continuum electrostatics

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{VdW} + \Delta G_{ele}$$
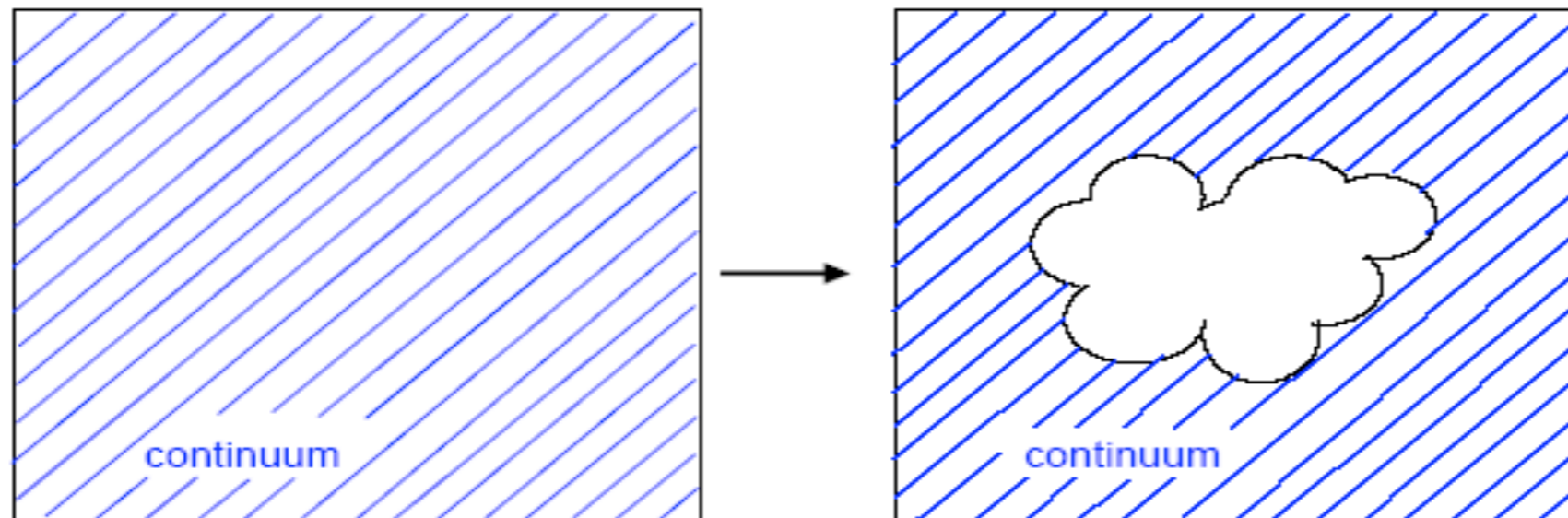
formation of cavity and VdW interactions of solute with solvent

$$\Delta G_{cav} + \Delta G_{VdW} = \sum_{i} c_i S_i.$$

modeled as sum over surface areas S_i of atoms

Continuum electrostatics
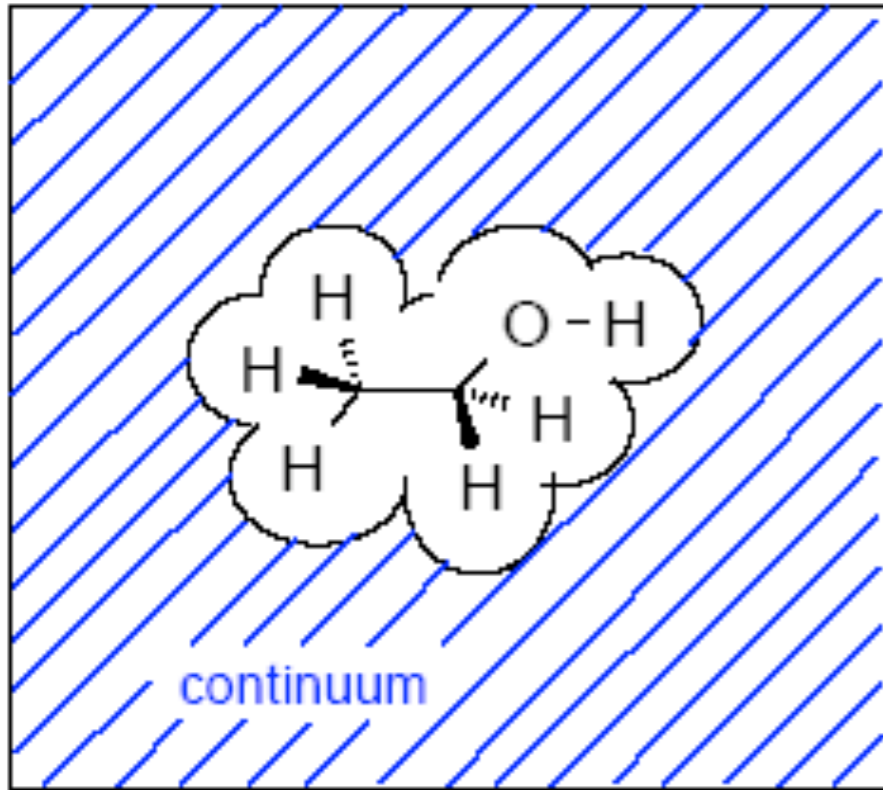
continuum

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{VdW} + \Delta G_{ele}$$

$$E_{ele} = q\phi(\vec{r})$$

Interaction of molecular charges with electrostatic potential Φ(r)

$$\Delta E_{ele} = q\phi_{solv}(\vec{r}) - q\phi_{vac}(\vec{r})$$

$$\phi_{rf}(\vec{r}) = \phi_{solv}(\vec{r}) - \phi_{vac}(\vec{r})$$

$$\Delta E_{ele} = q\phi_{rf}(\vec{r})$$

Continuum electrostatics
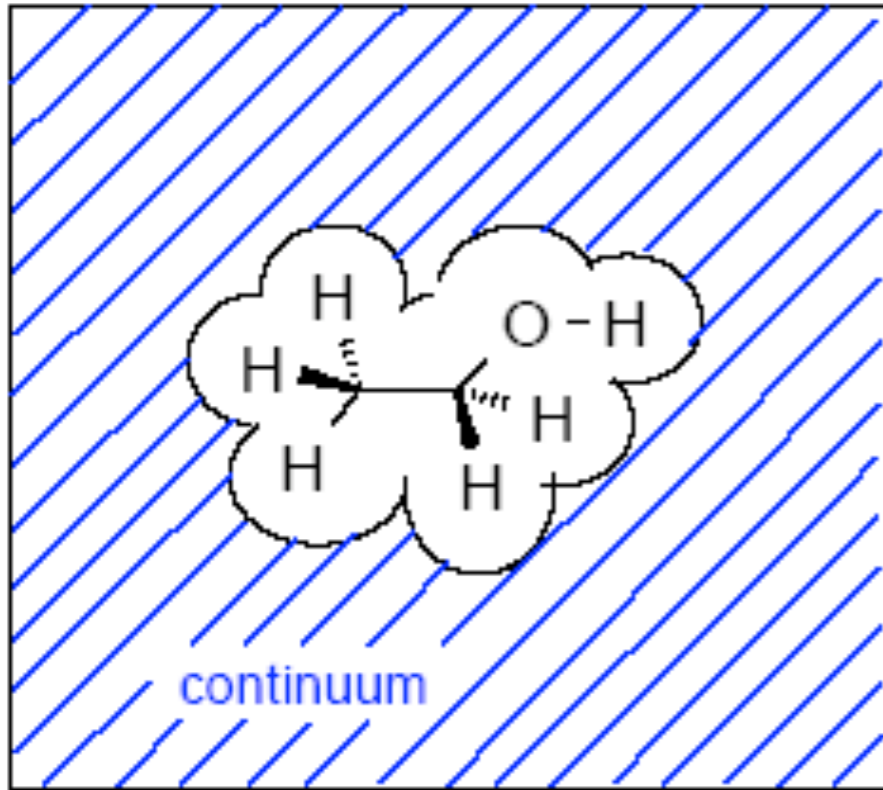
$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{VdW} + \Delta G_{ele}$$

Energy of charge q:

$$\Delta E_{ele} = q\phi_{rf}(\vec{r})$$

Move to free energy surface by 'sampling' over possible solvent contributions (entropy) and by including the pV term.

Continuum electrostatics
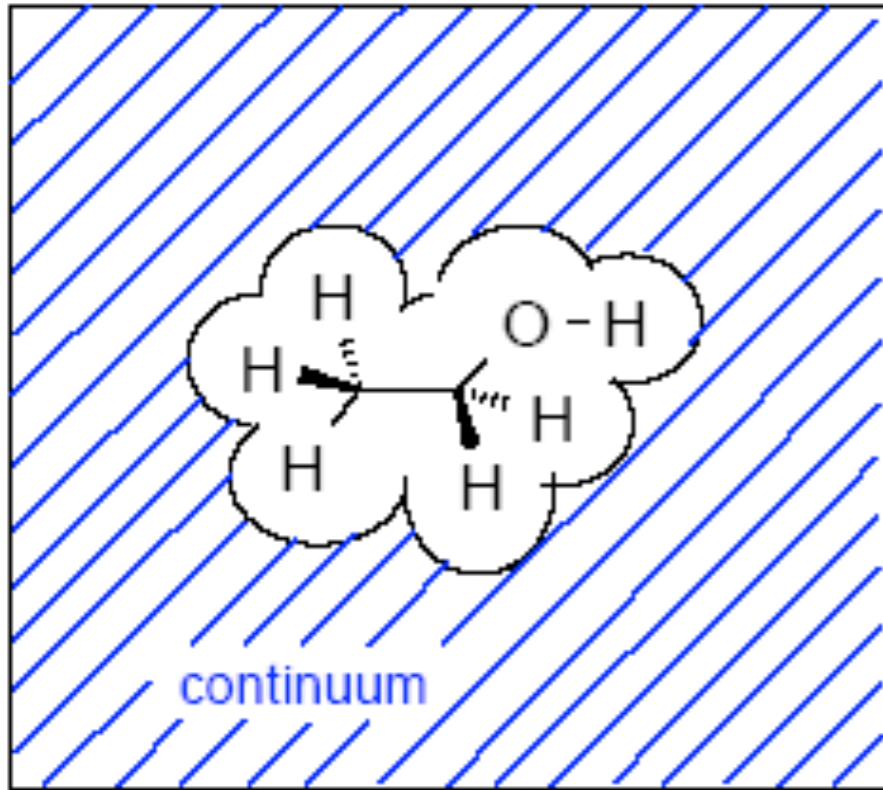
continuum

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{VdW} + \Delta G_{ele}$$

Energy of charge q:

$$\Delta E_{ele} = q\phi_{rf}(\vec{r})$$

Move to free energy surface by 'sampling' over possible solvent contributions (entropy) and by including the pV term.

This can be effectively done during parametrization of the model

Free enthalpy:

$$\Delta G_{ele} = q\phi_{rf}(\vec{r})$$
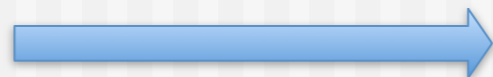
Calculate reaction field

Continuum electrostatics

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{VdW} + \Delta G_{ele}$$

Achievement:
we do not have to sample over all solvent degrees of freedom!

Free enthalpy:

$$\Delta G_{ele} = q\phi_{rf}(\vec{r})$$

Calculate reaction field

Continuum electrostatics: Generalized Born
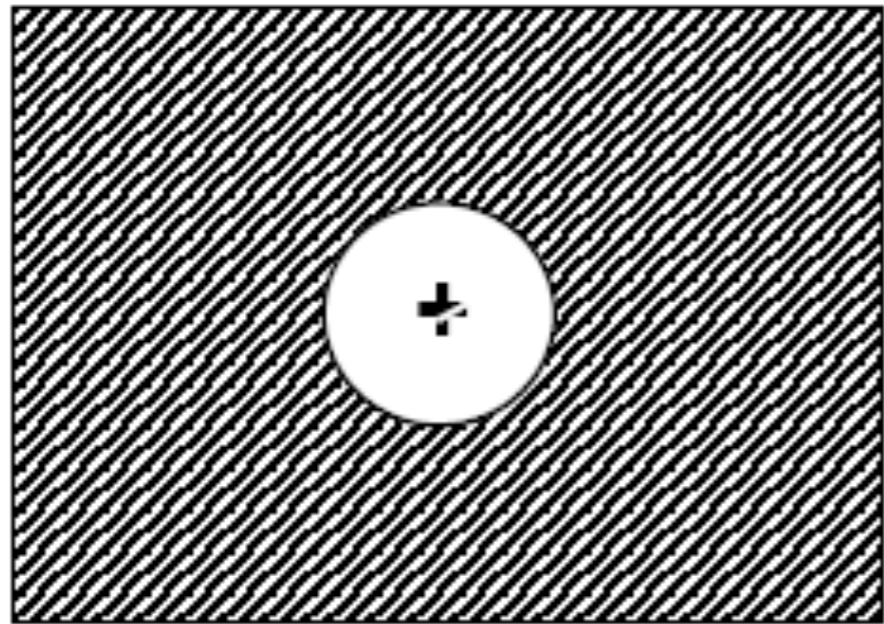
Born model (1920)

$$\Delta G_{ele} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon}\right)$$

Continuum electrostatics: Generalized Born

Born model (1920)

$$\Delta G_{ele} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon}\right)$$

Feig, Brooks, Current Opinion in Structural Biology 2004, 14:217–224

Continuum electrostatics: Generalized Born

Born model (1920)

$$\Delta G_{ele} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon}\right)$$

Generalized Born:

find empirical formular, which determines the Born radius of atom i inside the protein

ext

Feig, Brooks, Current Opinion in Structural Biology 2004, 14:217–224

Continuum electrostatics: Generalized Born

Born model (1920)

$$\Delta G_{ele} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon}\right)$$

Generalized Born:

find empirical formular, which determines the Born radius of atom i inside the protein

$$\Delta G^1_{ele} = -\left(1 - \frac{1}{\epsilon}\right)\sum_i \frac{q_i^2}{2a_i}$$

Feig, Brooks, Current Opinion in Structural Biology 2004, 14:217–224
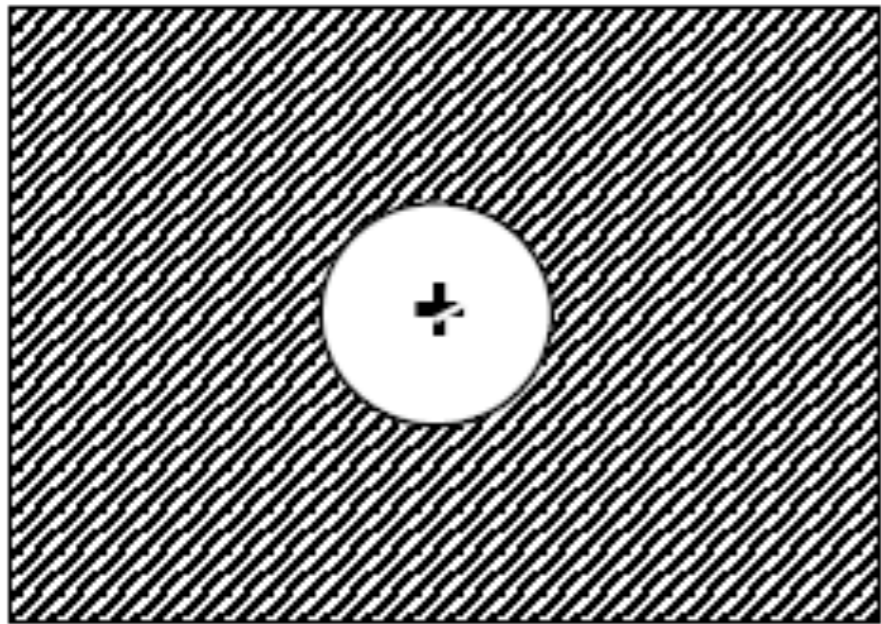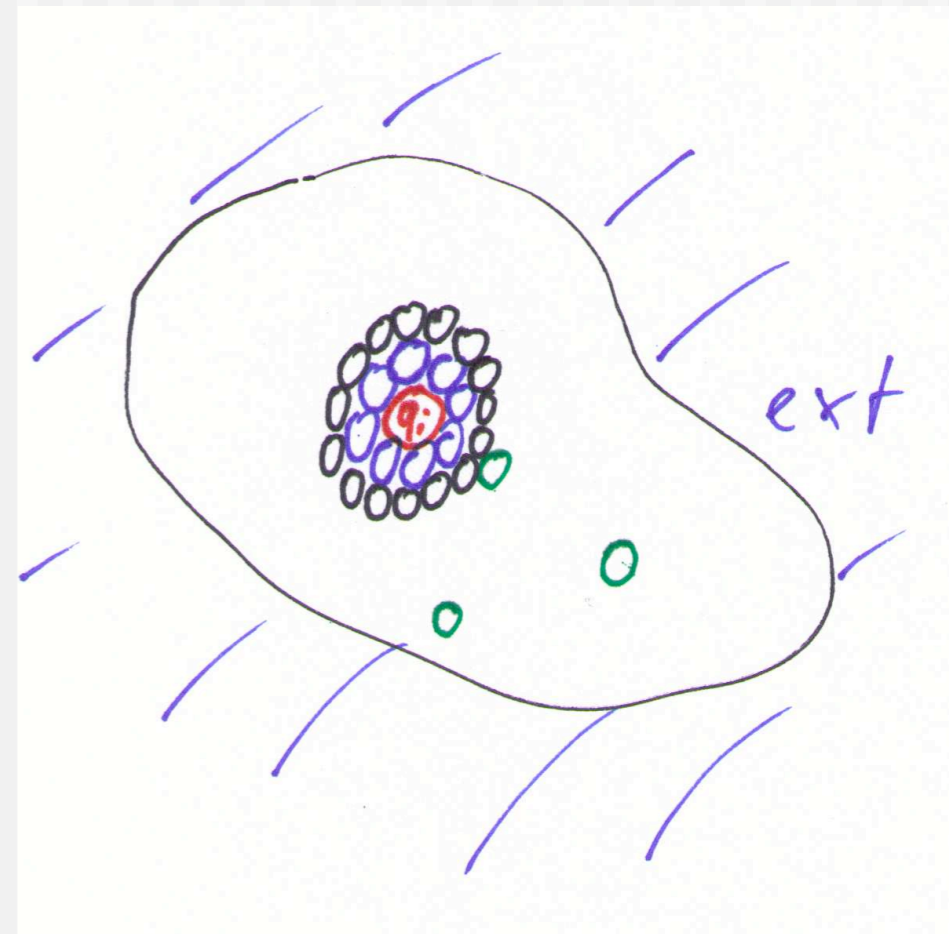
Continuum electrostatics: Generalized Born

Born model (1920)

$$\Delta G_{ele} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon}\right)$$

Generalized Born:

find empirical formular, which determines the Born radius of atom i inside the protein

$$\Delta G^1_{ele} = -\left(1 - \frac{1}{\epsilon}\right)\sum_i \frac{q_i^2}{2a_i}$$ + interaction $q_i q_j$

ext

Feig, Brooks, Current Opinion in Structural Biology 2004, 14:217–224
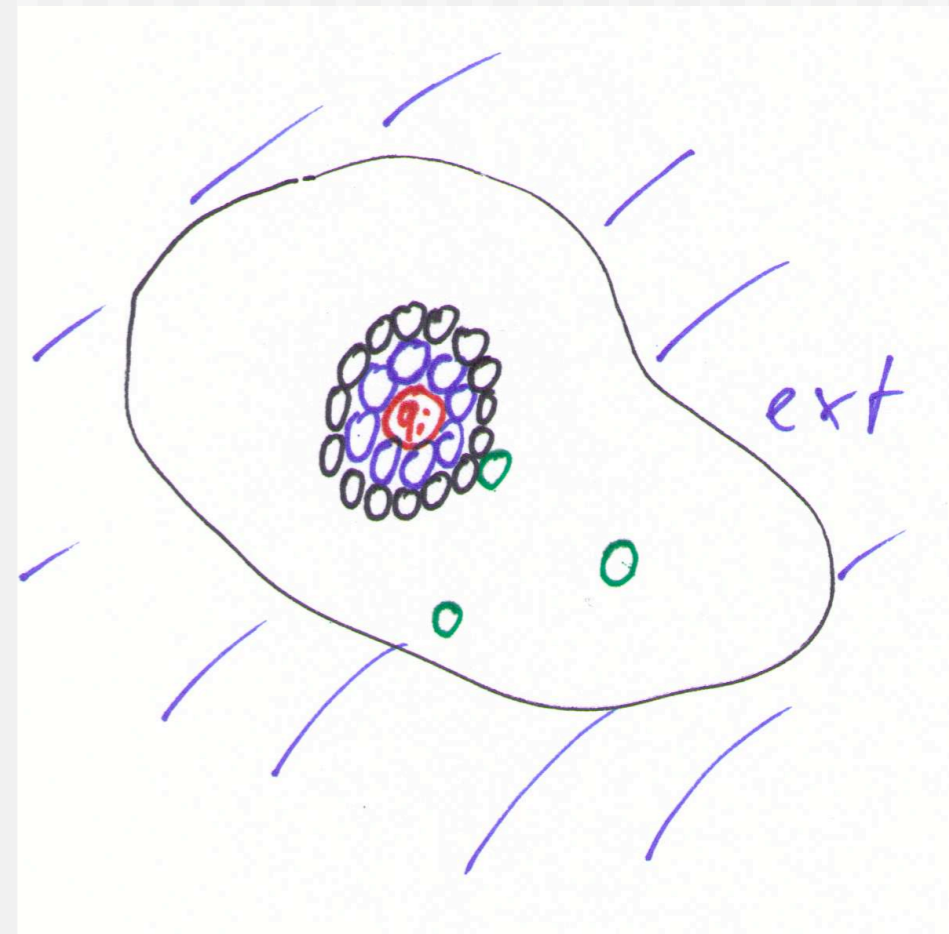
Continuum electrostatics: Poisson-Boltzmann

Poisson Boltzmann: solve Poisson eq. e.g. on a grid …

$$\nabla \epsilon \nabla \phi = -4\pi\rho$$
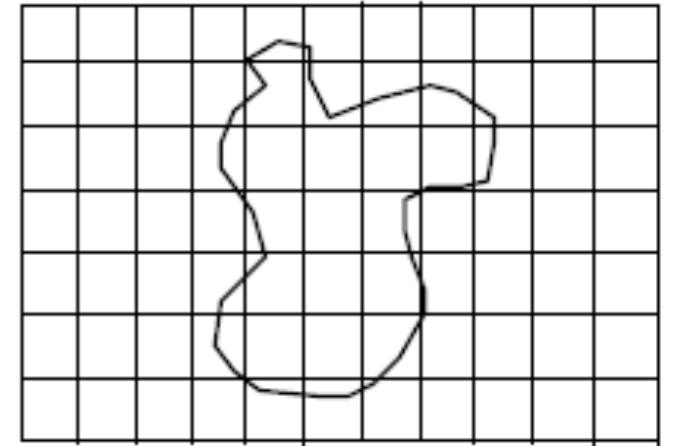
… with and without continuum around the molecule: get reaction field

$$\phi_{rf}(\vec{r}) = \phi_{solv}(\vec{r}) - \phi_{vac}(\vec{r})$$

Then put ions with (bulk) concentration n_i^0 and charge q_i according to Boltzmann distribution:

$$n_i(r) = n_i^0 e^{-q_i\phi(r)/kT}$$

Note: the PMF **W(r)** has been approximated by $\quad q_i\phi(r)$

Continuum electrostatics: Poisson-Boltzmann

For a 1:1 electrolyte, this leads to:

$$\nabla\epsilon\nabla\phi = -4\pi\rho + \sum_i q_i n_i^0 sinh[q_i\phi(r)/kT],$$

Can be linearized when $q_i\phi(r)$ small compared to kT

compute reaction field
from linearized PB equation

$$\nabla\cdot[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r})\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}).$$

Solvation free energy for set of
point charges q_i:

$$\Delta G_{elec} = \frac{1}{2}\sum_i q_i\phi_{rf}$$

Baker, Current Opinion in Structural Biology 2005, 15:137–143

Poisson Boltzmann (PB) vs. Generalized Born (GB)
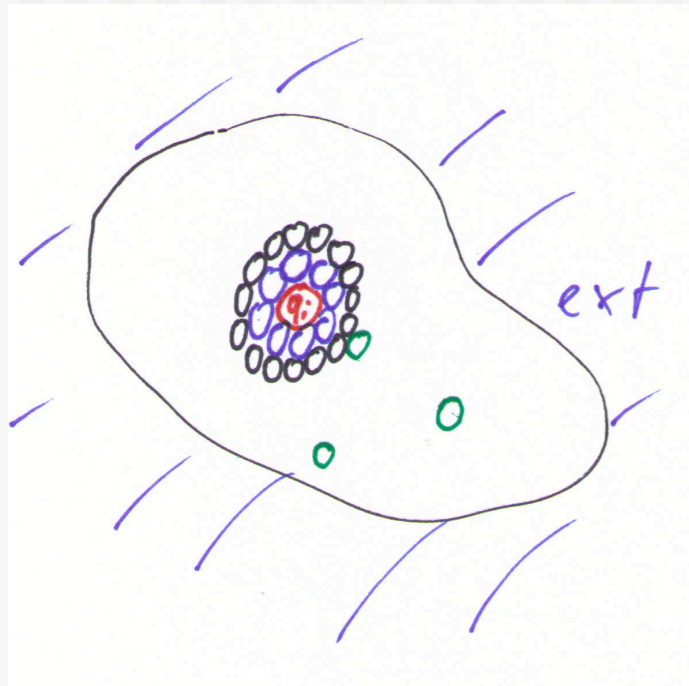
GB very cheap

- could do folding in principle
- however less accurate
- PB used to calibrate GB

First approximations:

- solvation→ charge scaling
- freezing vs. stochastic boundary
- size of movable MM?
- size of QM?
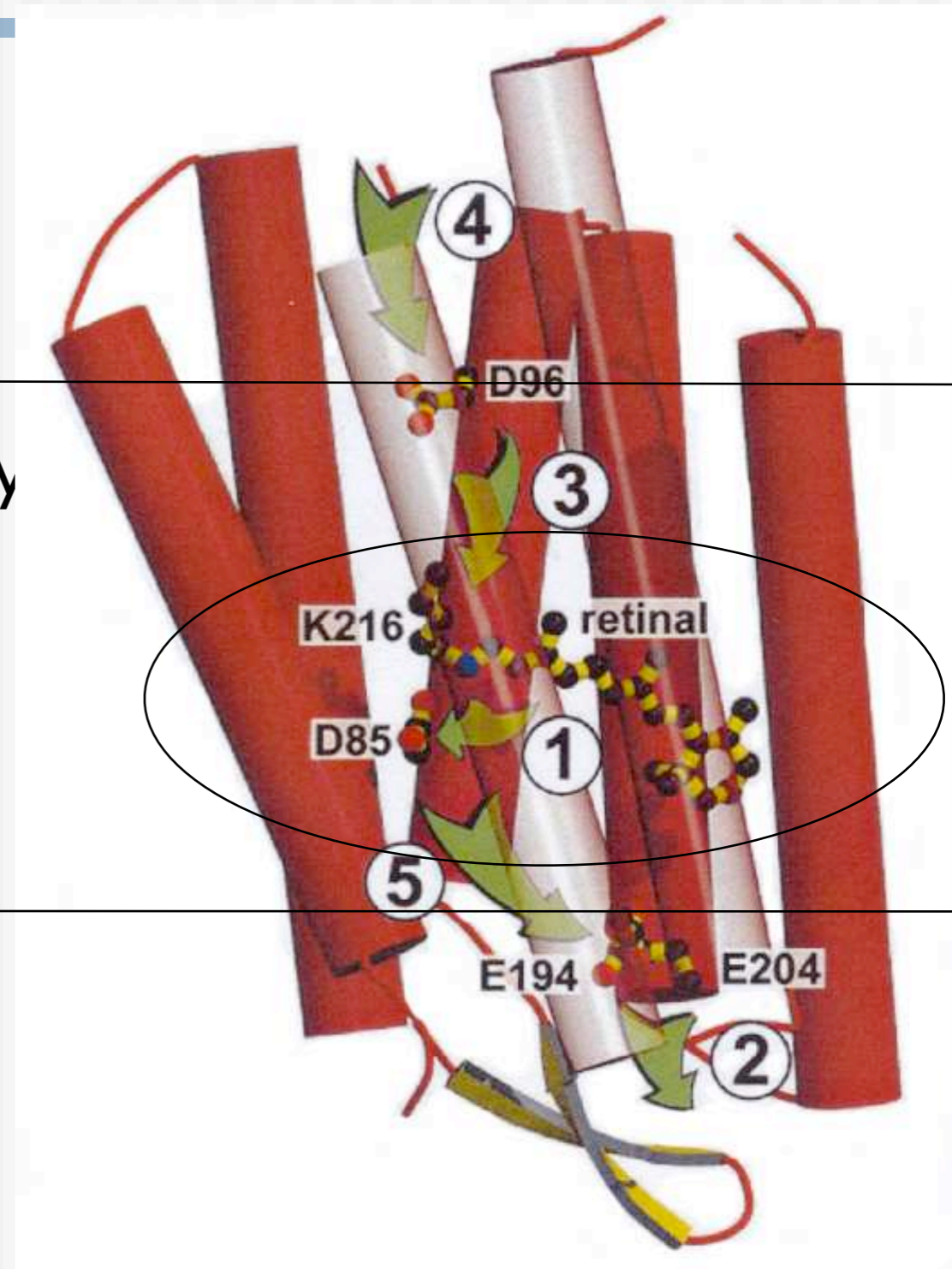
Charge scaling: T. Chem. Acc. 109 (2003) 118
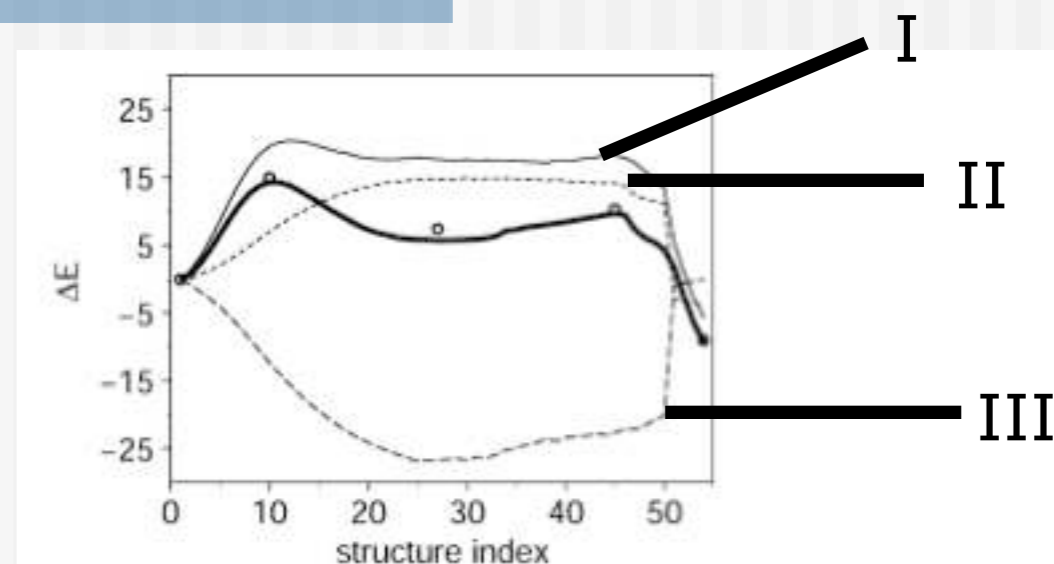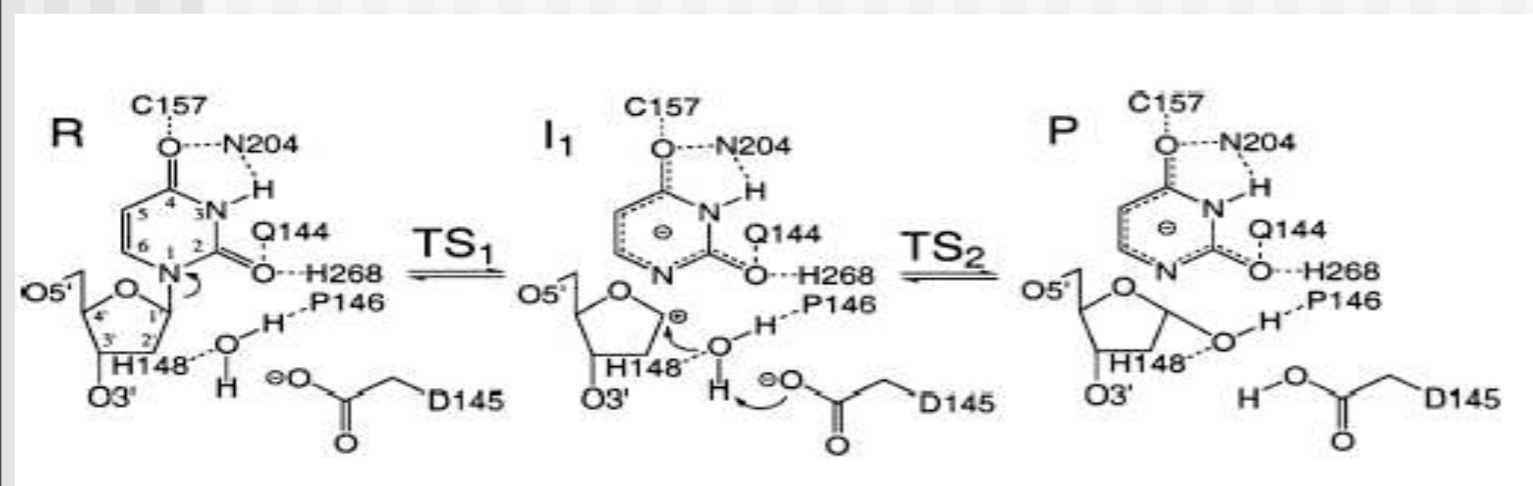
$$q_i' = \frac{\bar{\bar{\Phi}}_s(i \longrightarrow QM)}{\bar{\bar{\Phi}}_v(i \longrightarrow QM)} q_i = \frac{q_i}{\alpha_i}.$$

compute ESP of residue i at the QM region using PB with (**s**) and without (**v**) solvent

Fig. 3. Histogram of the scale factors for uracil-DNA glycosylase

Fig. 4. Reaction profile obtained with Mulliken charges representing the QM atoms. Total (*bold solid line*), step I (*thin solid line*), step IIa (*long dashed line*), and step III (*short dashed line*) energies. The total energies obtained for the extrema by using method IIb instead of IIa (Sect. II) are indicated by *circles*

energies are wrong, since compute with 'wrong charges
correction of QM/MM energies:

II • First, the interaction with the full MM charges has to be restored:

$$\delta W_{QM/MM}(\alpha_i \rightarrow 1) = \sum_{ij}(1 - \alpha_i)q_j\Phi_v(i \rightarrow QM)$$

III • Second, the solvation energy of the QM/MM system is computed from PB. Here, QM charges can be obtained from Mulliken population analysis or from ESP type charges.

Fig. 4. Reaction profile obtained with Mulliken charges representing the QM atoms. Total (*bold solid line*), step I (*thin solid line*), step IIa (*long dashed line*), and step III (*short dashed line*) energies. The

without CS:
profiles I+II: completely wrong

$$\delta W_{QM/MM}(\alpha_i \rightarrow 1) = \sum_{ij}(1 - \alpha_i)q_j\Phi_v(i \rightarrow QM)$$

III   • Second, the solvation energy of the QM/MM system is computed from PB. Here, QM charges can be obtained from Mulliken population analysis or from ESP type charges.

Generalized solvent boundary potential (GSBP)

PB very expensive, even for MM: can not be done extensively along MD simulations

outer region: fixed

inner region

Generalized solvent boundary potential (GSBP)
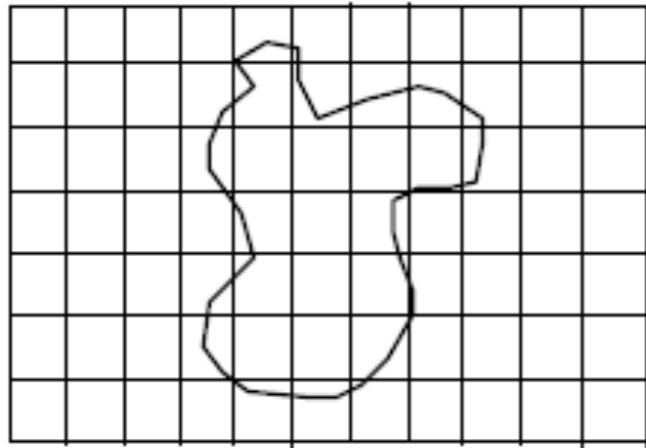
PB very expensive, even for MM: can not be done extensively along MD simulations

outer region: fixed

inner region

Generalized solvent boundary potential (GSBP by B. Roux)
▪ fix an outer region
▪ calculate reaction field and solvation free energy using a basis set for the inner region

Generalized solvent boundary potential (GSBP)

PB very expensive, even for MM: can not be done extensively along MD simulations

outer region: fixed

inner region

Generalized solvent boundary potential (GSBP by B. Roux)
- fix an outer region
- calculate reaction field and solvation free energy using a basis set for the inner region

➔ Very good for problems, where one can neglect the motion of the outer part, i.e. 'outer' atoms can be represented by average coordinates

Generalized solvent boundary potential (GSBP)

outer region: fixed

$(X_m, r_1 ... r_n)$

inner region

$(r_{n+1} ... r_N)$.

Beglov, Im, Roux (JCP 100, 9050, JCP 114, 2924)

Idea: 'Construct boundary potential, which relates the statistical properties of the infinite system to that of the finite (inner) system'

Potential of mean force (PMF) of inner system (frozen)

$W(X_m, r_1 ... r_n)$ by integration over 'outer' degrees of freedom $(r_{n+1} ... r_N)$.

$$e^{-\beta W(X_m, r_1 ... r_n)} = C \int dr_{n+1} ... dr_N e^{-\beta U(X_m, r_1 ... r_n, ... r_N)}$$

Generalized solvent boundary potential (GSBP)
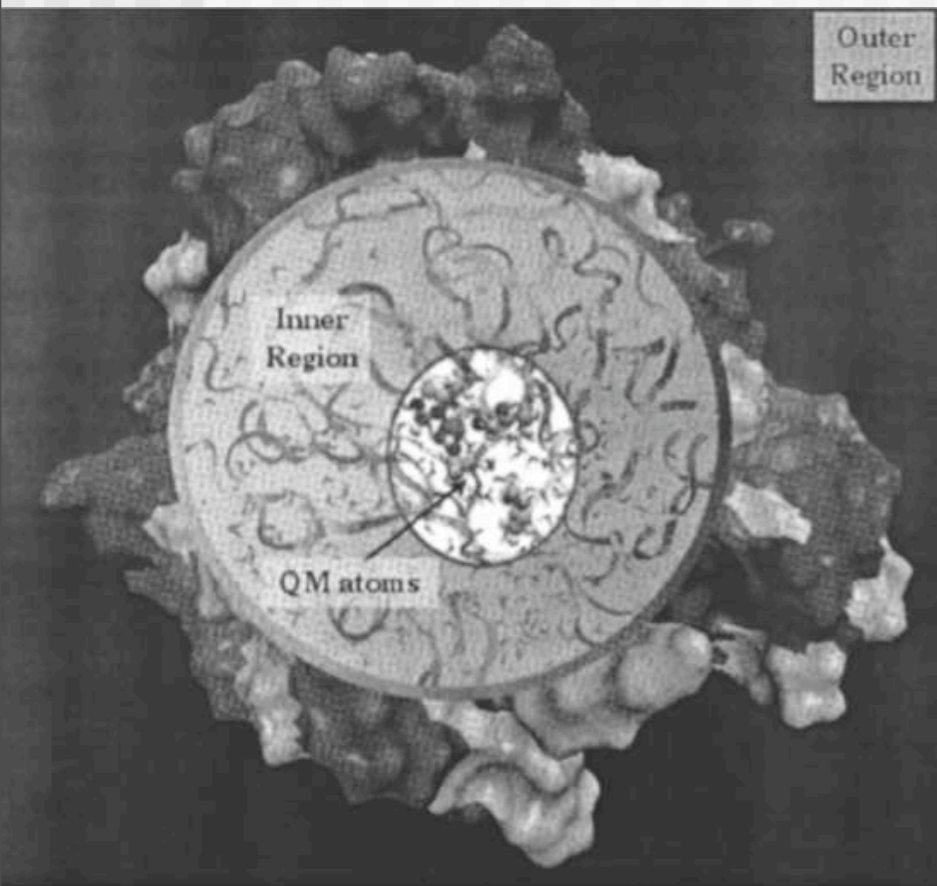
Potential of mean force (PMF) of inner system (frozen)

$W(X_m, r_1...r_n)$ by integration over 'outer' degrees of freedom.

U_{ii} potential energy of inner system

W– U_{ii} can be identified with **solvation free energy** of
solute + n solvent degrees of freedom,
frozen at coordinates $(X_m, r_1...r_n)$

$$\rho(x) = \rho^{(i)}(x) + \rho^{(o)}(x)$$

$\Delta W = W-U = (W_{ii} + W_{io} + W_{oo}) - (U_{ii} + U_{io} + U_{oo})$

Generalized solvent boundary potential (GSBP)

outer region: fixed

inner region

ΔW is the free energy needed to assemble the inner system for a chosen conformation in the presence of the environment:

$$\Delta W = \Delta W_{cr} + \Delta W_{vdW} + \Delta W_{elec}$$

ΔW$_{cr}$ insert (e.g. fixed) sphere, configurational constriction

ΔW$_{vdW}$ contains VdW interactions between inner and outer region

$$\Delta W_{elec}(\mathbf{X}) = \Delta W_{elec}{}^{(oo)}(\mathbf{X}) + \Delta W_{elec}{}^{(io)}(\mathbf{X}) + \Delta W_{elec}{}^{(ii)}(\mathbf{X})$$

const.

Solvation free energy for the macromolecule

$$\Delta W_{elec}(\mathbf{X}) = \Delta W_{elec}^{(oo)}(\mathbf{X}) + \Delta W_{elec}^{(io)}(\mathbf{X}) + \Delta W_{elec}^{(ii)}(\mathbf{X})$$

Various components using a basis set expansion:

25-50 Å

inner

$\epsilon = 80$

$\epsilon = 6$ $\epsilon = 4$

$\epsilon = 2$

54 Å

$$\Delta W_{elec}^{(io)} = \sum_{\alpha \in inner} q_\alpha \phi_{rf}^{(o)}(\mathbf{r}_\alpha)$$

$$\Delta W_{elec}^{(io)}(\mathbf{X}) + U_{elec}^{(io)}(\mathbf{X}) = \sum_{\alpha \in inner} q_\alpha \phi_s^{(o)}(\mathbf{r}_\alpha)$$

1) compute reaction field
from linearized PB equation before MD

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r})\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}).$$

save and use it to compute the (io)
terms along MD of inner atoms

Solvation free energy for the macromolecule

25-50 Å

inner

$\epsilon = 80$

$\epsilon = 6$ $\epsilon = 4$

$\epsilon = 2$

54 Å
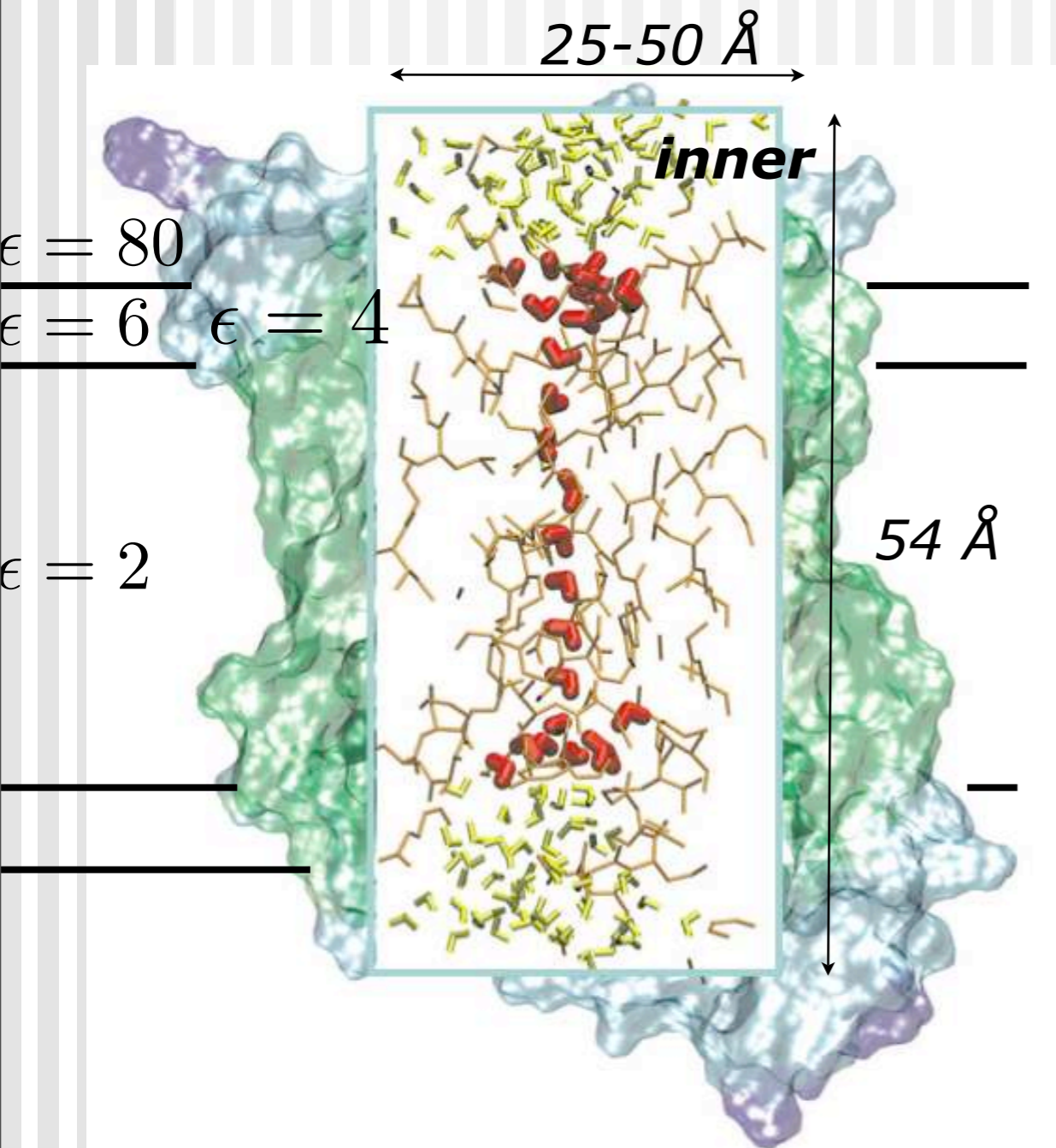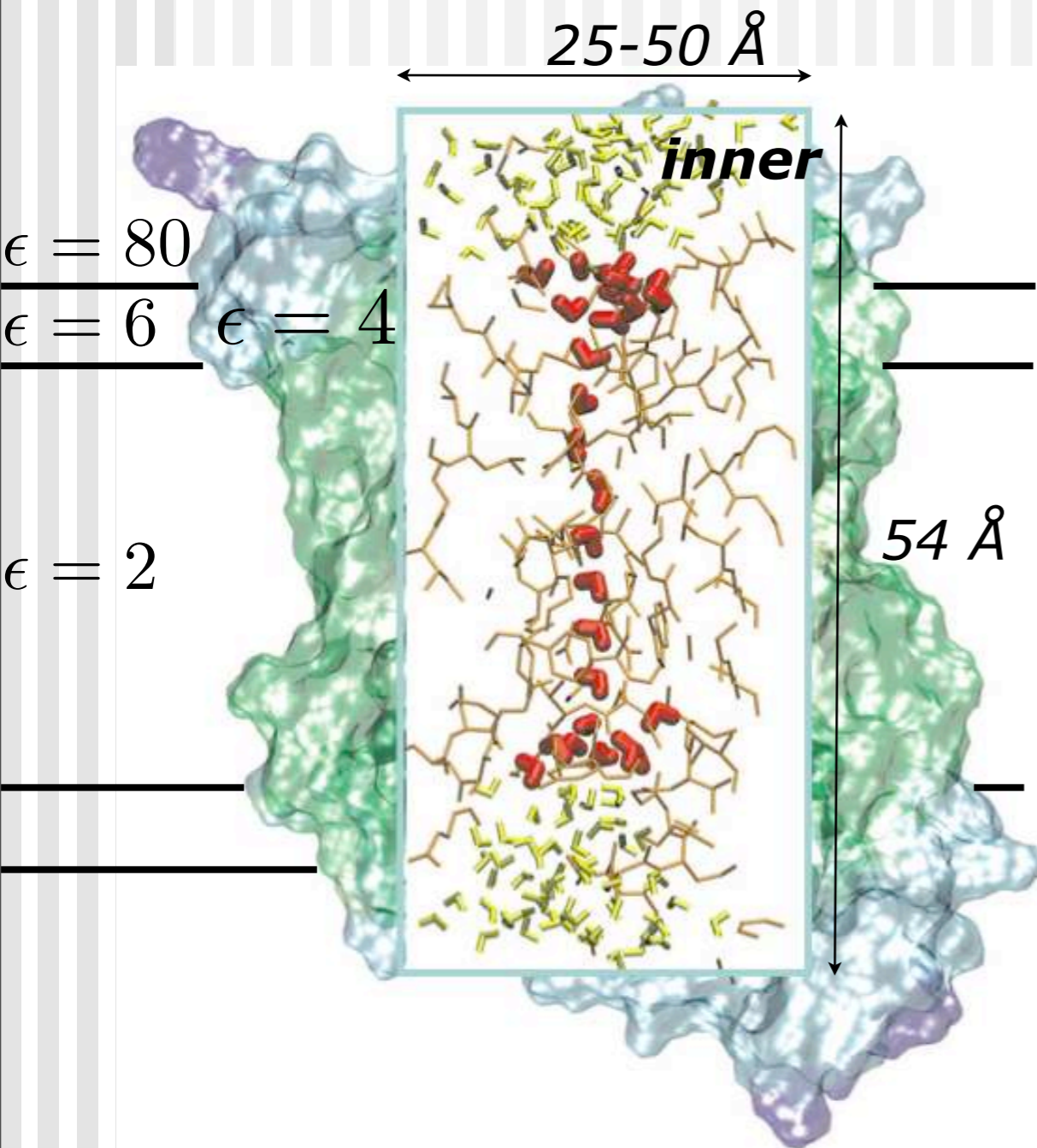
The (ii) is the complicated term, since it arises from the inner atoms and would have to be update every MD step

$$\Delta W_{elec}^{(ii)} = \sum_{\alpha \in inner} q_\alpha \phi_{rf}^{(i)}(\mathbf{r}_\alpha)$$

or, more general:

$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3 r$$

introduce $b_n(\boldsymbol{r})$: basis functions,
        depend on the system geometry

$$\rho^{(i)}(x) = \sum_m c_m b_m(r_\alpha)$$

Solvation free energy for the macromolecule

$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3r$$

Solvation free energy for the macromolecule

$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3r$$

introduce $b_n(\boldsymbol{r})$: basis functions,
depend on the system geometry

$$\rho^{(i)}(r) = \sum_m Q_m b_m(r)$$

Solvation free energy for the macromolecule

$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3 r$$

introduce $b_n(\boldsymbol{r})$: basis functions,
depend on the system geometry

$$\rho^{(i)}(r) = \sum_m Q_m b_m(r)$$

on the other hand

$$\rho^{(i)}(r) = \sum_{\alpha \epsilon inner} q_\alpha \delta(r - r_\alpha)$$

Solvation free energy for the macromolecule

$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3r$$

on the other hand

$$\rho^{(i)}(r) = \sum_{\alpha \epsilon inner} q_\alpha \delta(r - r_\alpha)$$

introduce $b_n(\boldsymbol{r})$: basis functions, depend on the system geometry

$$\rho^{(i)}(r) = \sum_m Q_m b_m(r)$$

Generalized multipole moment:

$$Q_n = \sum_{\alpha \in inner} q_\alpha b_n(\mathbf{r}_\alpha)$$

Solvation free energy for the macromolecule

$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3 r$$

on the other hand

$$\rho^{(i)}(r) = \sum_{\alpha \in inner} q_\alpha \delta(r - r_\alpha)$$

$$\phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) = \sum_m Q_m \phi_{rf}^{(i)}(b_m, \epsilon_i, \epsilon_o)$$

introduce $b_n(\boldsymbol{r})$: basis functions,
depend on the system geometry

$$\rho^{(i)}(r) = \sum_m Q_m b_m(r)$$

Generalized multipole moment:

$$Q_n = \sum_{\alpha \in inner} q_\alpha b_n(\mathbf{r}_\alpha)$$

Solvation free energy for the macromolecule

$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3 r$$

on the other hand

$$\rho^{(i)}(r) = \sum_{\alpha \in inner} q_\alpha \delta(r - r_\alpha)$$

$$\phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) = \sum_m Q_m \phi_{rf}^{(i)}(b_m, \epsilon_i, \epsilon_o)$$

$$\Delta W_{elec}^{(ii)}(\mathbf{X}) = \frac{1}{2} \sum_{mn} Q_m M_{mn} Q_n$$

introduce $b_n(\boldsymbol{r})$: basis functions, depend on the system geometry

$$\rho^{(i)}(r) = \sum_m Q_m b_m(r)$$

Generalized multipole moment:

$$Q_n = \sum_{\alpha \in inner} q_\alpha b_n(\mathbf{r}_\alpha)$$

Rx field matrix (computed **once**):

$$M_{nm} = \int d\mathbf{r}\ b_n(\mathbf{r})\phi_{rf}(\mathbf{r}; b_m(\mathbf{r}))$$

Solvation free energy for the macromolecule

1)
$$\Delta W_{elec}^{(io)}(\mathbf{X}) + U_{elec}^{(io)}(\mathbf{X}) = \sum_{\alpha \in inner} q_\alpha \phi_s^{(o)}(\mathbf{r}_\alpha)$$

2)
$$\Delta W_{elec}^{(ii)} = \frac{1}{2} \int \rho^i \phi_{rf}^{(i)}(\rho^i, \epsilon_i, \epsilon_o) d^3 r$$

$$\Delta W_{elec}^{(ii)}(\mathbf{X}) = \frac{1}{2} \sum_{mn} Q_m M_{mn} Q_n$$

Rx field matrix (computed **once**):

$$M_{nm} = \int d\mathbf{r} \; b_n(\mathbf{r}) \phi_{rf}(\mathbf{r}; b_m(\mathbf{r}))$$

this is the costly part:

compute for all basis functions the matrix elements in advance:

Membrane systems

$\epsilon = 80$

$\epsilon = 6$

$\epsilon = 2$

25 Å

54 Å
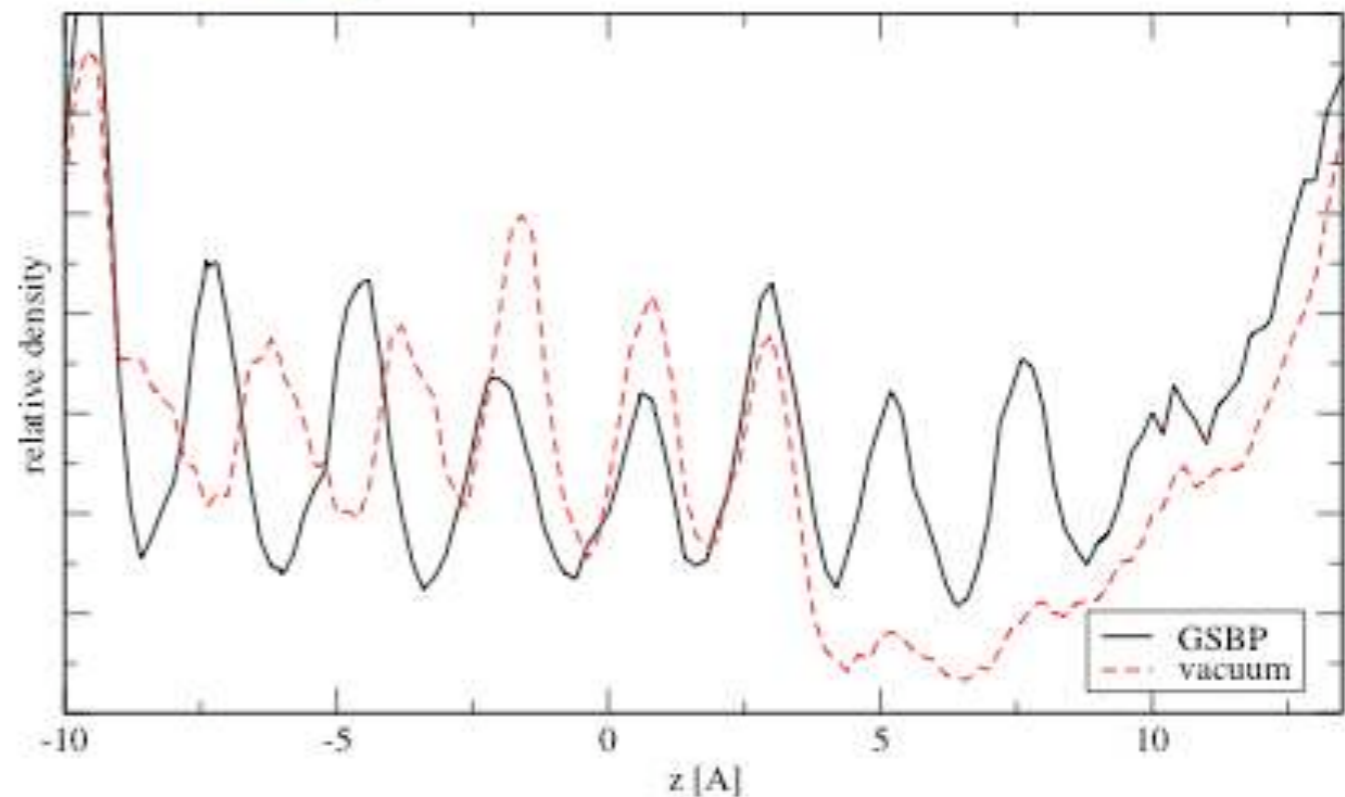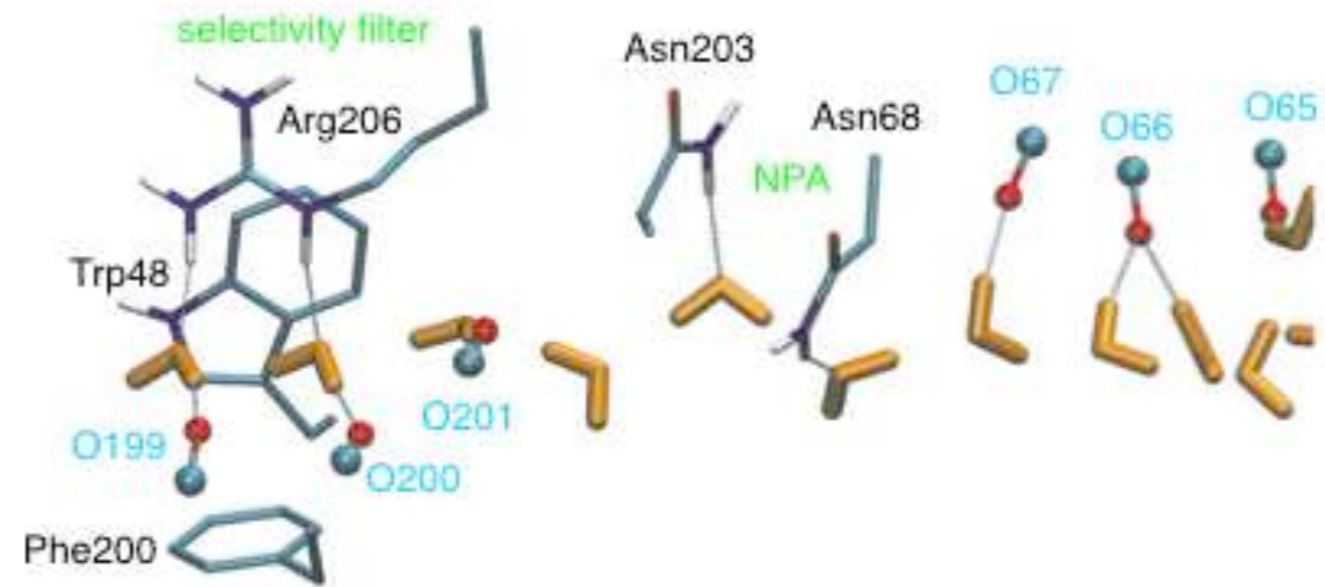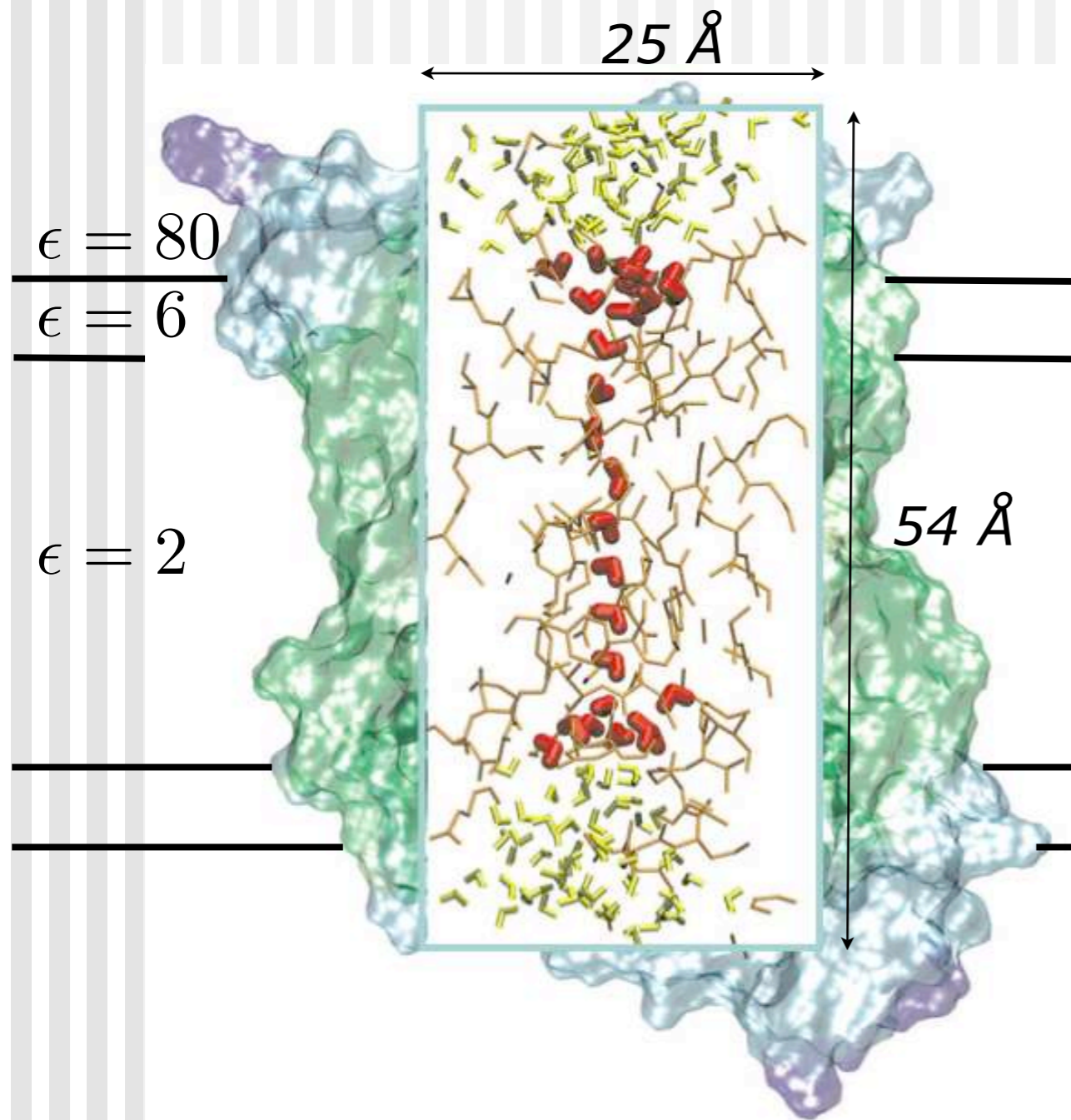
QM/MM implementation
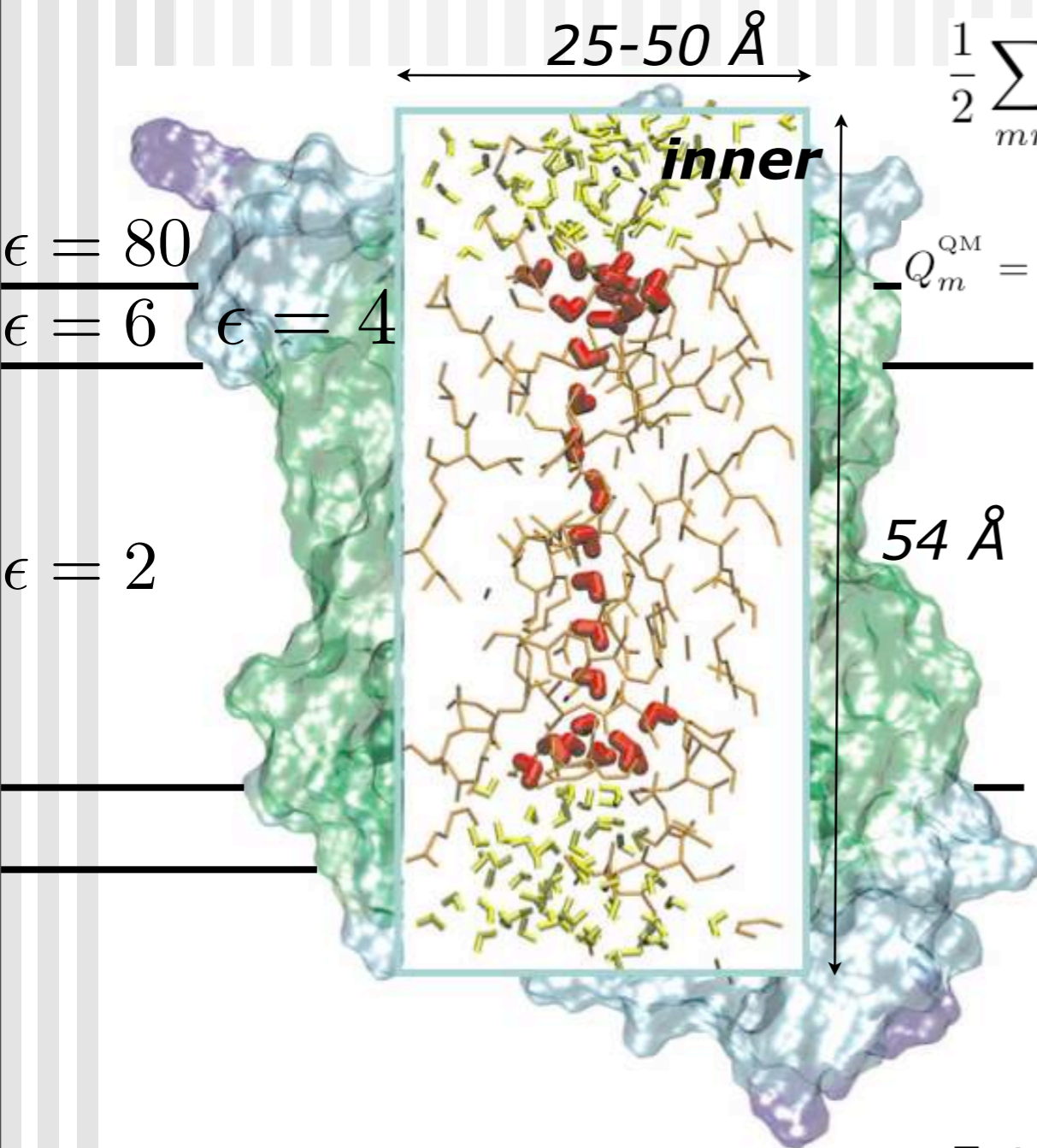
SCC-DFTB energy contribution:

$$\frac{1}{2}\sum_{mn} Q_m^{\mathrm{QM}} M_{mn}^* Q_n^{\mathrm{QM}} + \sum_{mn} Q_m^{\mathrm{QM}} M_{mn}^* Q_n^{\mathrm{MM}} + \int d\mathbf{r}\, \rho^{\mathrm{QM}}(\mathbf{r})\, \phi_s^{(o)}(\mathbf{r})$$

$$Q_m^{\mathrm{QM}} = \int d\mathbf{r}\, \rho^{\mathrm{QM}}(\mathbf{r})\, b_m(\mathbf{r}) = \int d\mathbf{r} \sum_{A\in\mathrm{QM}} \Delta q^A\, \delta(\mathbf{r}-\mathbf{R}_A)\, b_m(\mathbf{r}) = \sum_{A\in\mathrm{QM}} \Delta q^A\, b_m(\mathbf{R}_A)$$

SCC-DFTB matrix element contribution

$$\frac{1}{2} S_{\mu\nu} \sum_{B\in\mathrm{QM}} \left[\Gamma_{CB} + \Gamma_{DB}\right] \Delta q^B + \frac{1}{2} S_{\mu\nu} \left[\Omega(\mathbf{R}_C) + \Omega(\mathbf{R}_D)\right]$$

$$\Gamma_{AB}(\mathbf{R}_A, \mathbf{R}_B) \equiv \sum_{mn} b_m(\mathbf{R}_A) M_{mn} b_n(\mathbf{R}_B)$$

$$\Omega(\mathbf{R}_A) \equiv \sum_{m,n} b_m(\mathbf{R}_A)\, M_{mn}\, Q_n^{\mathrm{MM}}$$

Extension to other QM methods is straightforward

Effective QM/MM simulations for complex systems

Continuum region

MM region

QM region
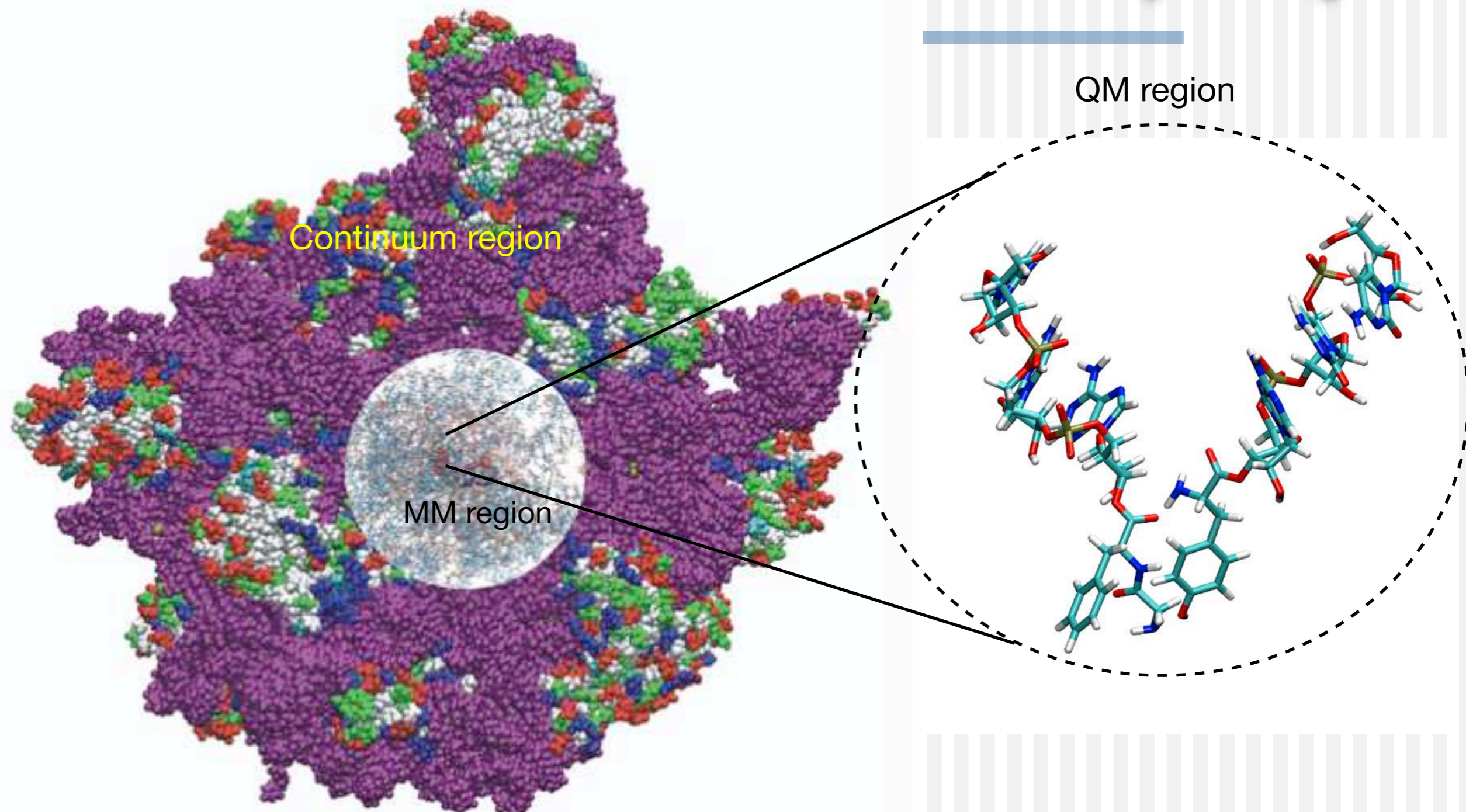
Generalized Boundary Condition, Electrostatics in QM/MM: P. Schaefer, D. Riccardi, QC, *J. Chem. Phys.* 123, 014905 (**2005**)*; J. Phys. Chem. B.* 109, 17715 (**2005**)

Membrane-proteins: P. König, N. Ghosh, QC, *J. Phys. Chem. A (Truhlar Issue) 110, 548* (**2006**) [ccO, aqua porin, Hv channel]

van der Waals in QM/MM: D. Riccardi, G. Li, QC *J. Phys. Chem. B* 108, 6467 (**2004**)

QM/MM free energy perturbation: Li, G. QC, *J. Phys. Chem. B* 107, 8643 (**2003**); 107, 14521 (**2003**)

Approximate DFT methods (Proton affinity, H-bonding, Phosphate, metals, ET): Y. Yang, H. Yu, J. Yu, M. Elstner, QC, *J. Phys. Chem. B Feature Article*, 110, 6458 (2006); *J. Phys. Chem. A 111, 10861* (2007)

Efficiency
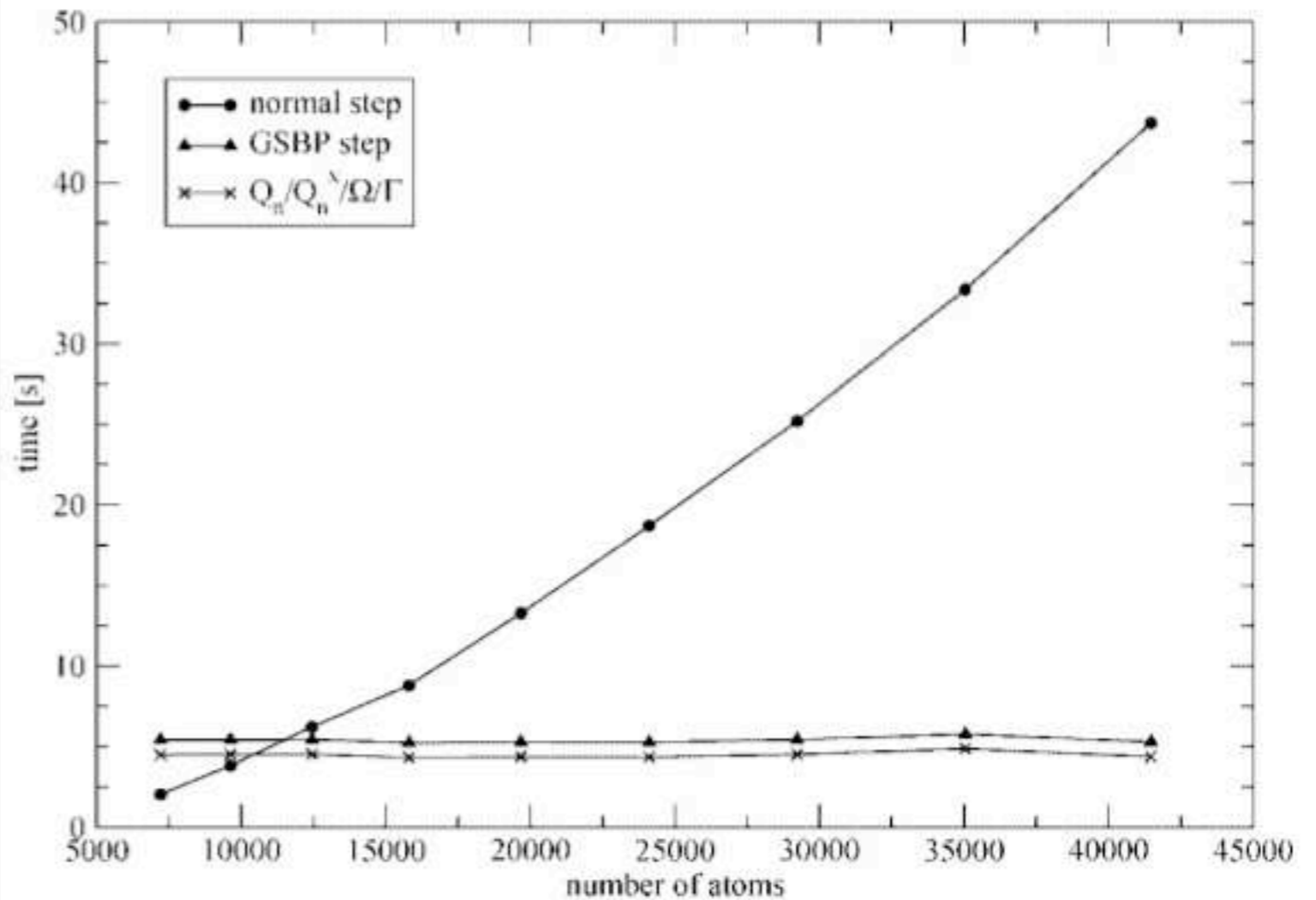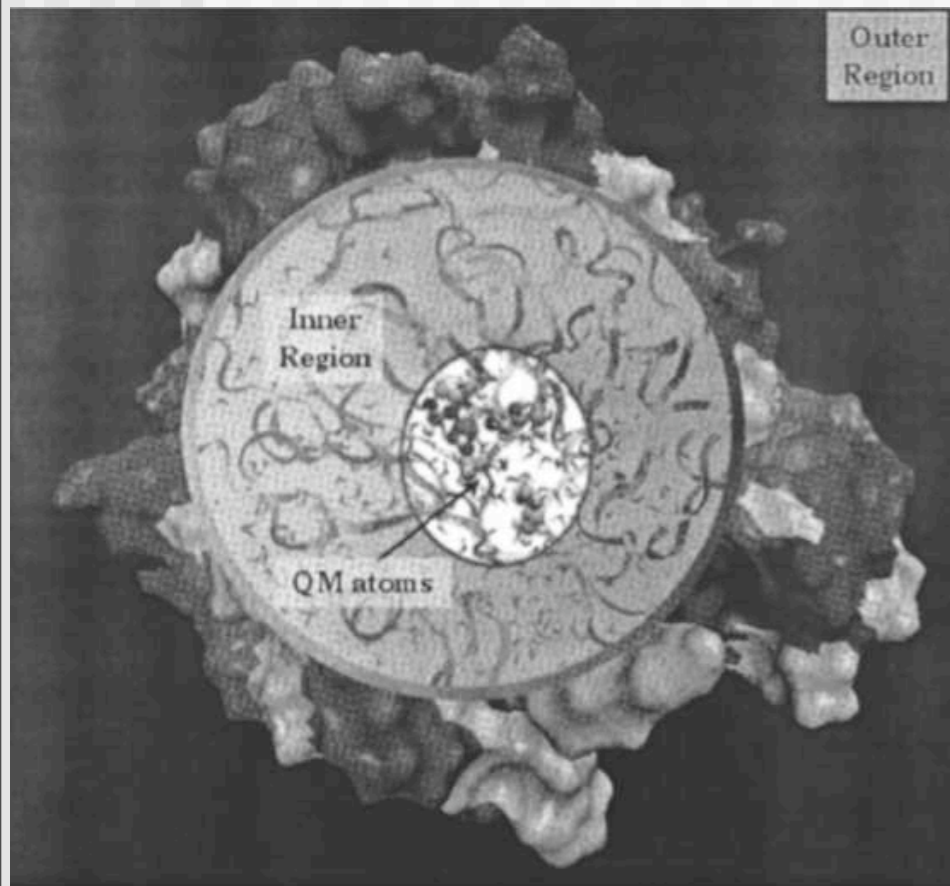
Threonine in a water droplet:
- inner region: 18 A (2700 atoms)
- total system: 7200-42.000 atoms

Benighaus & Thiel, JCTC 4 (2008) 1600.

Dynamics in electronically excited states and proton transfer
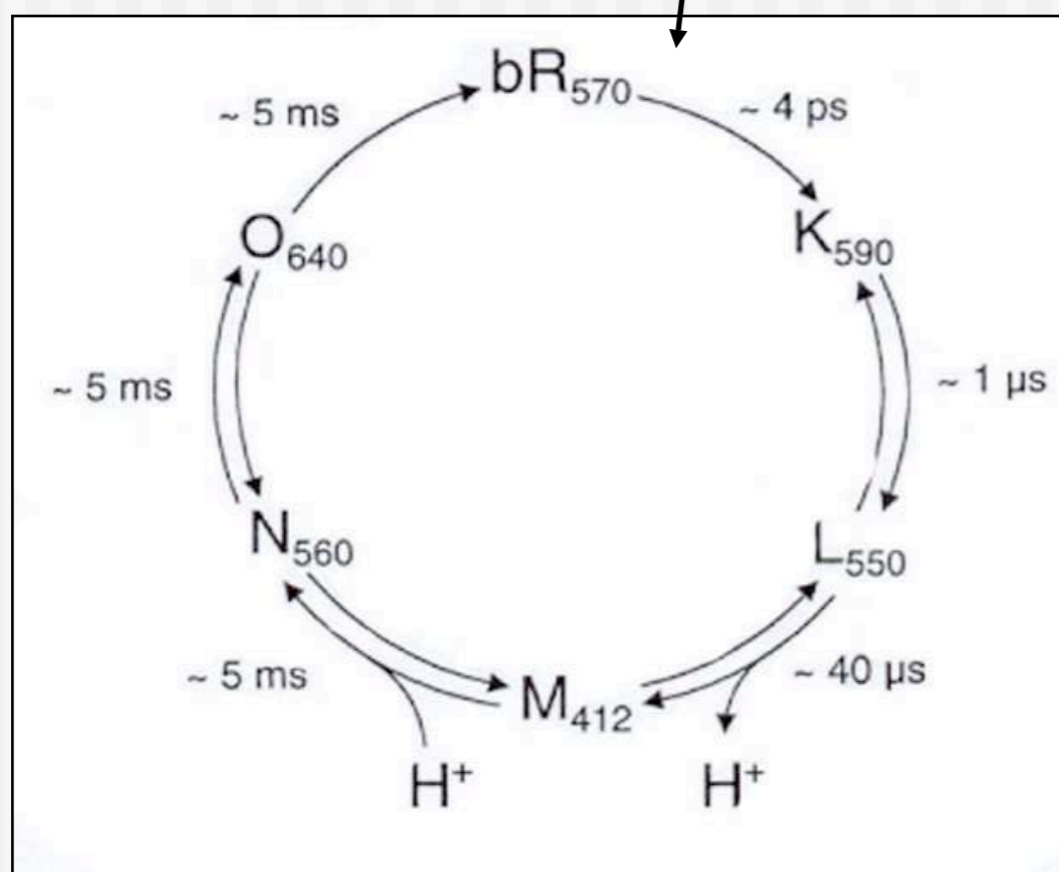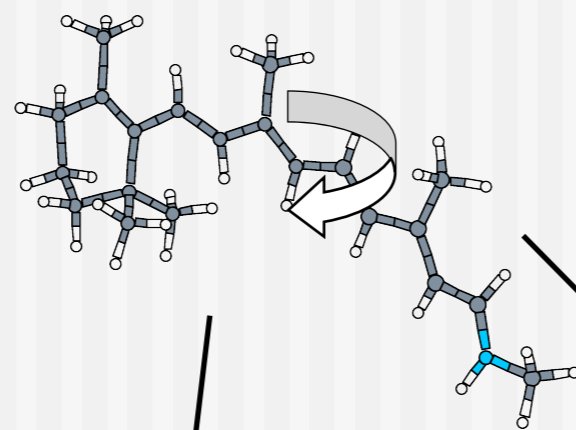
Bondar et al., JACS 44 (2004) 14668, Structure 12 (2004) 1281

Dynamics in electronically excited states and proton transfer

Bondar et al., JACS 44 (2004) 14668, Structure 12 (2004) 1281
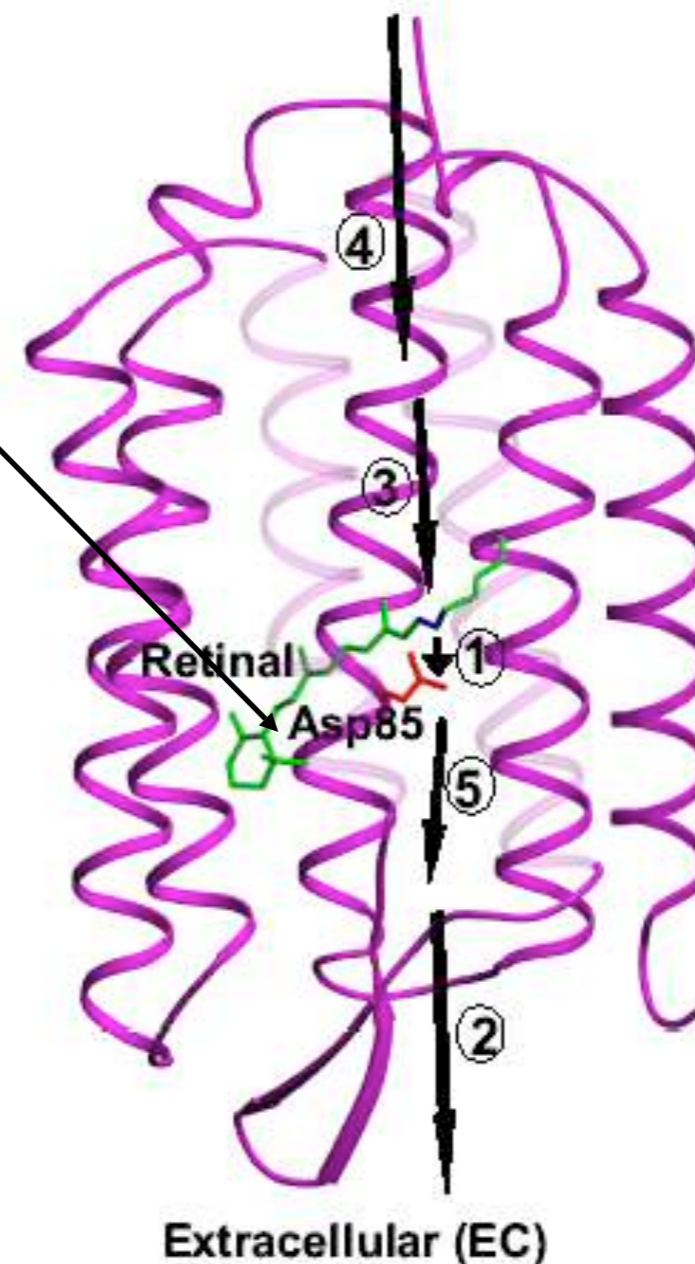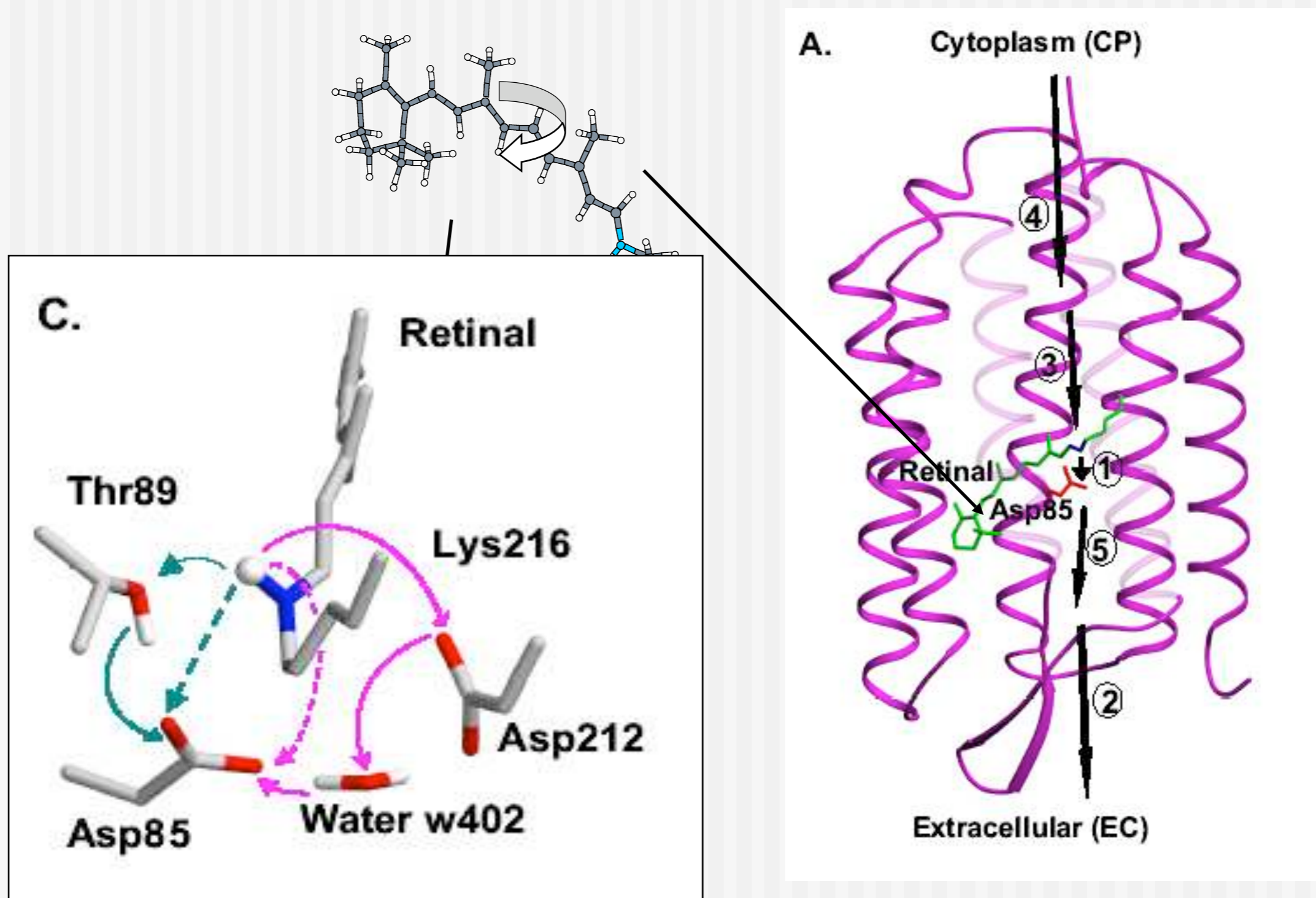
Proton transfer in bR
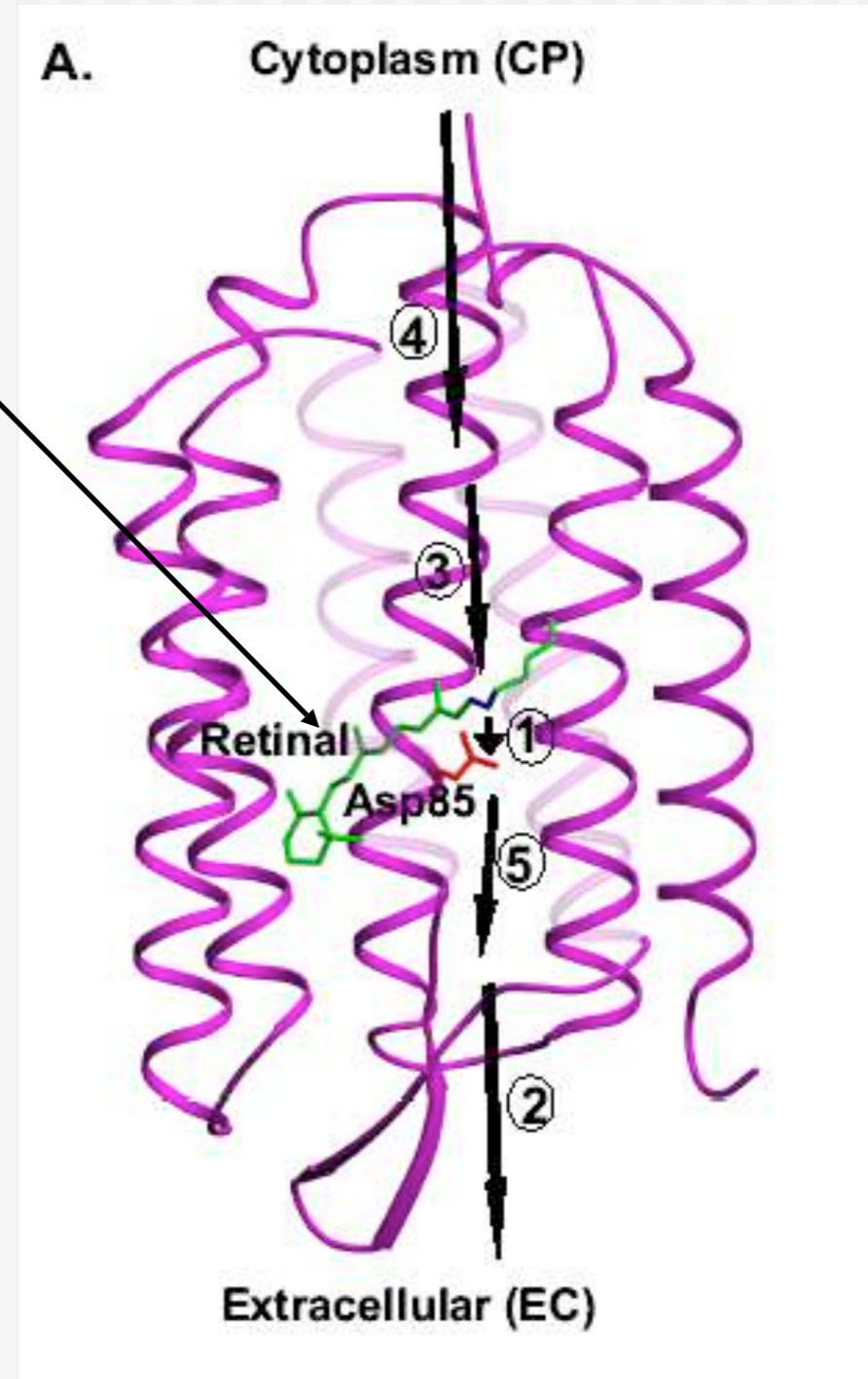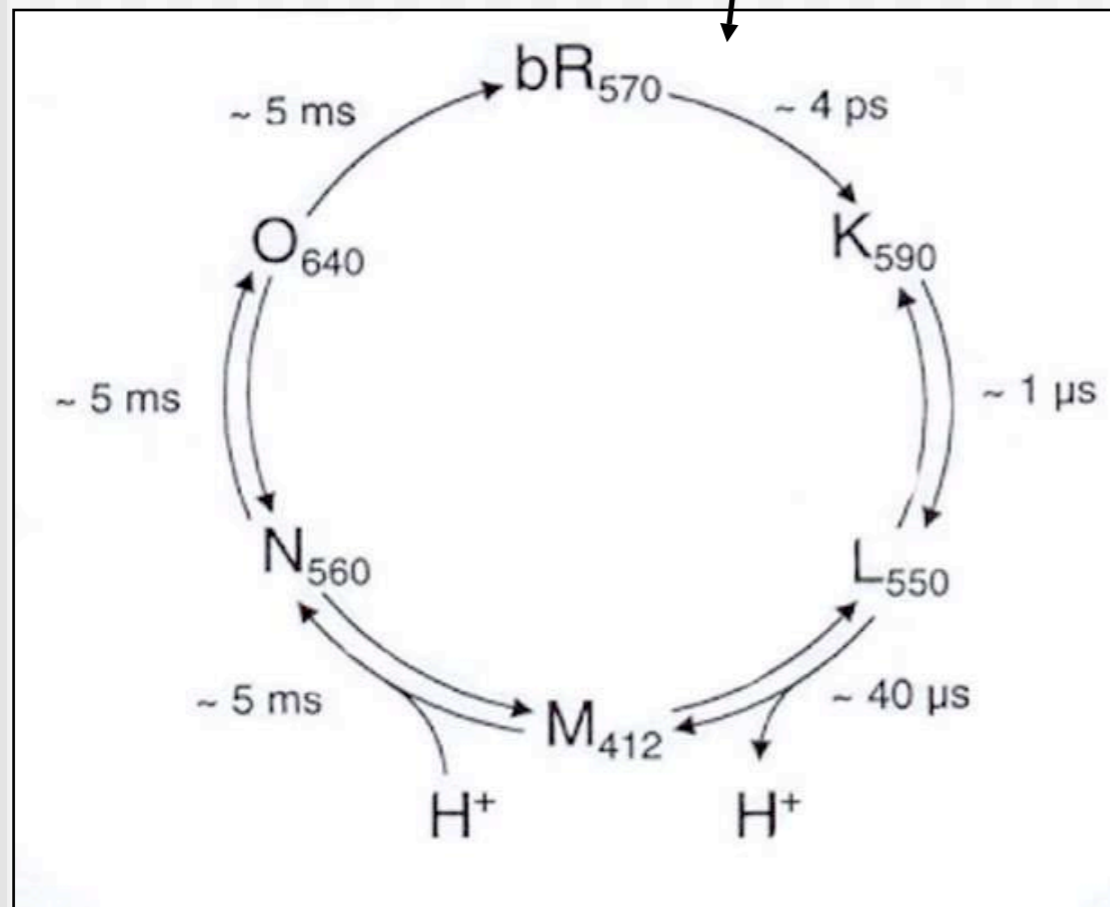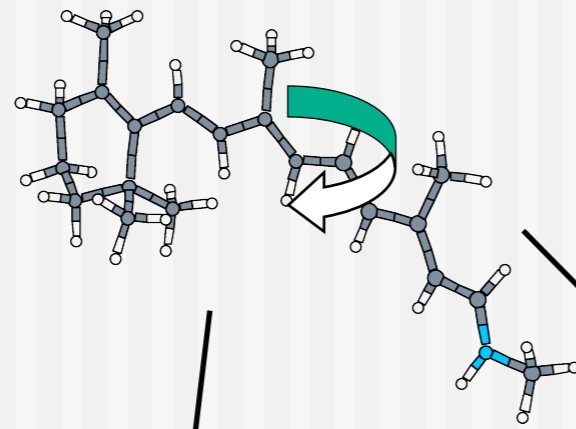
1ˢᵗ step: Photo-isomerisation
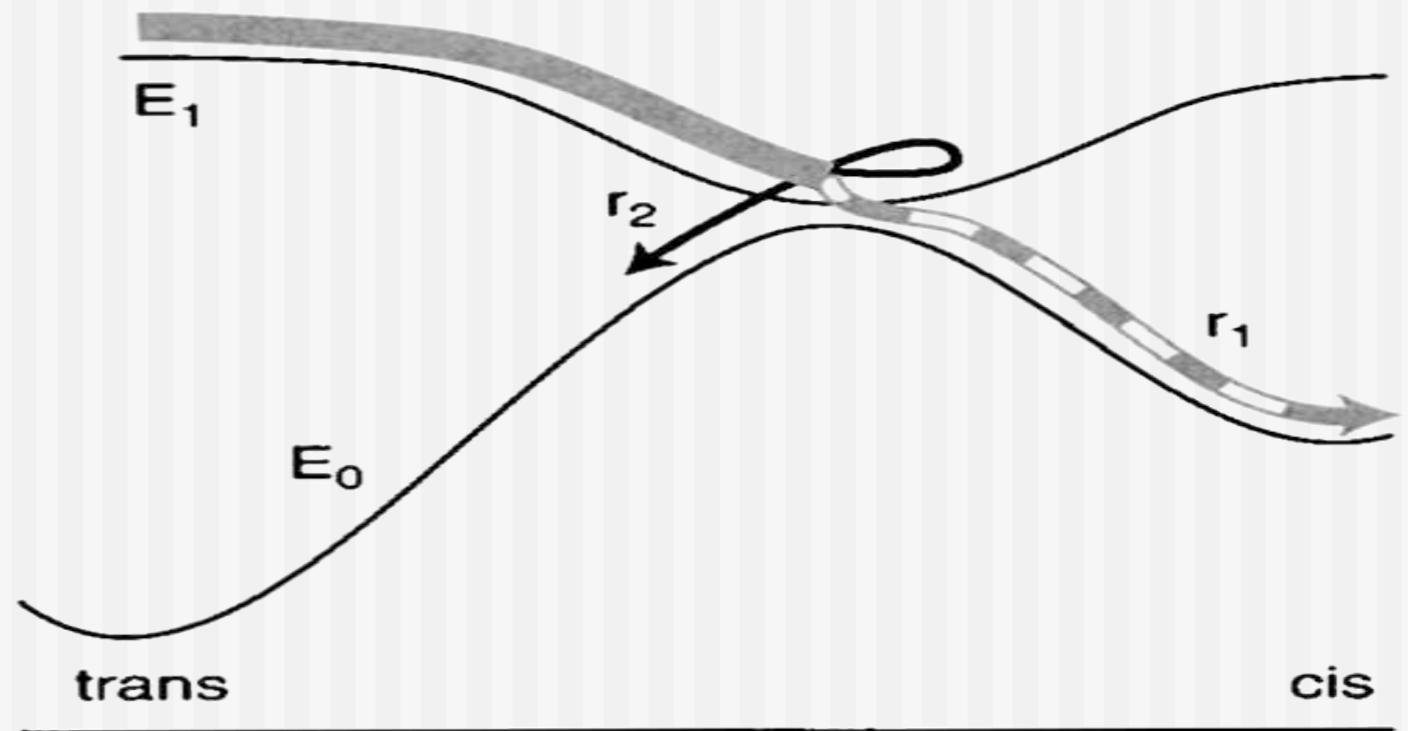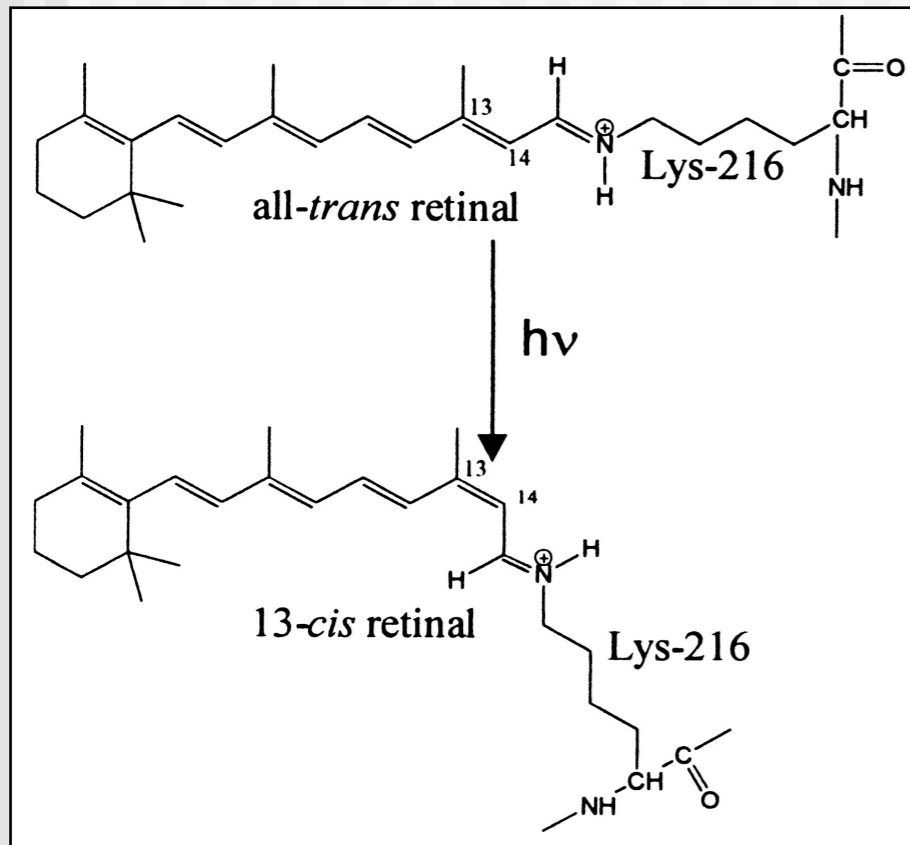
Reaction coordinate, efficiency

alcohol	not specific (11 cis)		10 ps	0.1
bR	trans	13 cis	0.5 ps	0.65

1ˢᵗ step: Photo-isomerisation
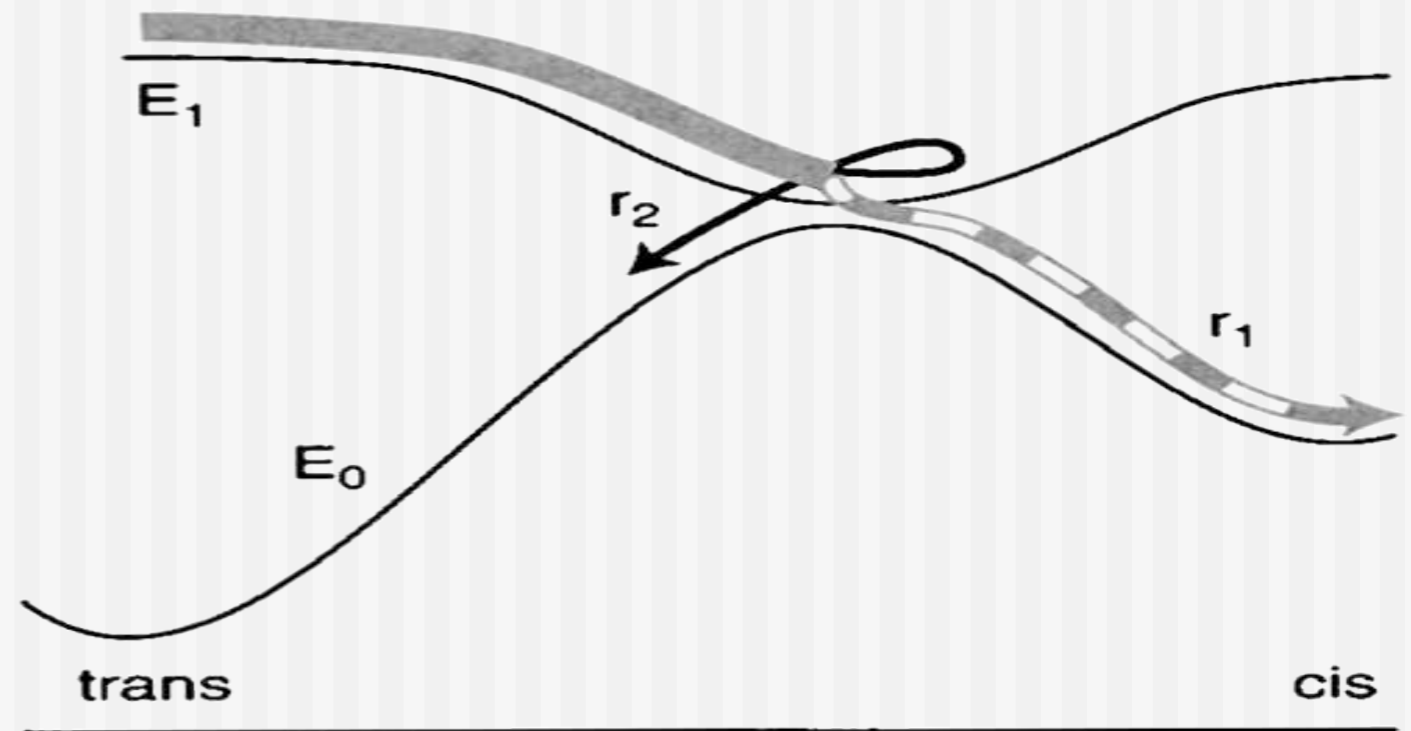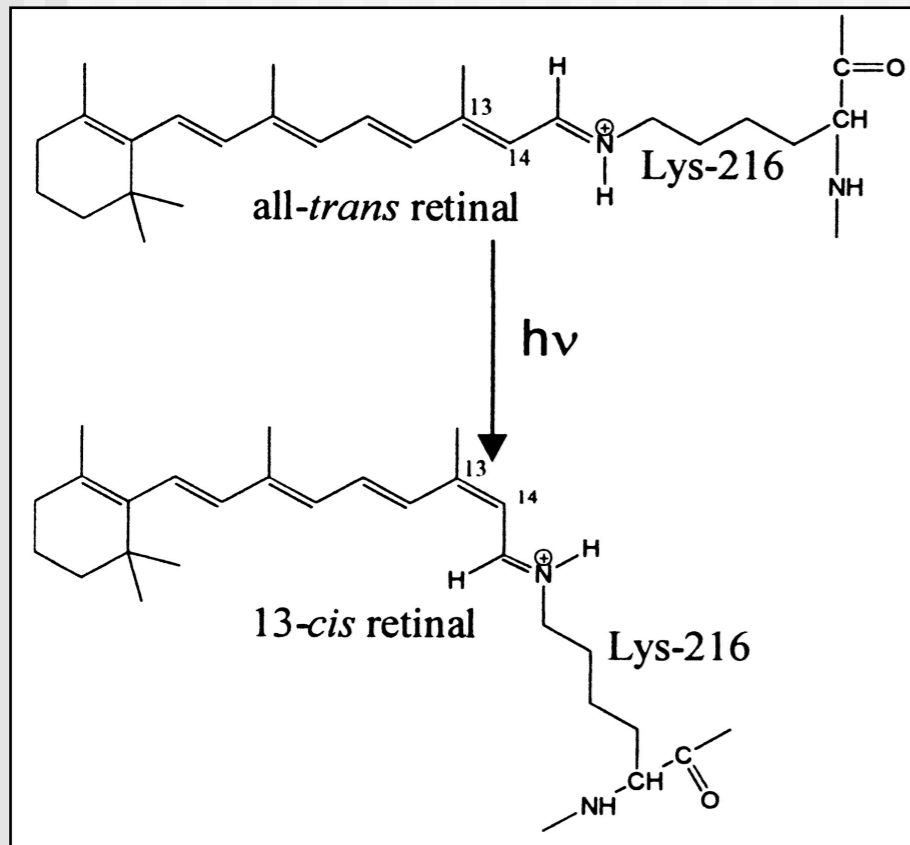
Reaction coordinate, efficiency

alcohol	not specific (11 cis)	10 ps	0.1
bR	trans	13 cis	0.5 ps	0.65

- 2 modes: C-C stretch and torsion
- specificity, speed?
- rotation in limited space:
  simple rotation?
- structure in K

How is the energy stored in K?
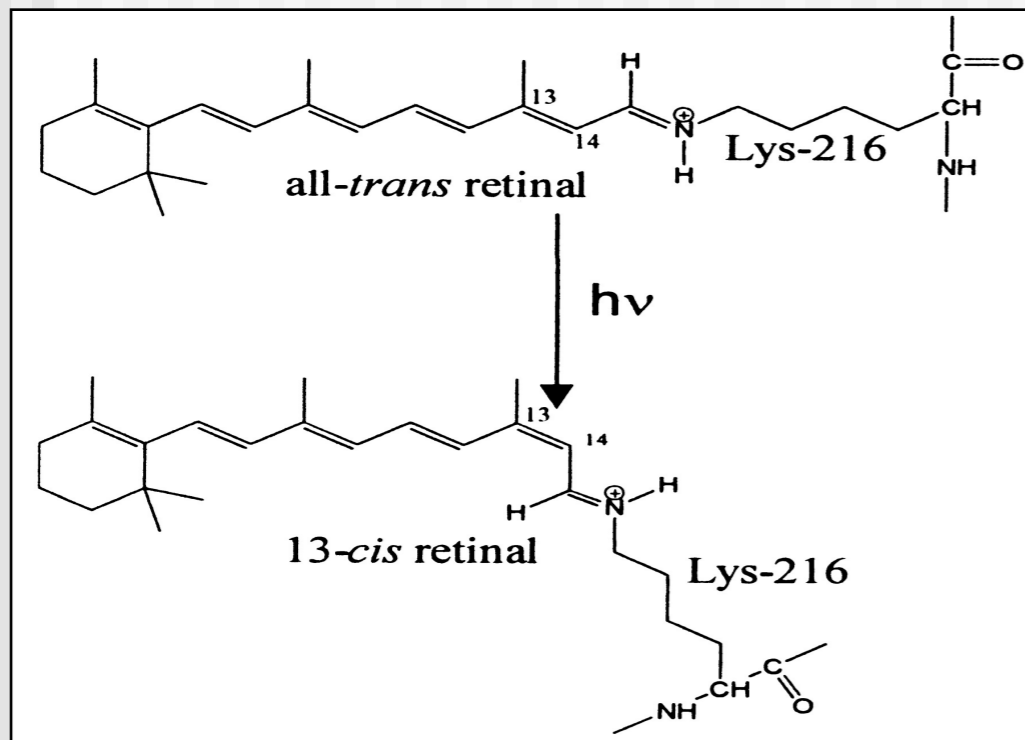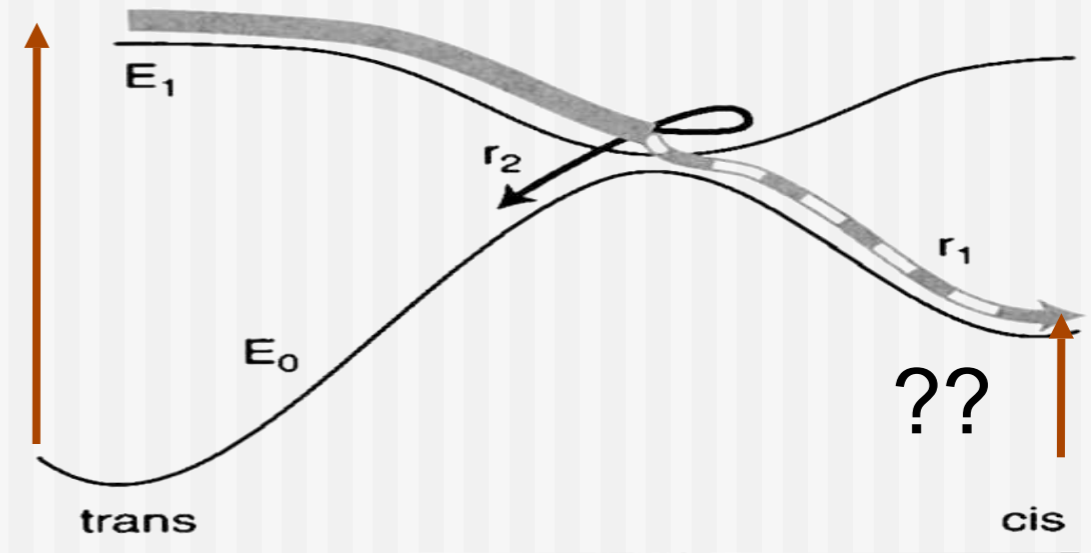
Rotation in limited space

Rotation in limited space
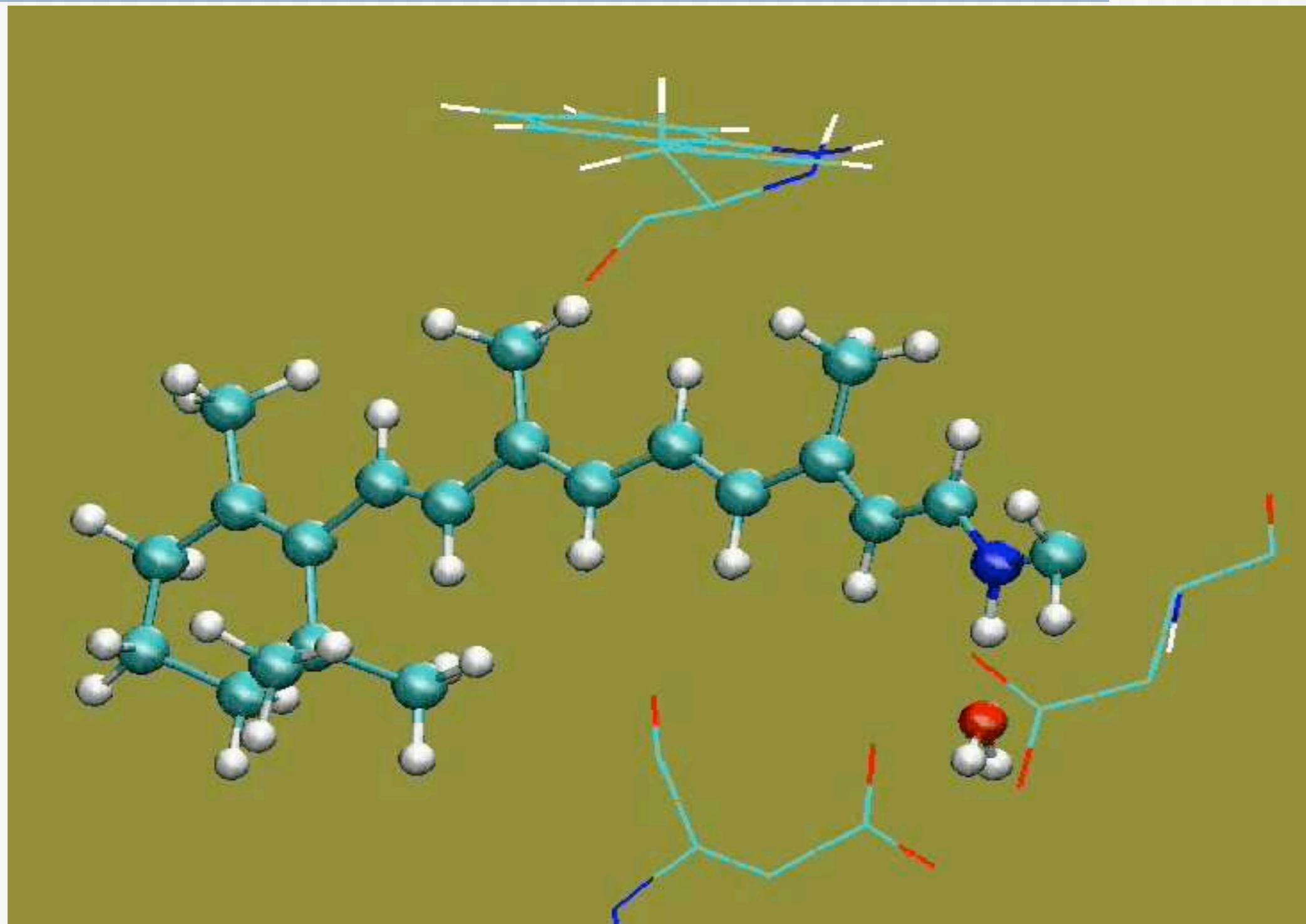
Energy scheme: bR-K-L
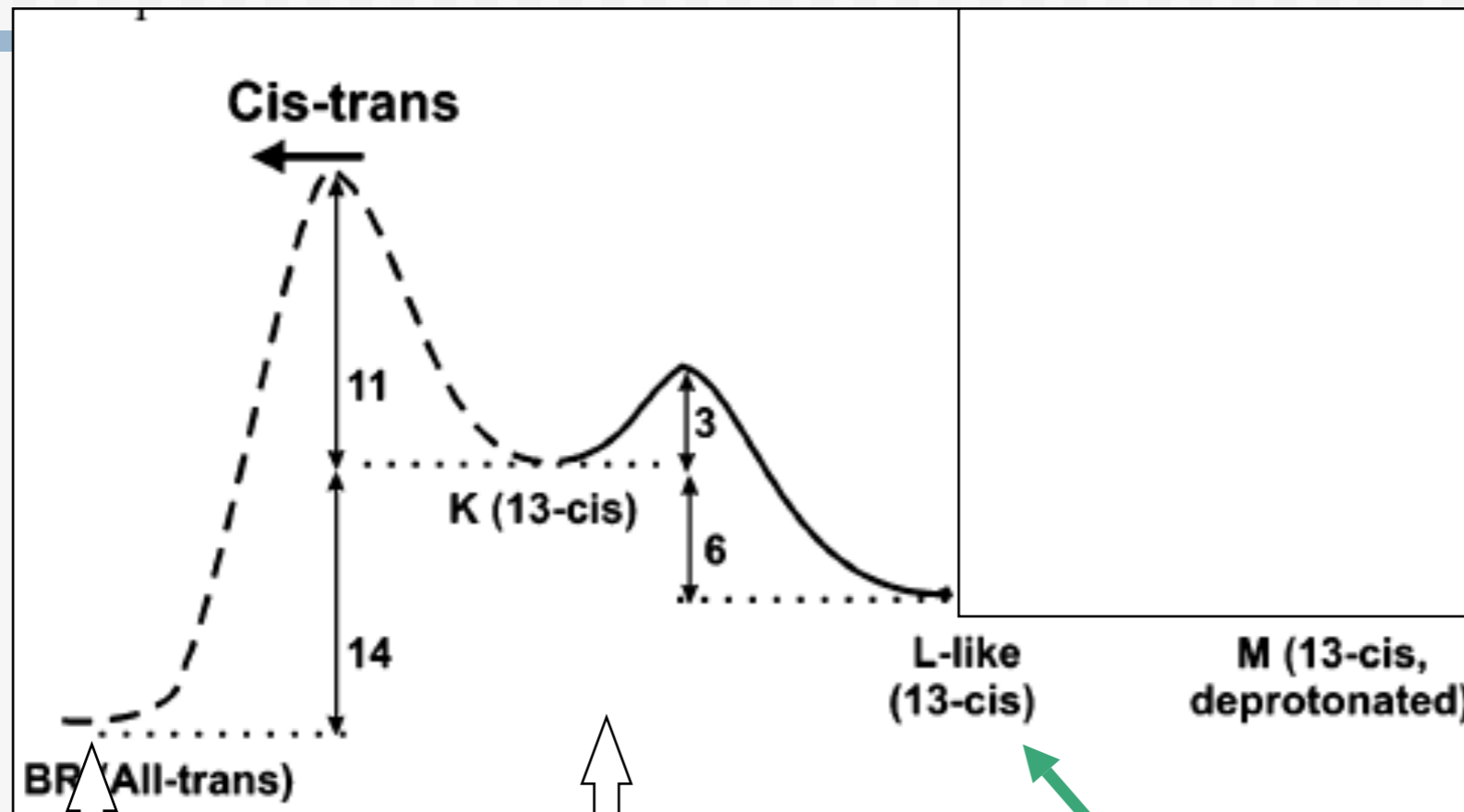
QM/MM: with protein interaction

14kcal~ 0.6 eV

0.3eV: twist

0.3eV: H-bonding
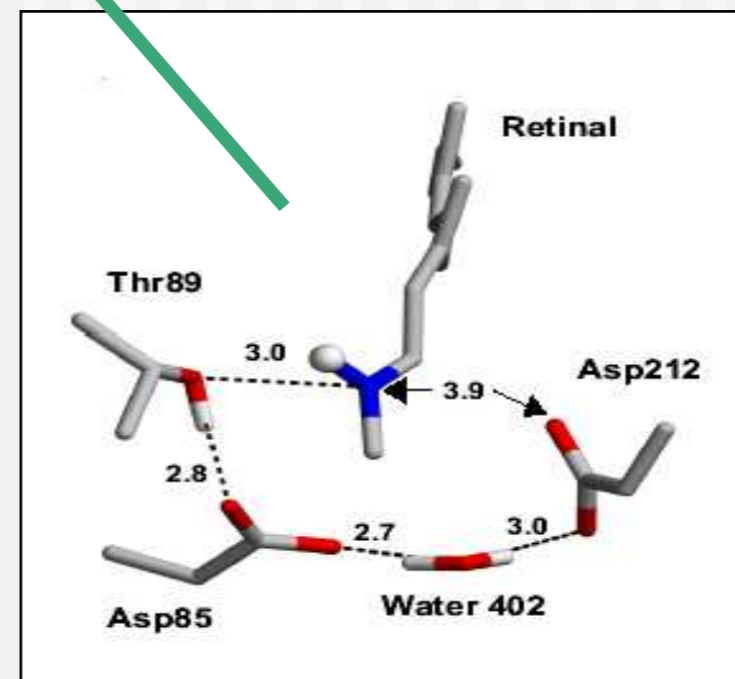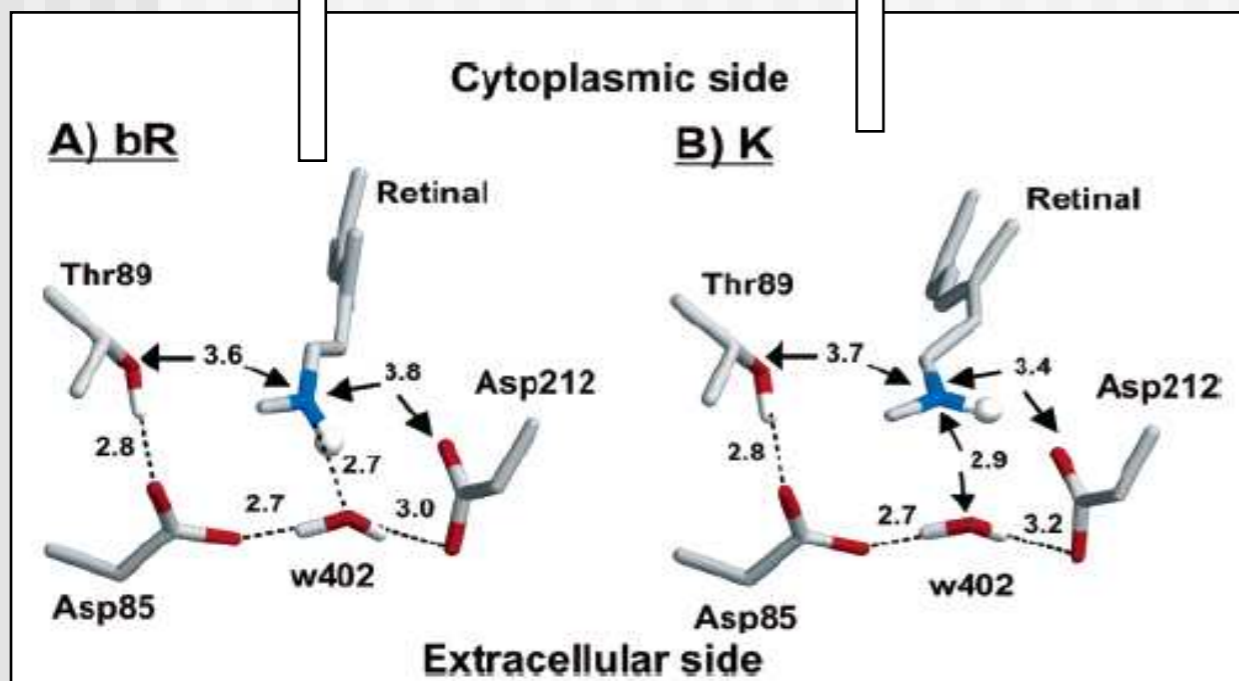
Change of pK_a → deprotonation

Proton transfer: bacteriorhodopsin
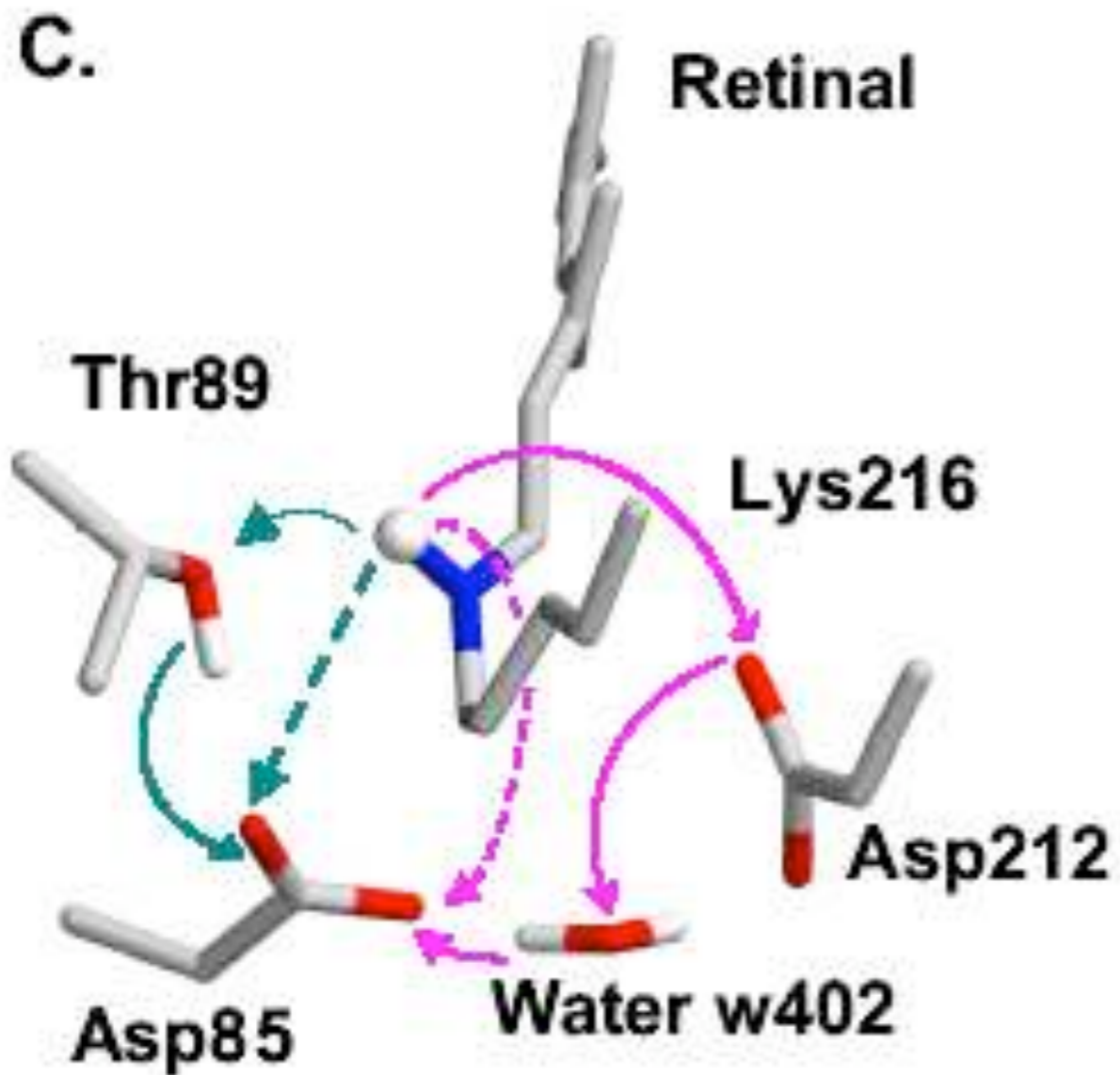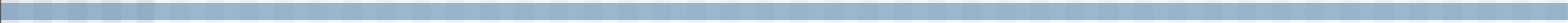
Energy scheme: bR-K-L

QM/MM: with protein interaction

14kcal~ 0.6 eV

0.3eV: twist

0.3eV: H-bonding
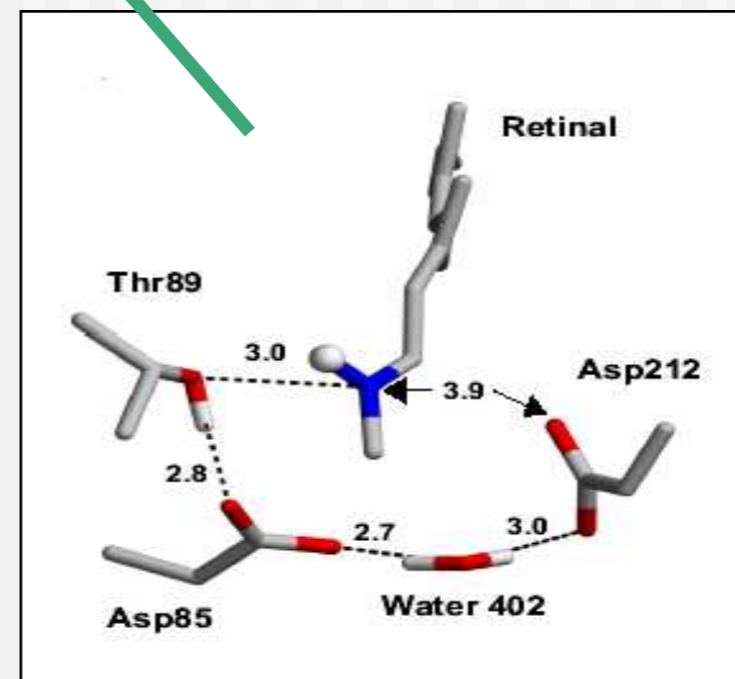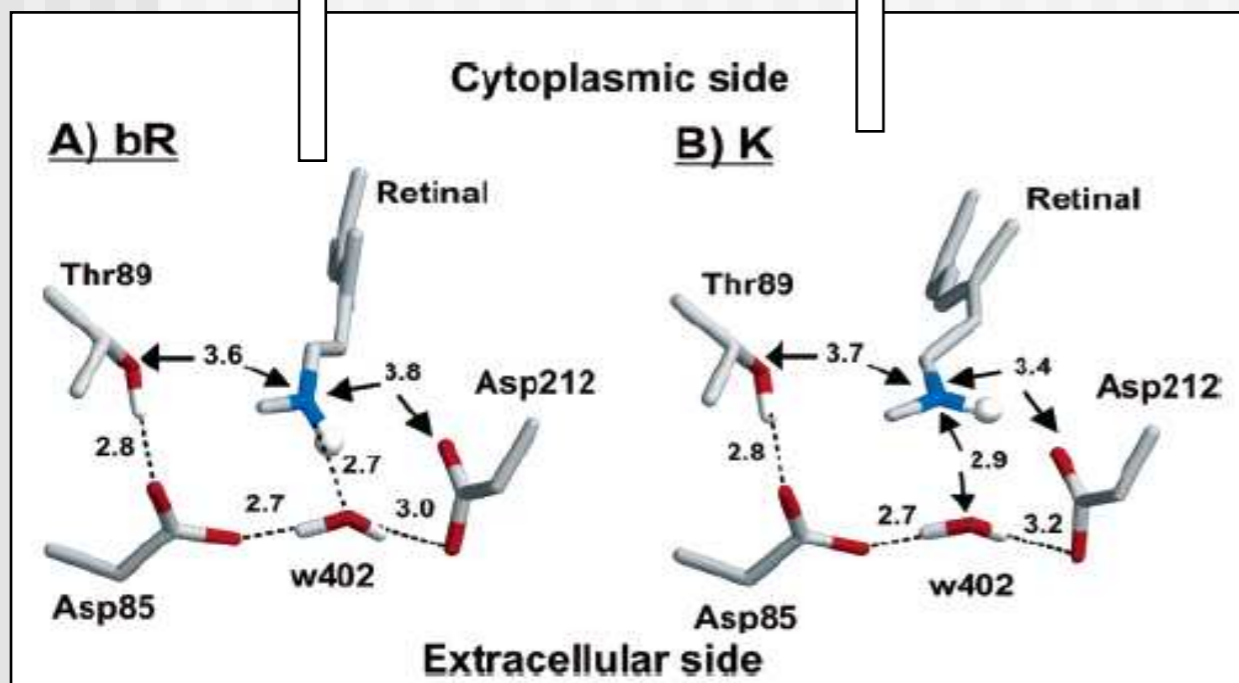
Change of pK_a → deprotonation

Directionality in the early photocycle steps
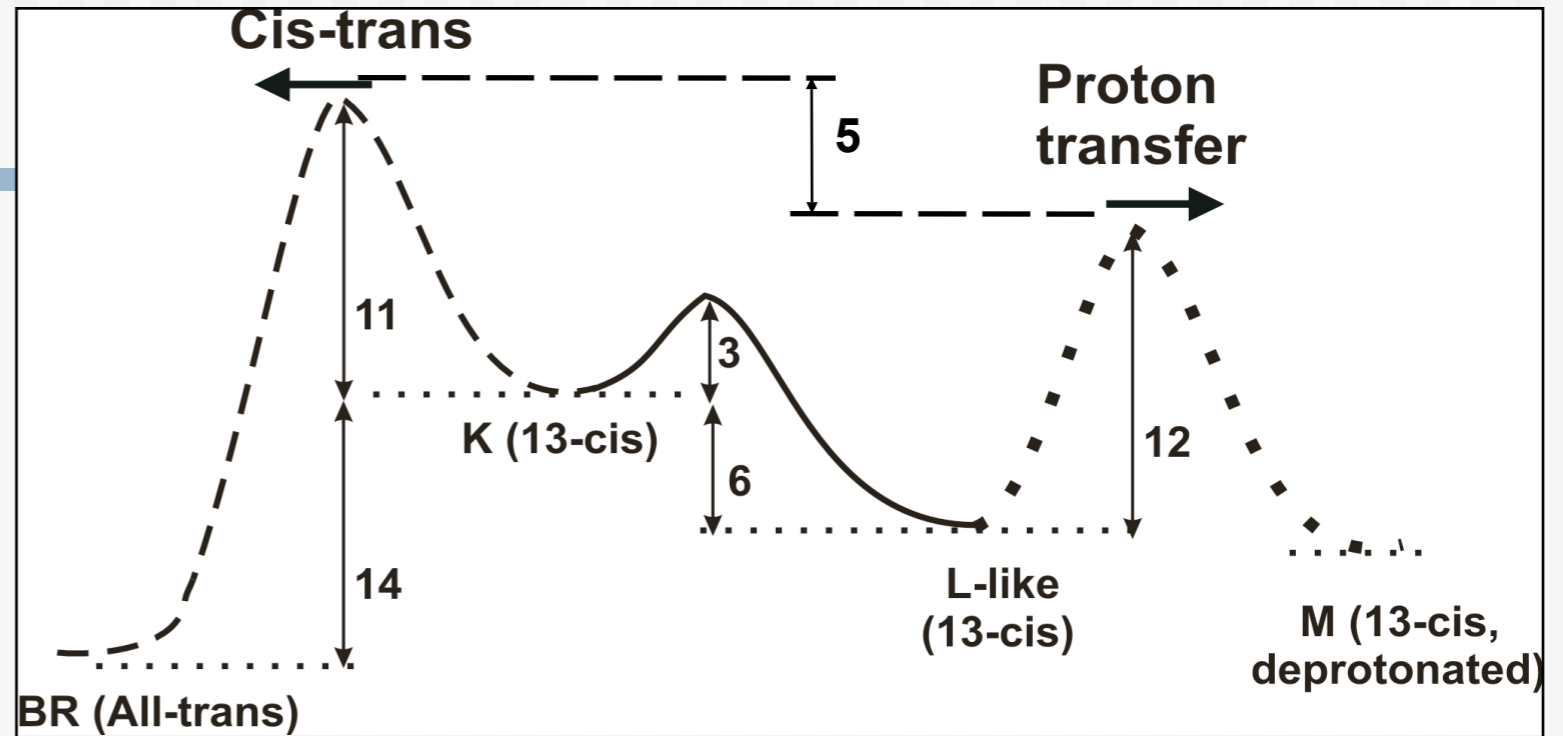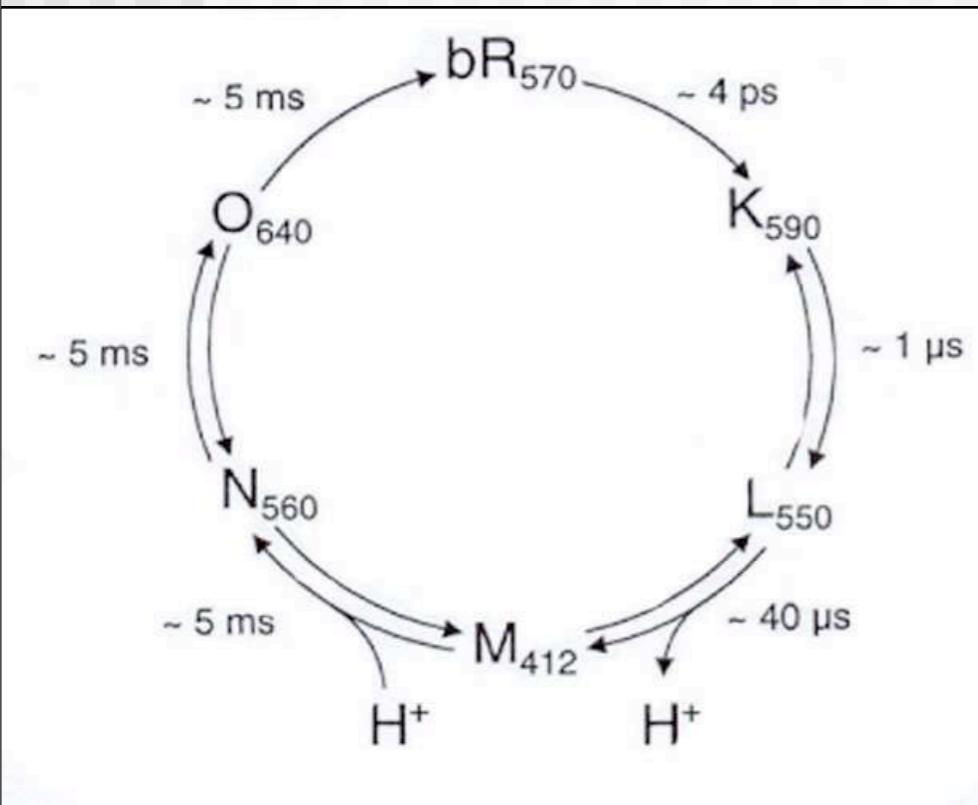
Energy storage

7 kcal/mol in retinal twist + 7kcal/mol in perturbation of h-bonding interactions

- enough energy is stored to drive the photocycle:
  membrane potential: 7kcal/mole

- thermal cis-trans isomerization is energetically unfavourable

- balance between productivity and energy storage
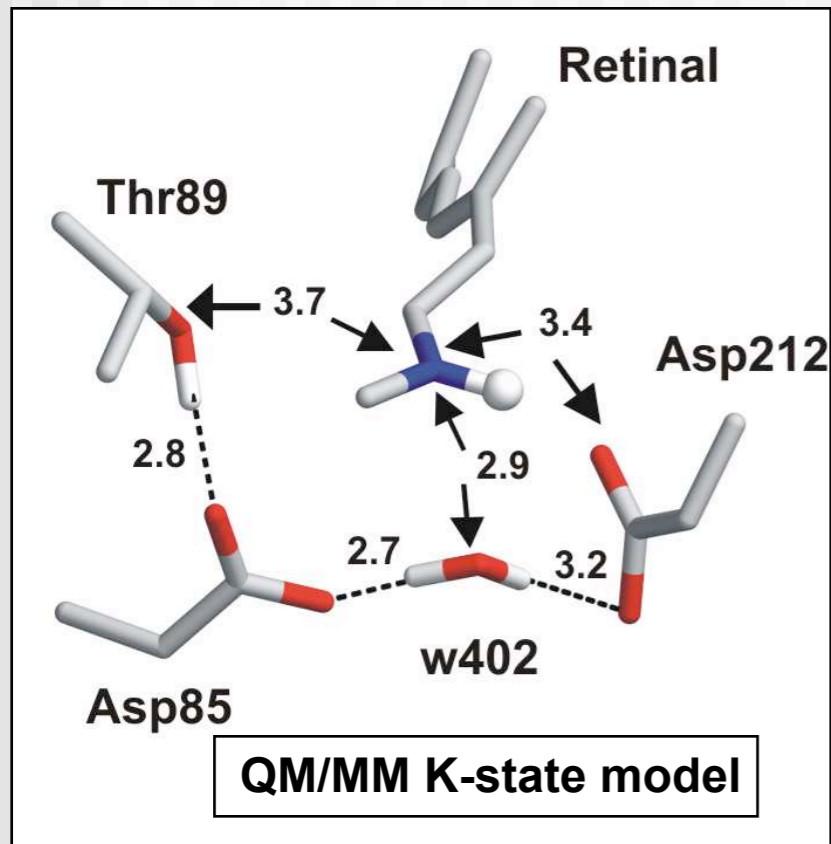
Why no back-proton transfer?
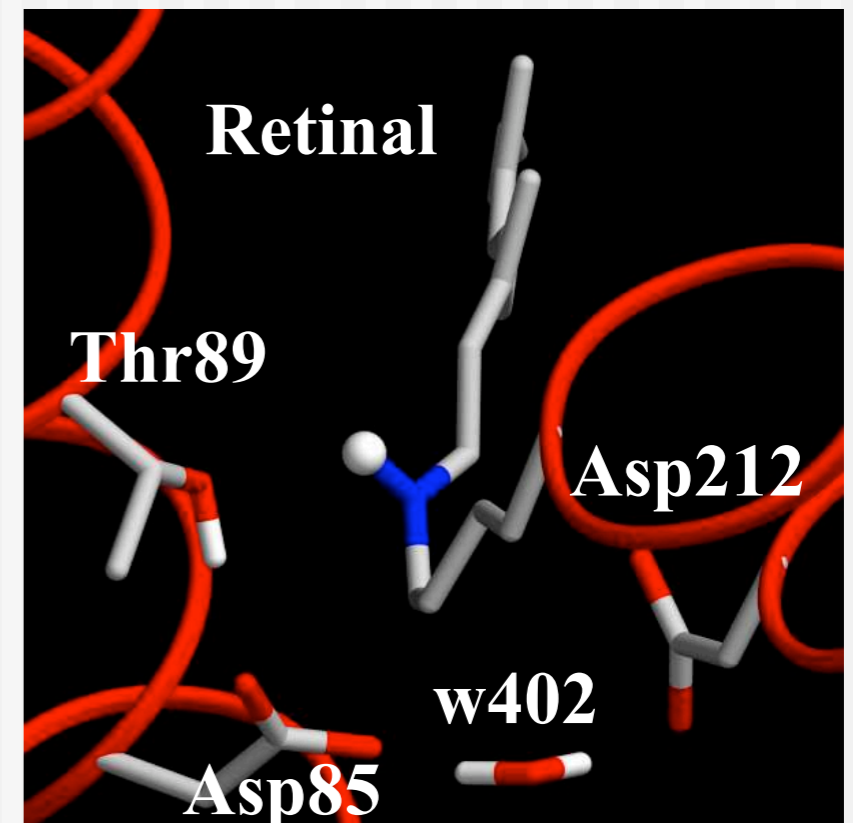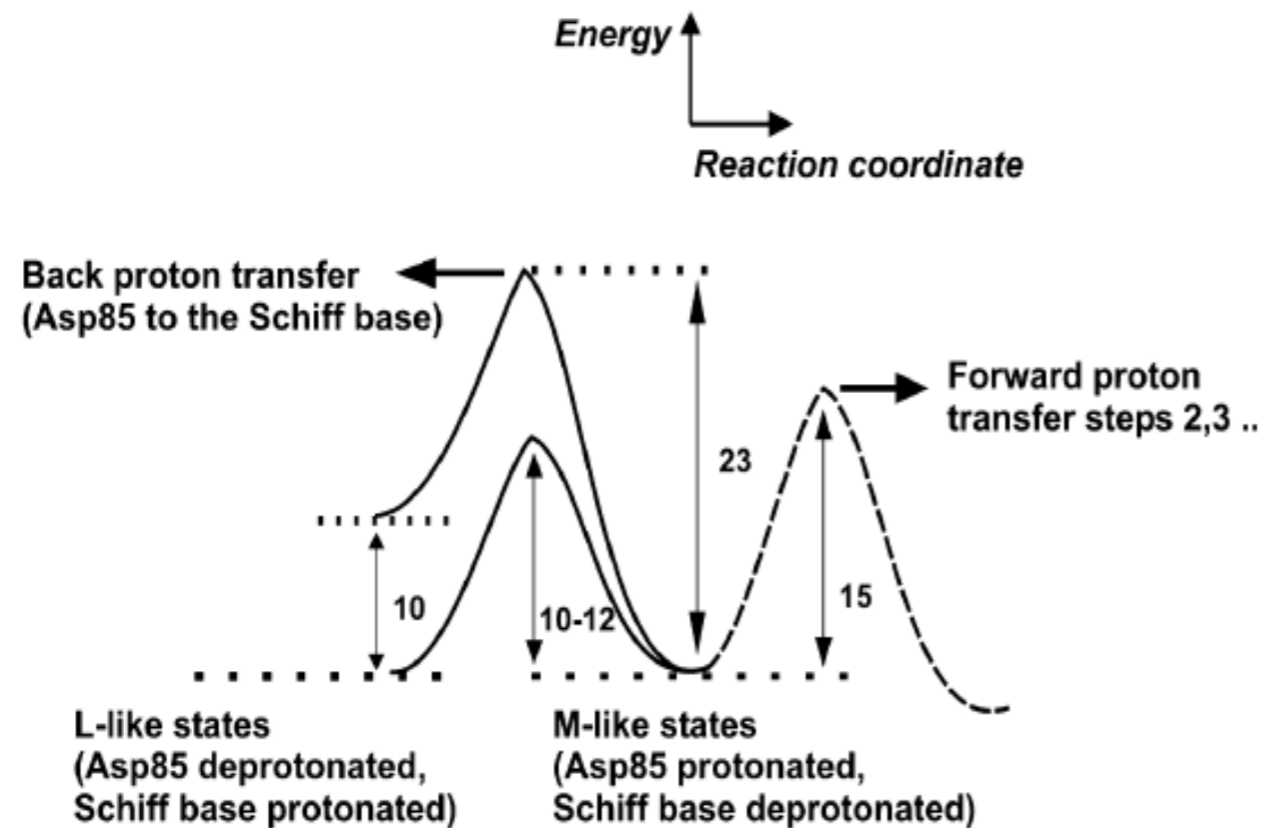
what causes the 'energy-drop' between M_1-M_2 intermediates?

➔ Changes in structure: early M vs. late M
- Thr89-Asp85 distance
- # waters
- deprotonation of 'proton release group' (next PT site)

Step 2: 'Proton release'
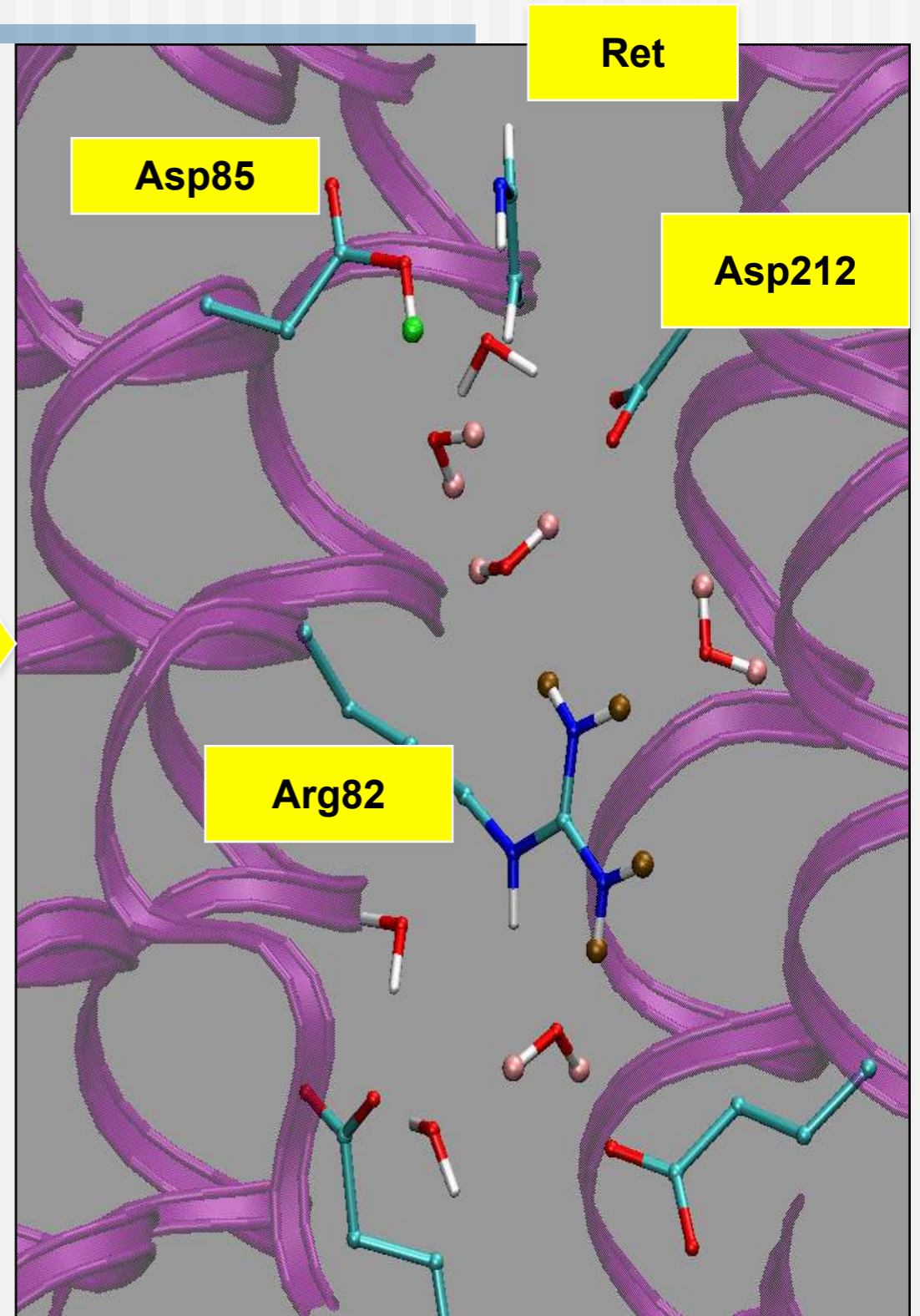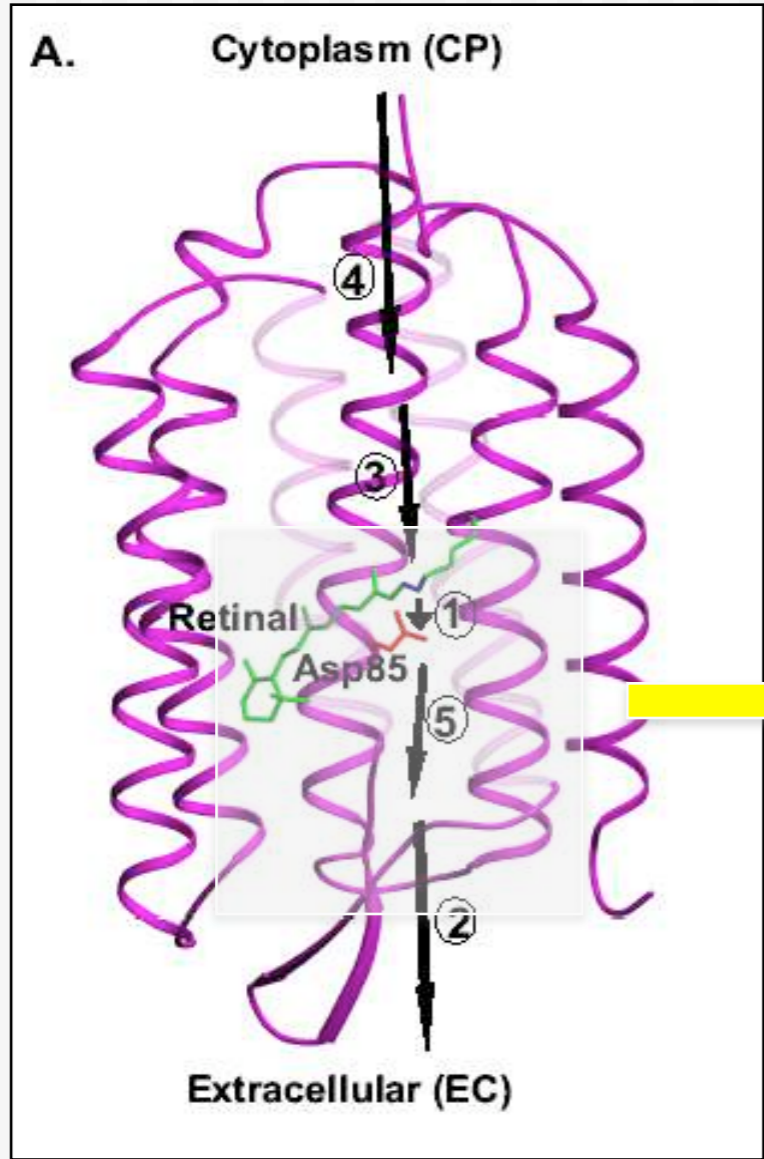
'Proton release'
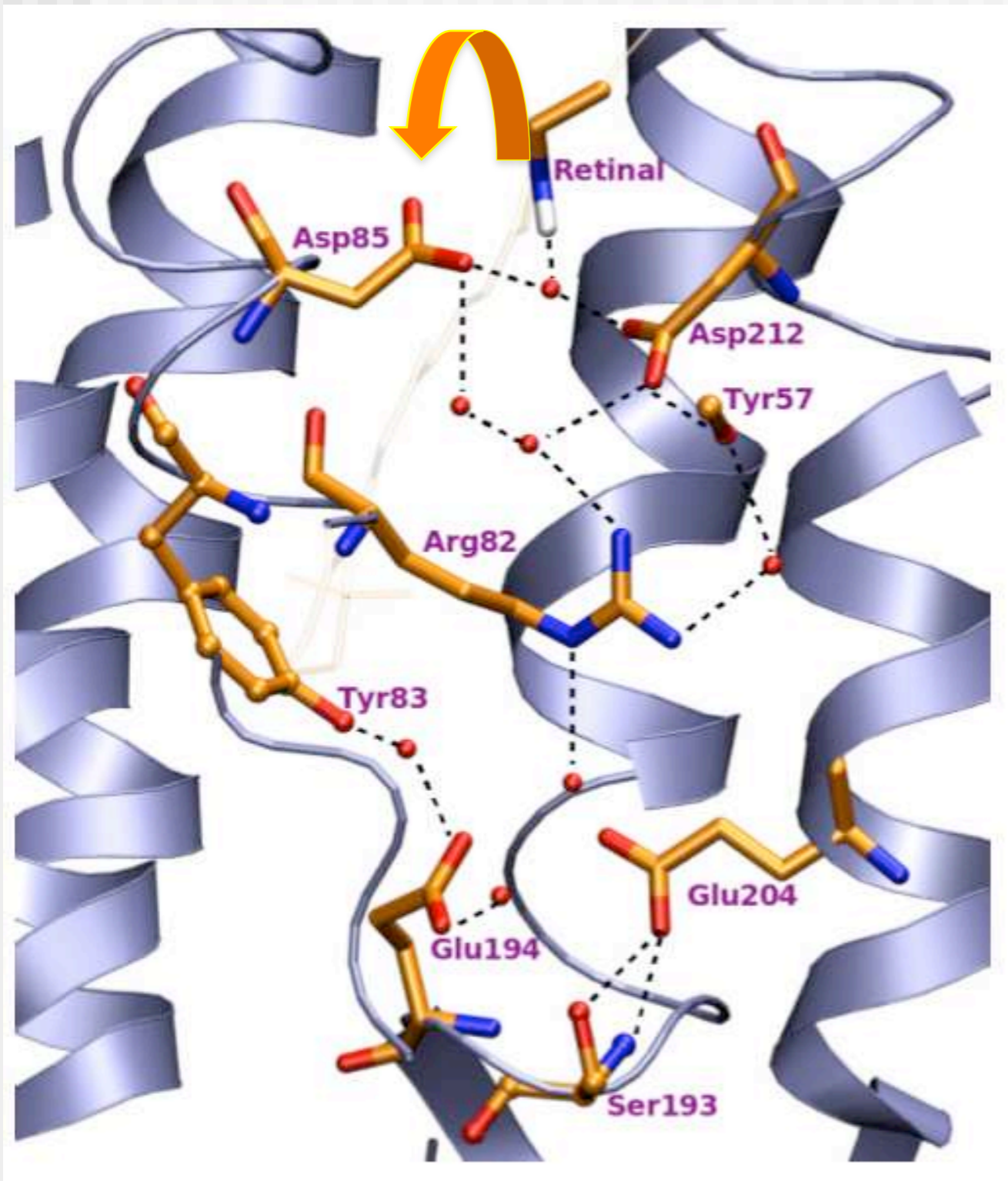
1) proton transfer
2) destabilization of H-bonded network
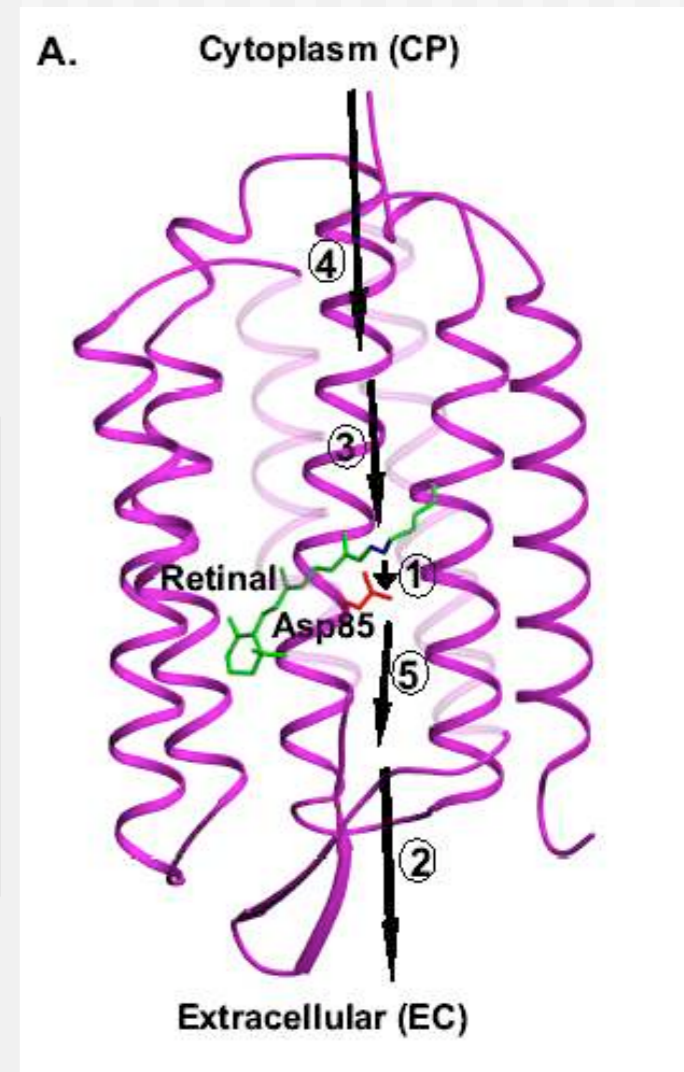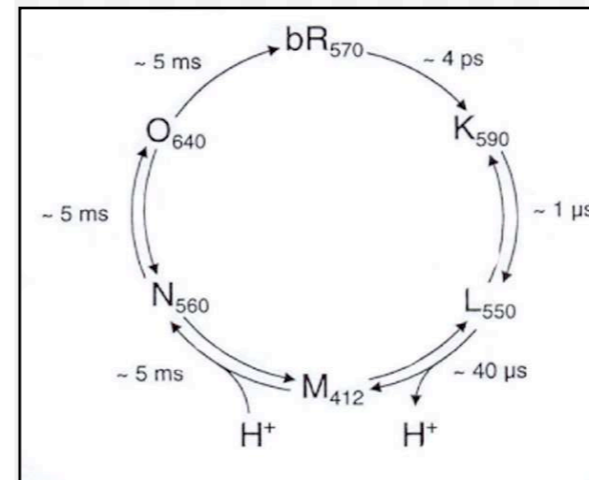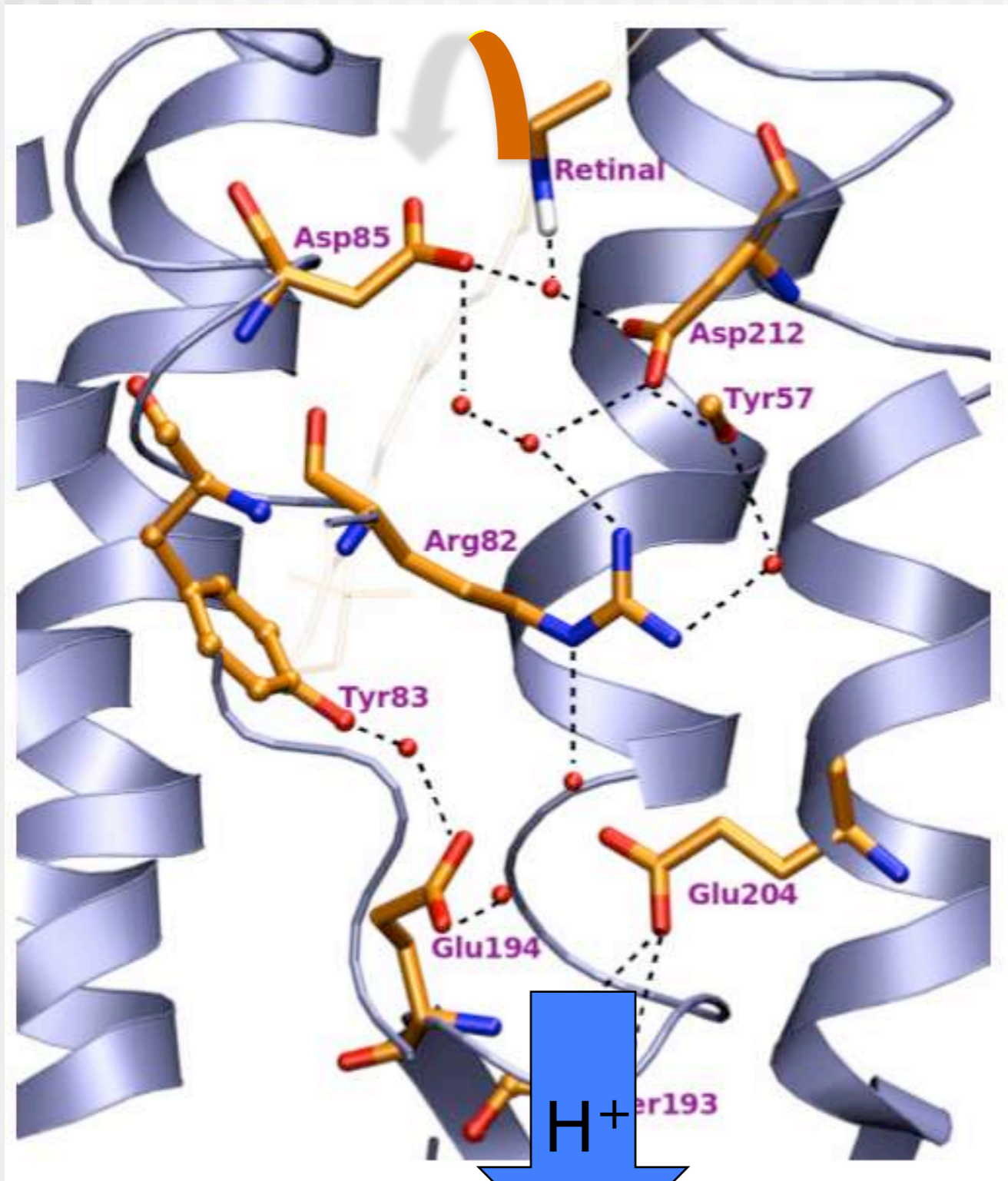
'Proton release'

1) proton transfer
2) destabilization of H-bonded network
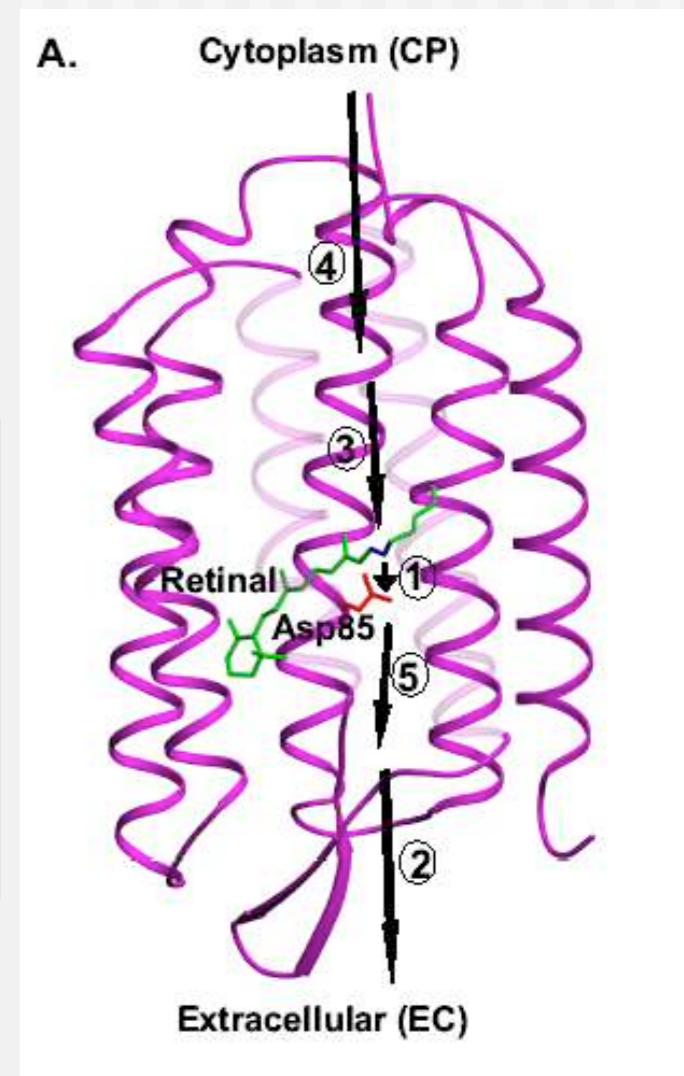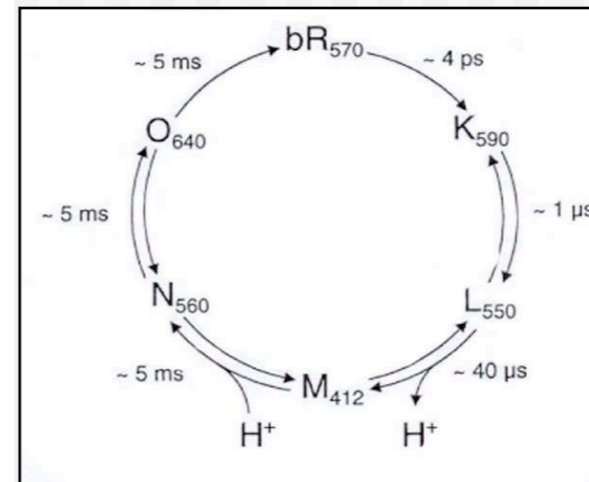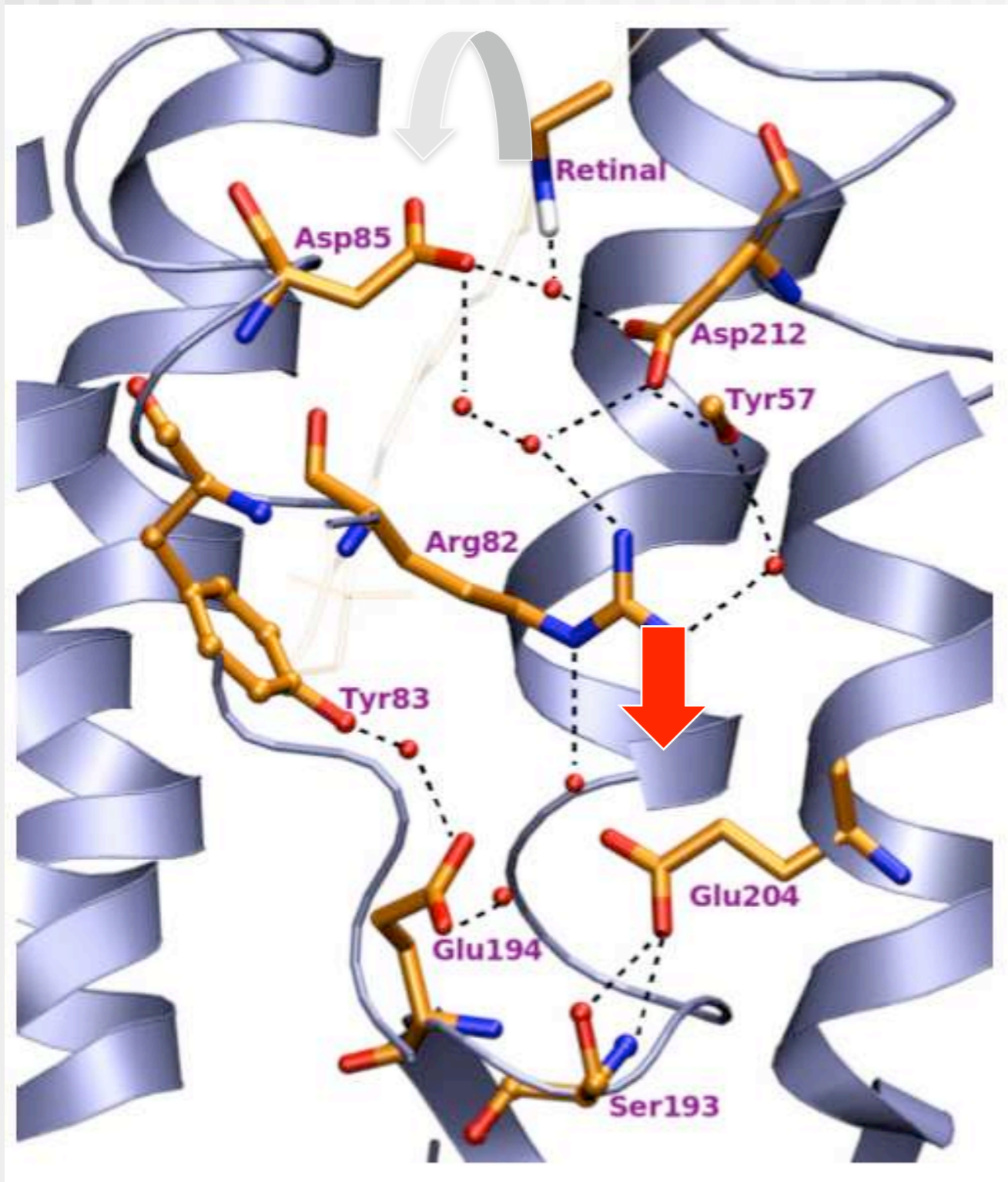
'Proton release'

1) proton transfer
2) destabilization of H-bonded-network
3) ➔ Arginin moves 'down

Question:
what comes first:
Arg movement or
proton relase?

➔ MD and free energy profiles

Arginine dynamics

1) proton transfer
2) destabilization of H-bonded-network
3) ➔ Arginin moves 'down

Arginine dynamics
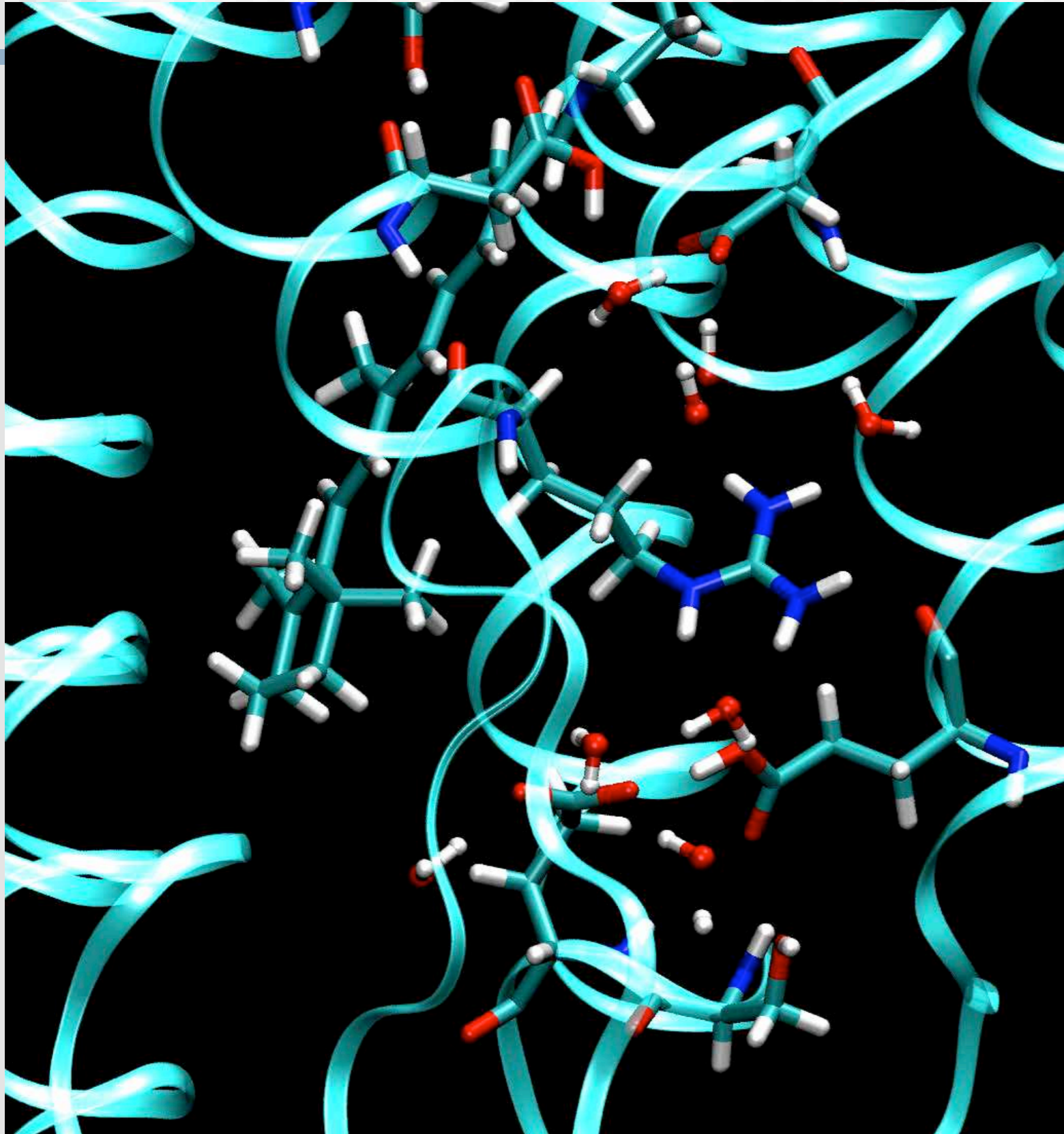
1) proton transfer
2) destabilization of H-bonded-network
3) → Arginin moves 'down

'Free energy profile'
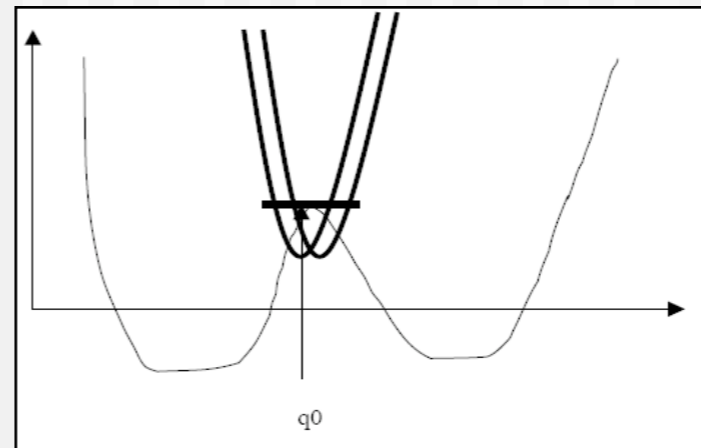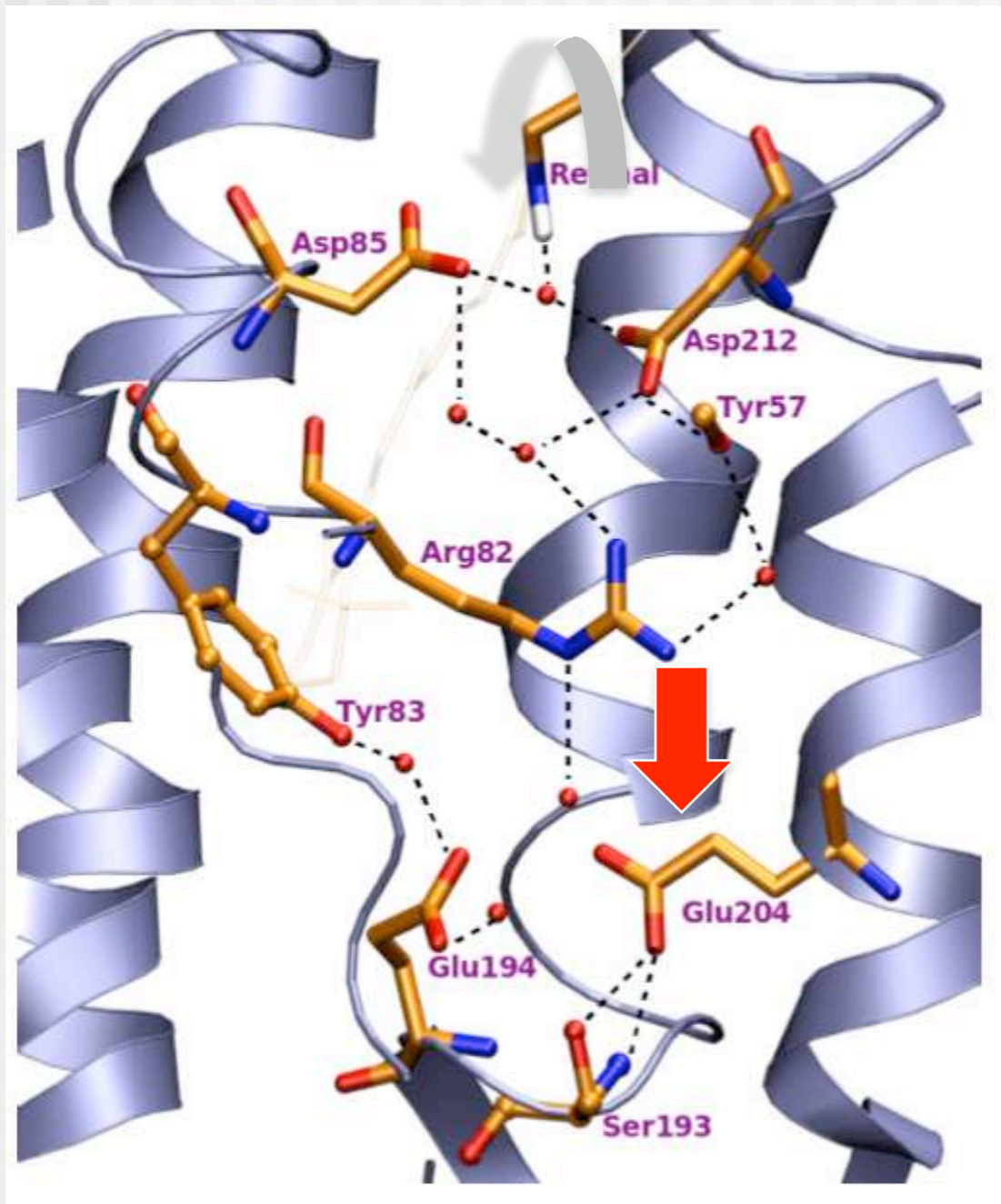
umbrella sampling:

$$F(q) = -kT \ln P(q)$$
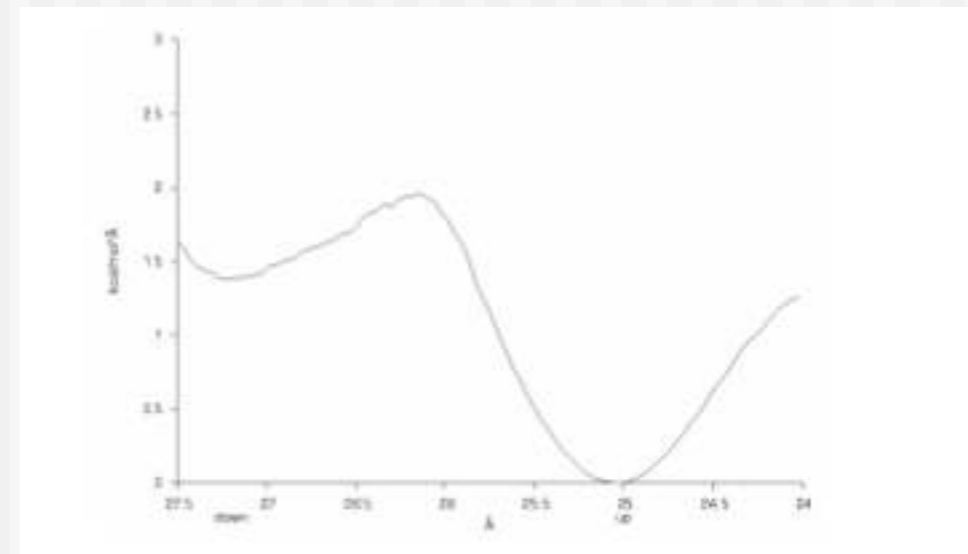
product   **reactant**

Clemens et al., to be published

'Proton release'
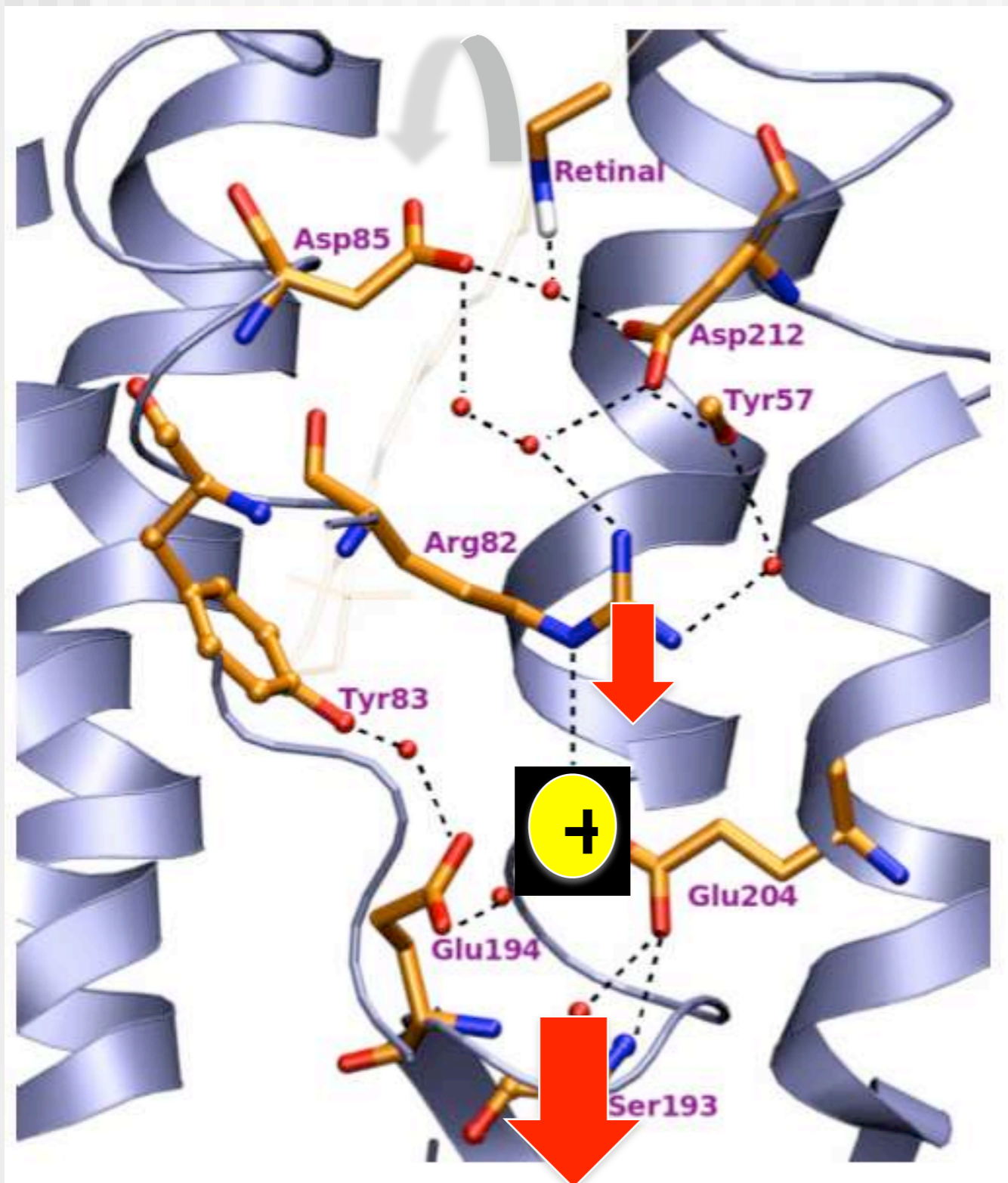
1) proton transfer
2) destabilization of H-bonded-network
3) ➔ Arginin moves 'down

- deprotonation:

  a) positive charge of Arg destabilzes proton at PRG
  b) Glu's are 'moved apart' ➔ change of pK_a

➔ pK_a calculations

'Proton release'
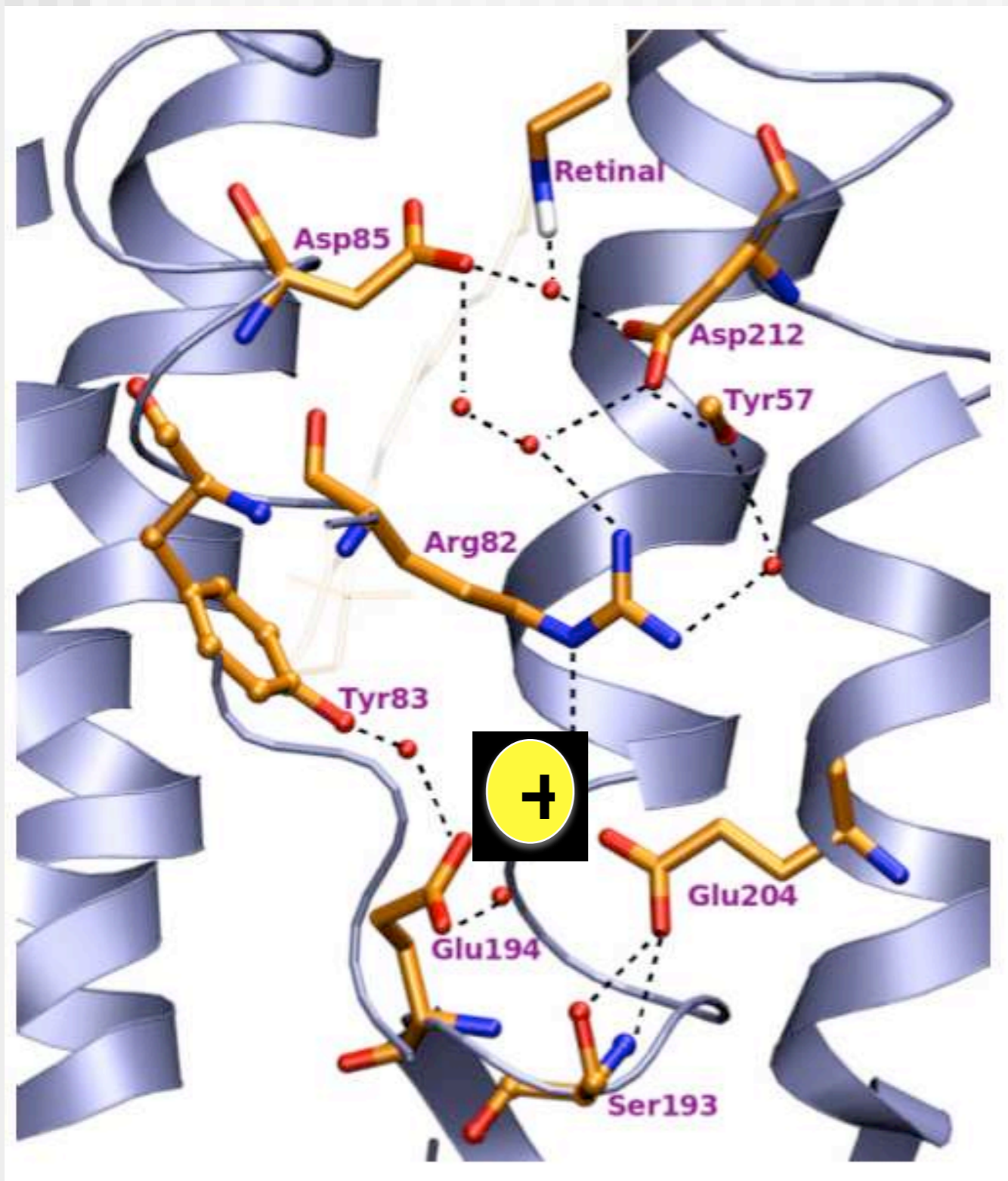
1) proton transfer
2) destabilization of H-bonded-network
3) ➔ Arginin moves 'down
4) => positive charge of Arg 'pushes' proton out.

unclear up to now:
What is the 'proton release group?

'Proton release group' (PRG): first suggestion one glu or even both glutamates?

- FTIR **difference spectra**: PRG is **not** Glu (194 or 204)!

- no signal from C=O band between 1700-1770 cm^{-1} detected after release!
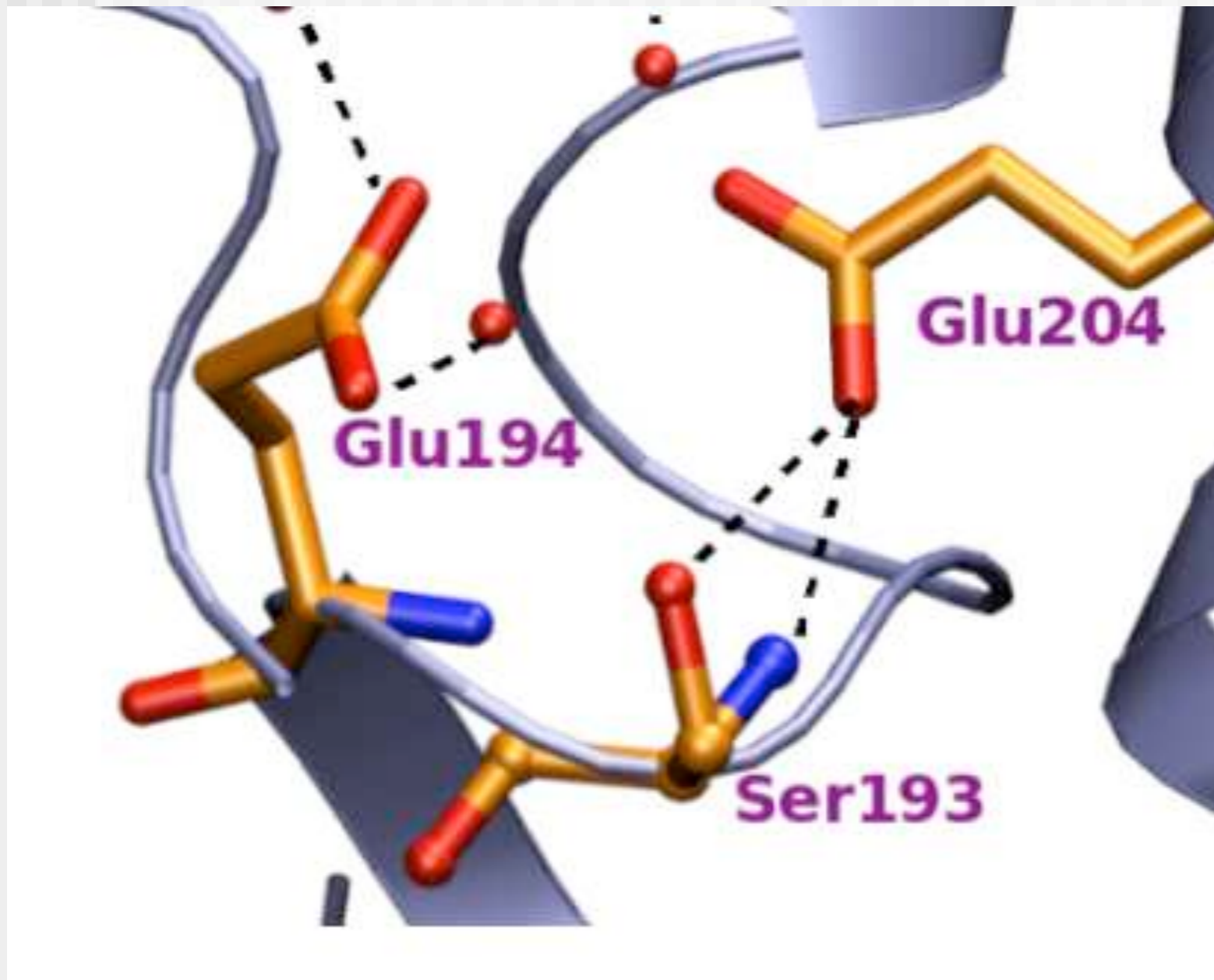
'Proton release group' (PRG): first suggestion one glu or even both glutamates?

- FTIR **difference spectra**: PRG is **not** Glu (194 or 204)!

- no signal from C=O band between 1700-1770 cm^{-1} detected after release!
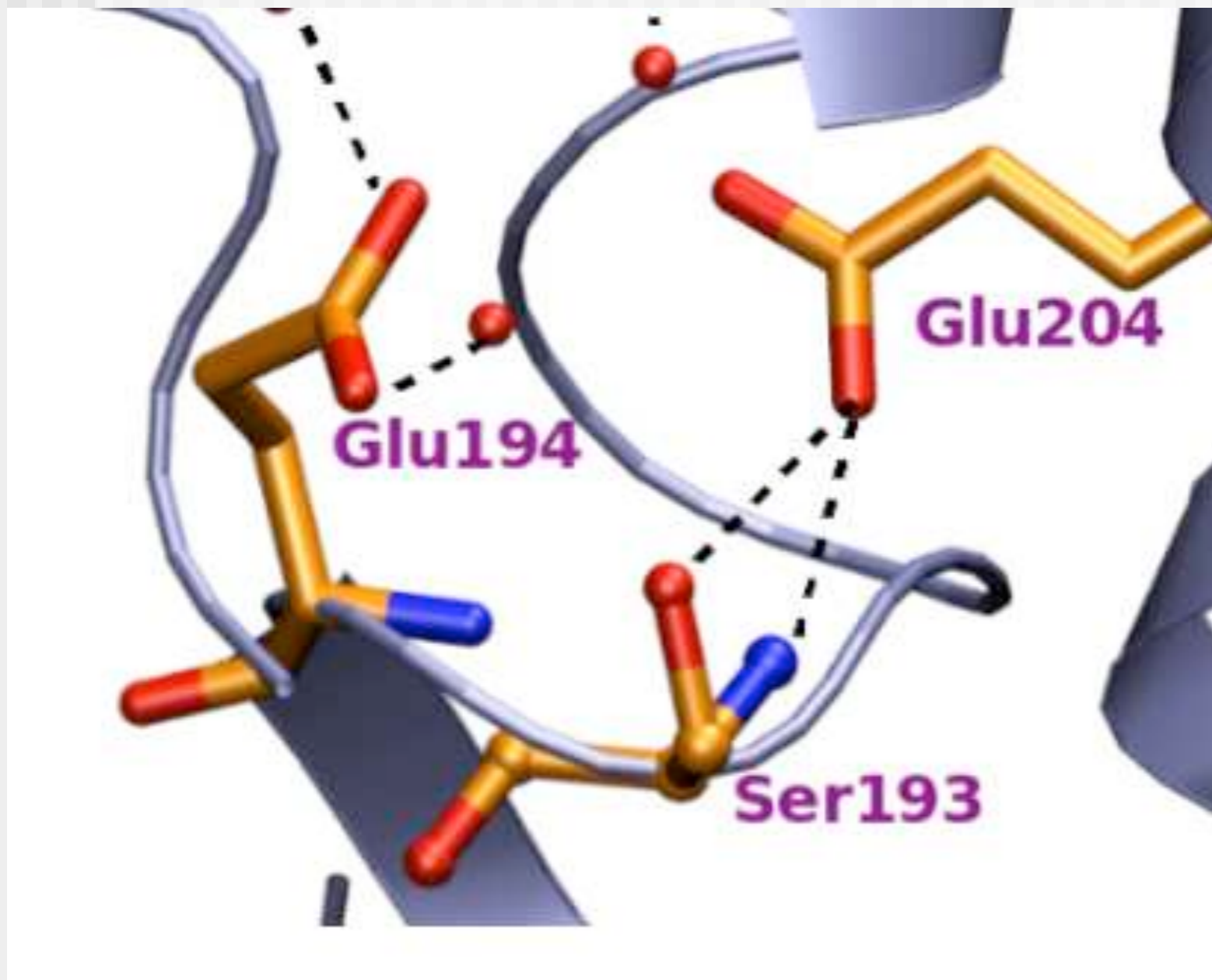
1700-1770 cm^{-1}

proton release

1350cm^{-1}
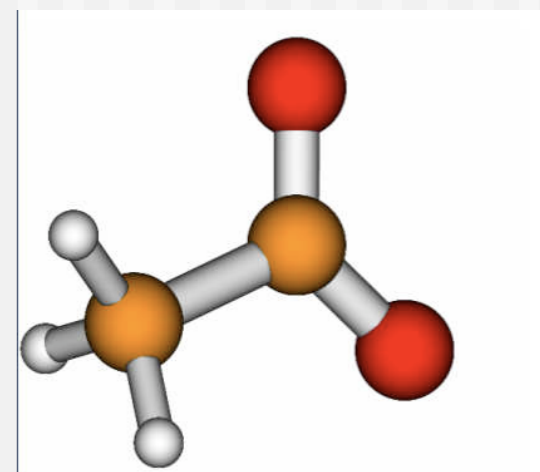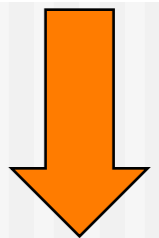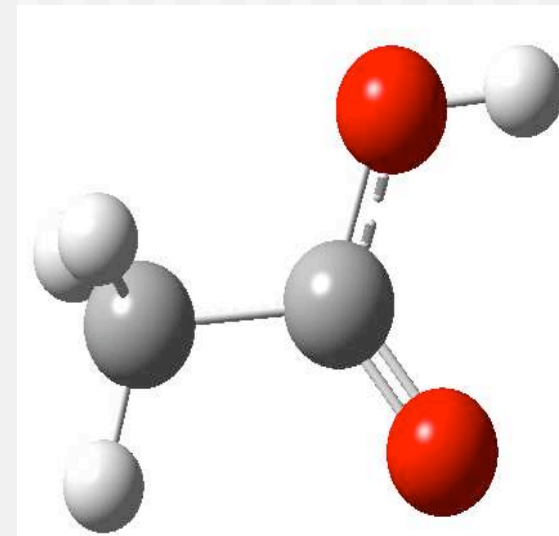
PRG: a protonated water cluster?

(Nature 439:109, PNAS 102:3633)

Arg82

Thr205

Tyr83

Glu204

Glu194

Ser193
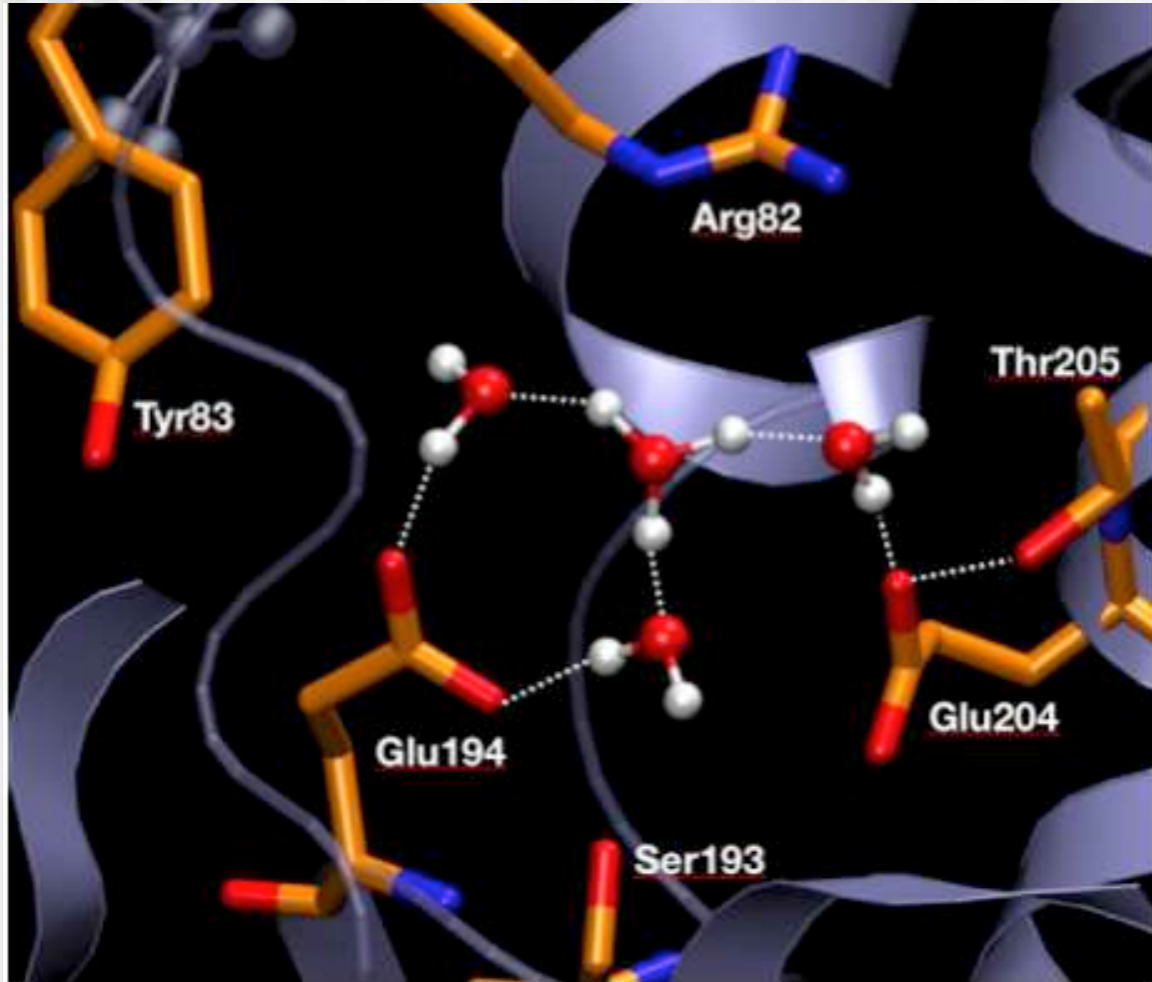
- FT-IR: continuum band between 1700 cm^{-1} and 2100 cm^{-1} measured

- characteristic for Zundel-complex

Zundel Ion
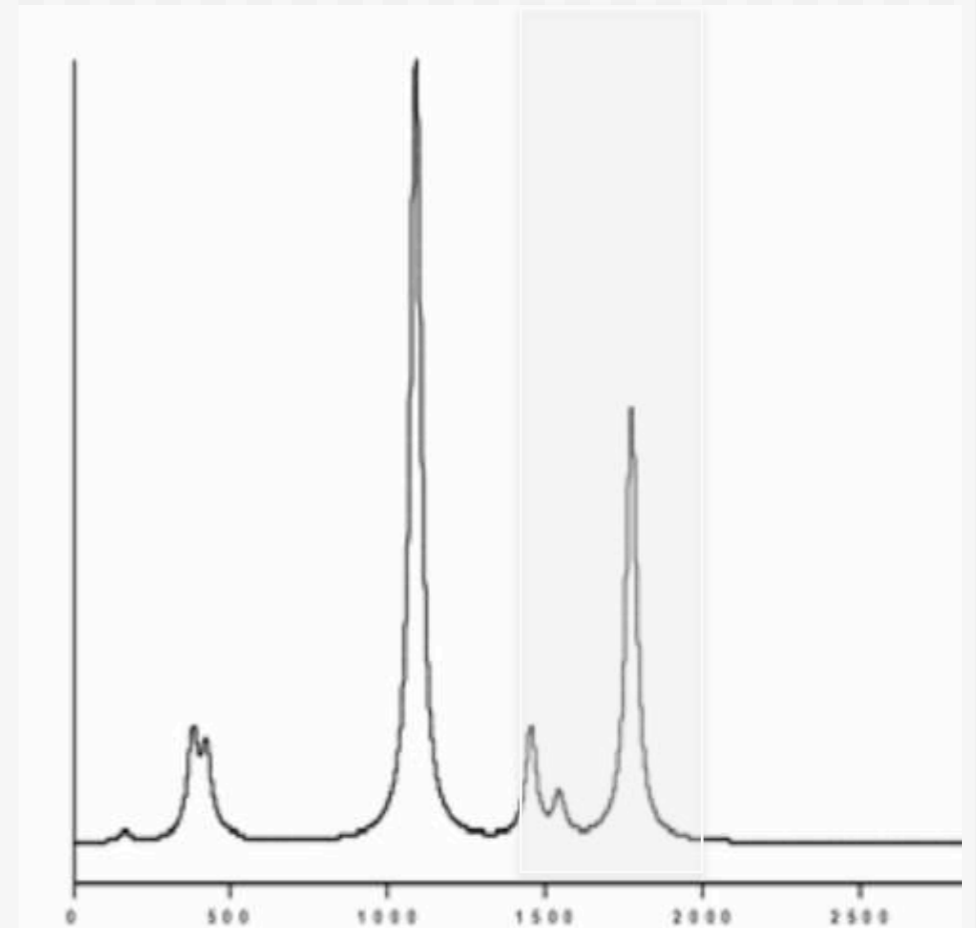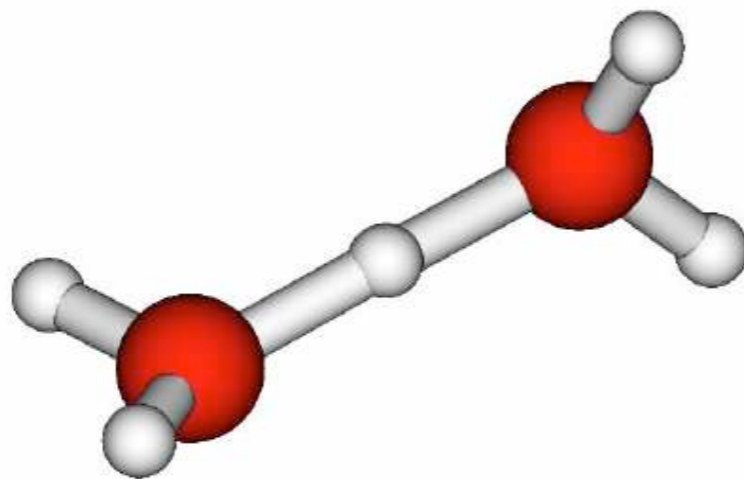
Computational support for protonated cluster
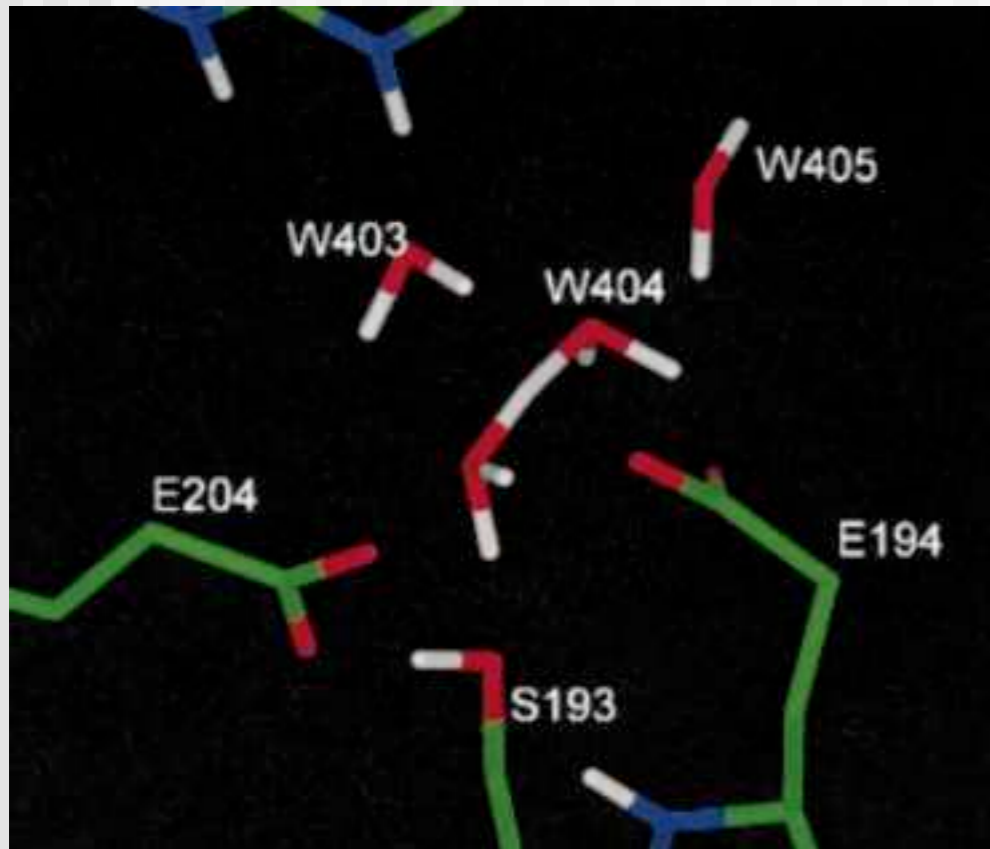
electrostatic calculations

- proton may sit on
  a water cluster

Spassov et al., J. Mol. Biol. (2001) 312, 203
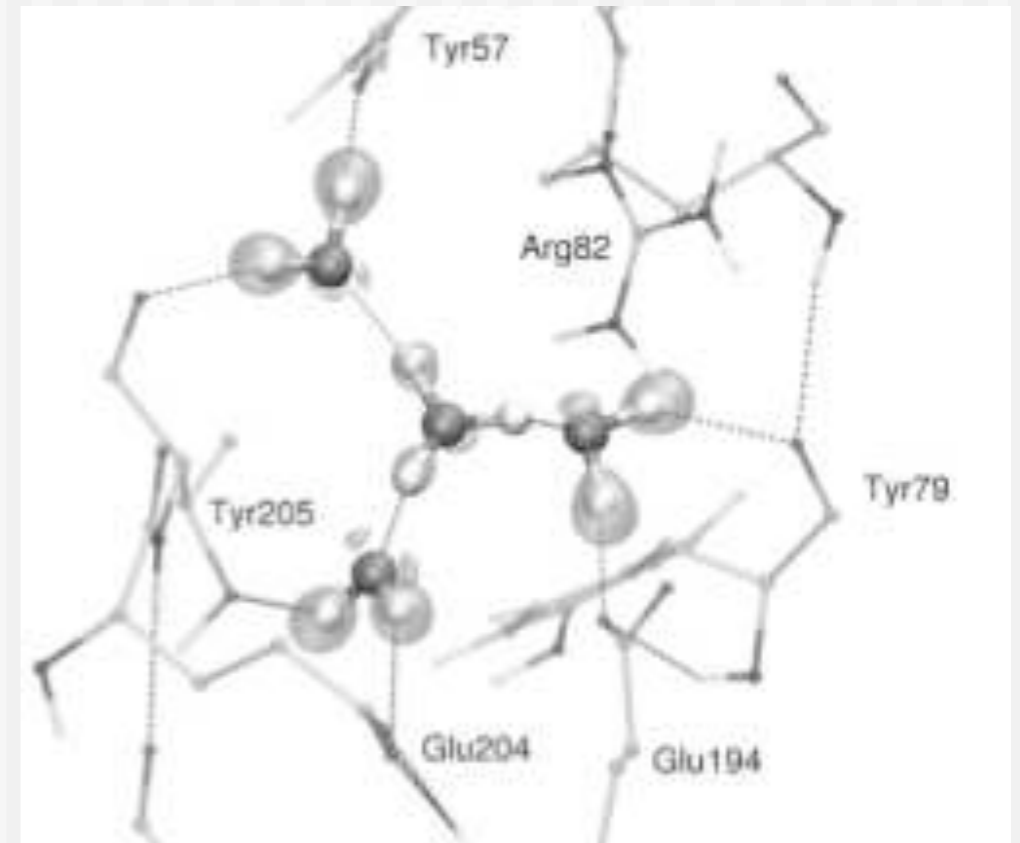
QM/MM CPMD

- continuum band reported
- constraints on cluster
- Glu's not QM

Rousseau et al., Ang. 2004, 116,4908

QM/MM MD calculations: include the Glu's in QM
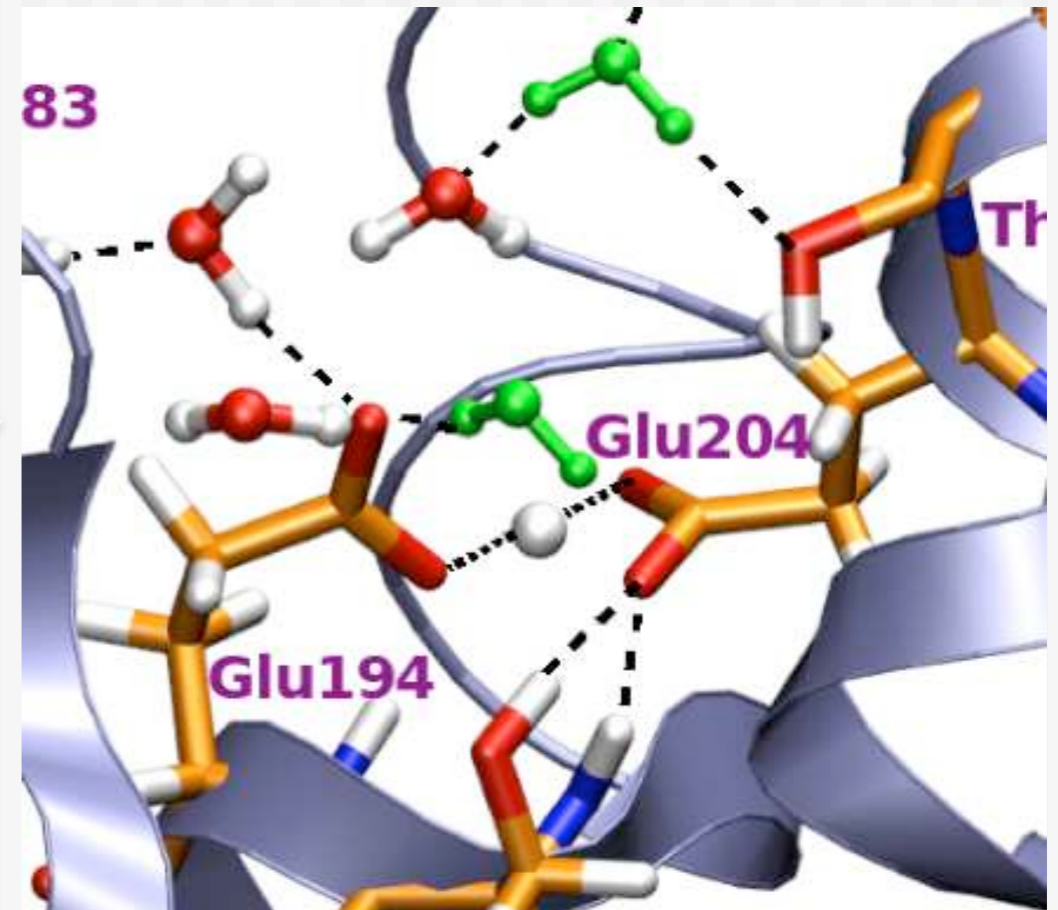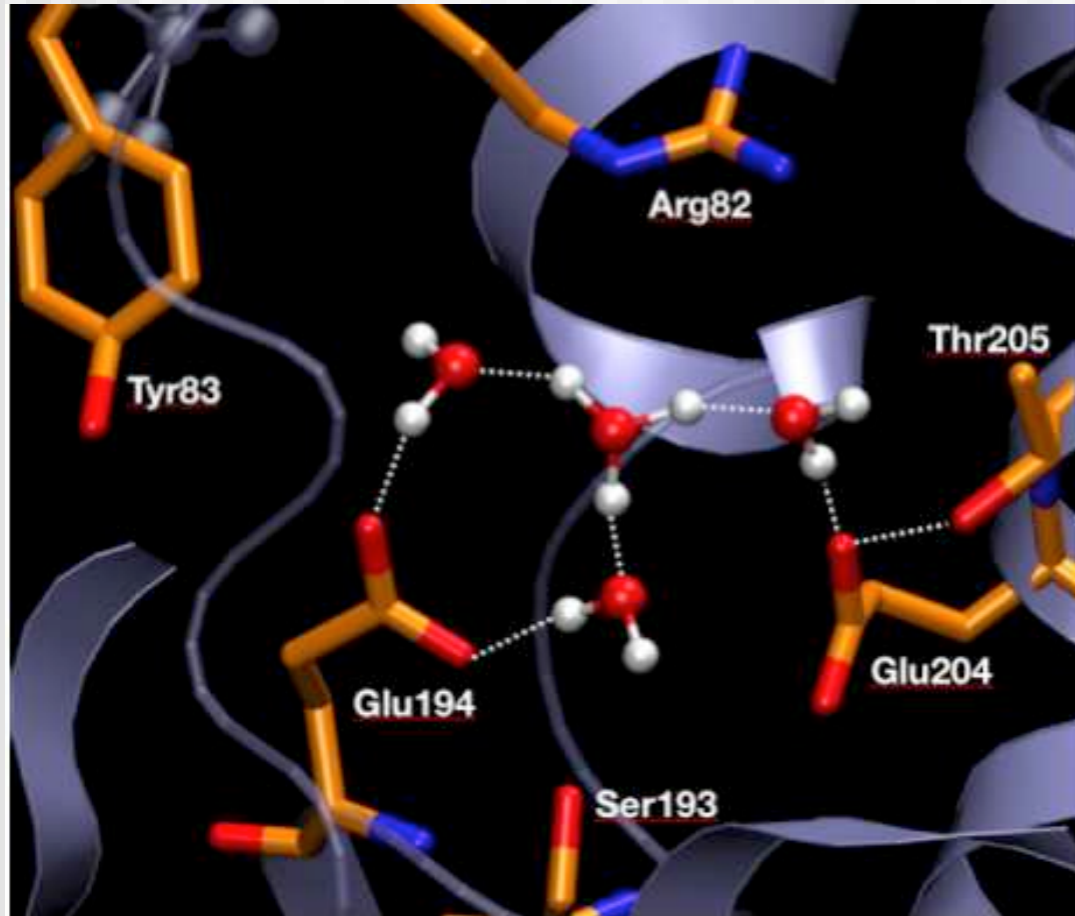
method: QM/MM/GSBP

DFT-B3LYP/CHARMM/GSBP 2ps dynamics

SCC-DFTB/CHARMM/GSBP 2ns dynamics (for IR spectra)

• proton leaves water cluster within 1ps

• stable conformation between Glu194 and Glu204

Phatak et al., PNAS 105, 19672

QM/MM MD calculations
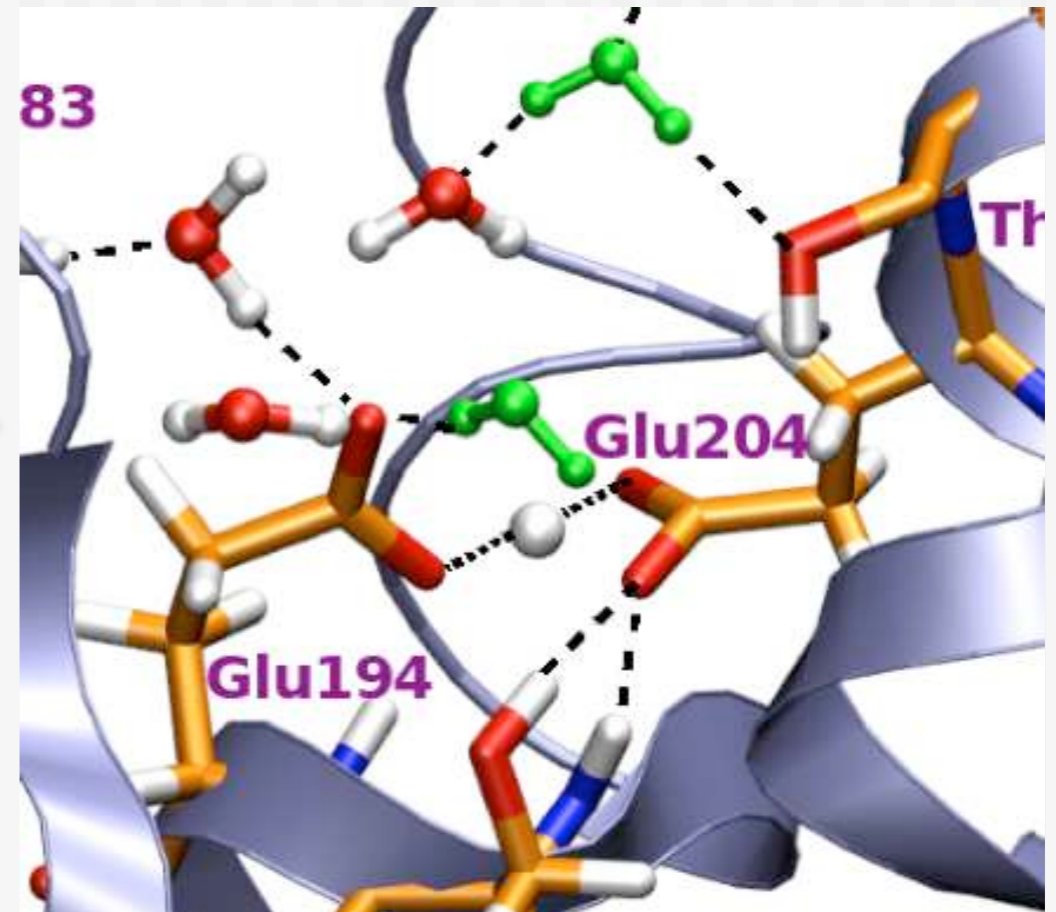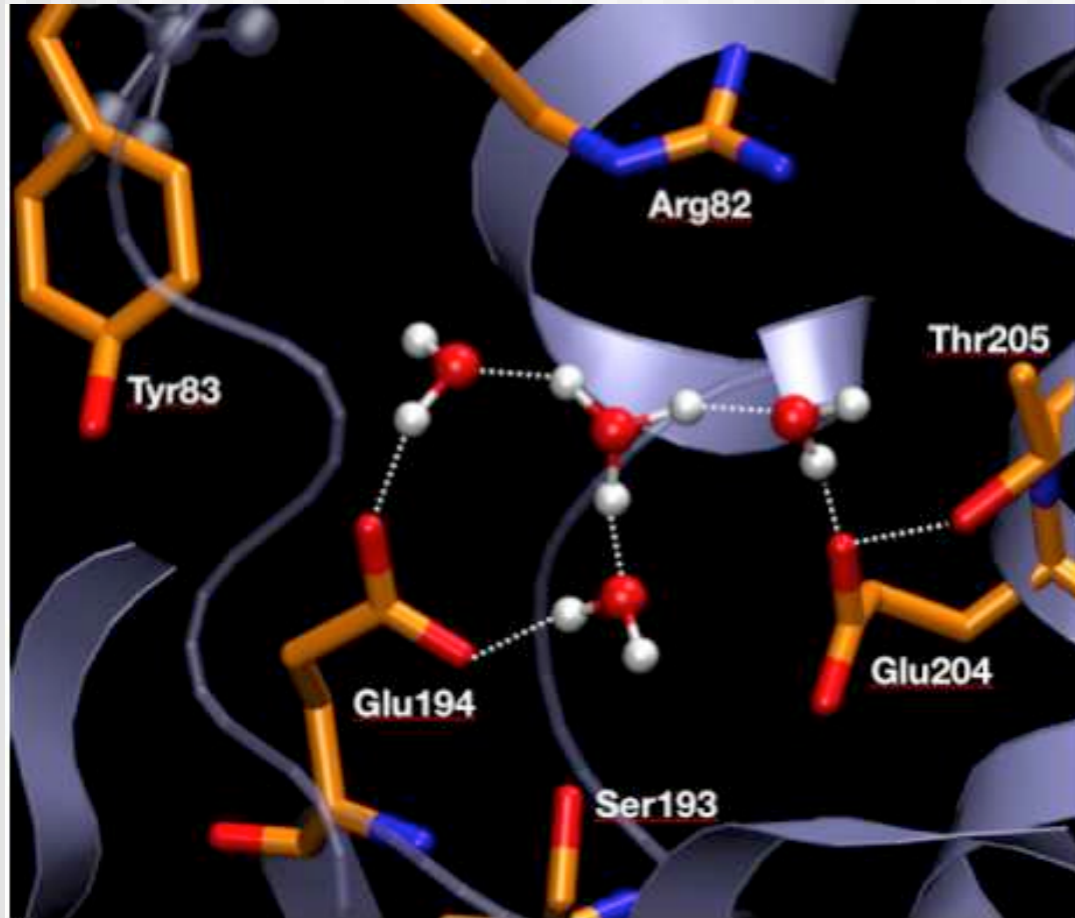
➔3 problems/questions:

1. Is this compatible with crystal structures?

2. can we explain the missing C=O signature in IR spectra?

3. can we explain the continuum modes?

Phatak et al., PNAS 105, 19672

1) crystal structures
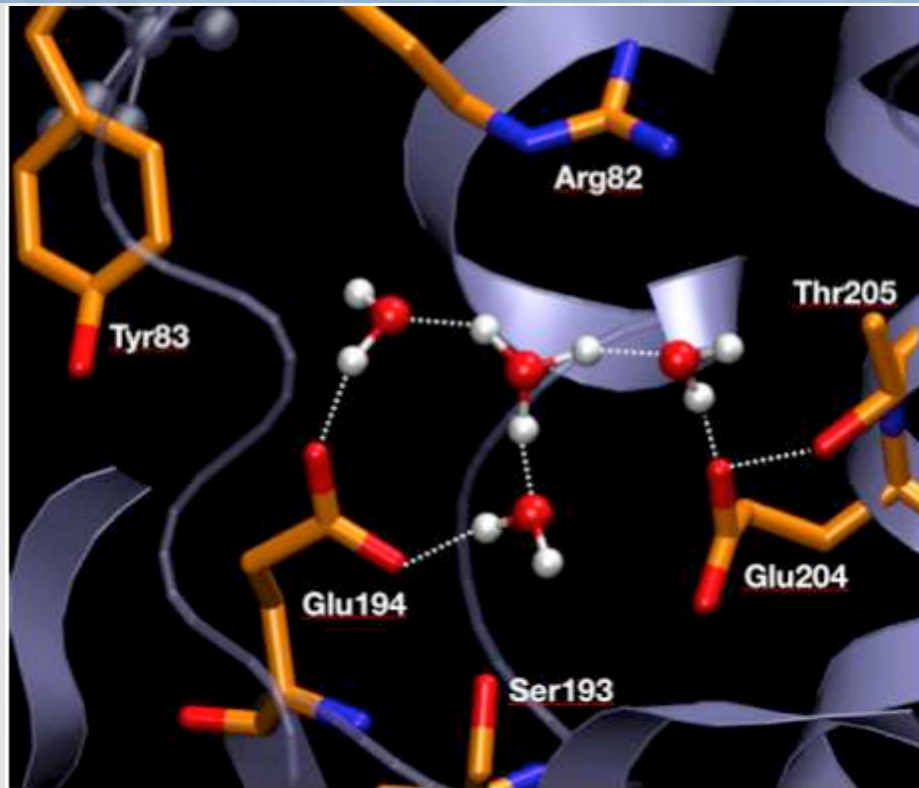
this model is **NOT** consistent with X–ray data

O–O distance too large: R=4.5 A

State	PDB code (Res.[b])	Glu194(O^ϵ)- Glu204(O^ϵ) distance (Å)
Ground	1C3W (1.55 Å) (25)	3.0
Ground	1QHJ (1.90 Å) (40)	2.4
Ground	1KGB (1.65 Å) (41)	2.5
Ground	1IW6 (2.30 Å) (42)	2.5
Ground	1C8R (1.80 Å) (43)	2.6
L	1E0P (2.10 Å) (44)	3.1
L	1O0A (1.62 Å) (27)	3.1
L	1UCQ (2.40 Å) (45)	2.6
L	1VJM (2.30 Å) (46)	2.4
L	2NTW (1.53 Å) (26)	4.2
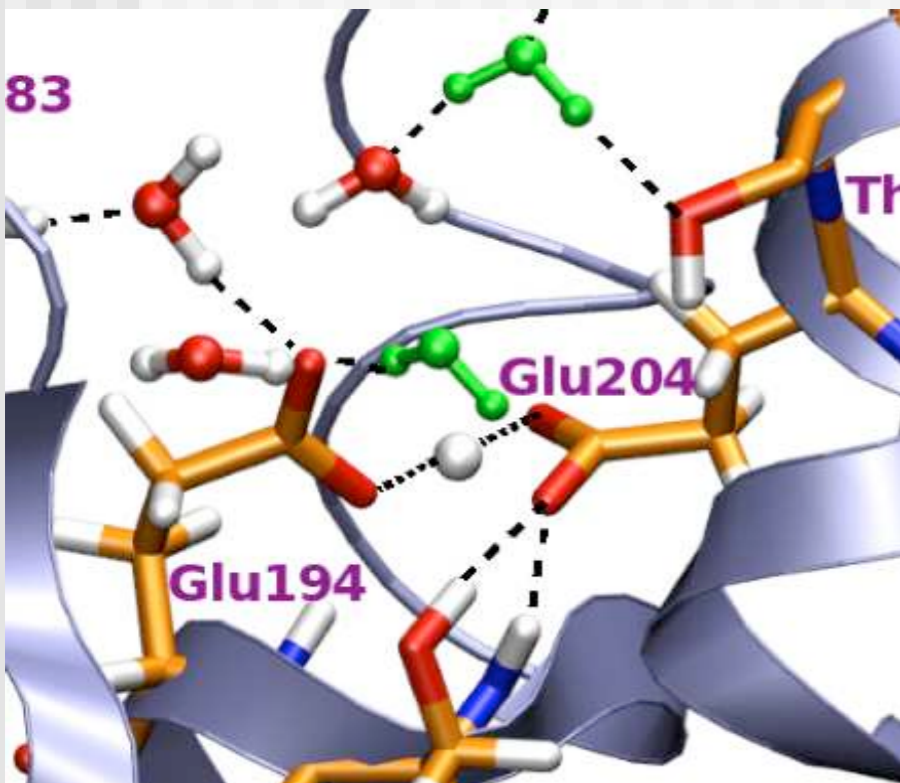
this one fits: O-O distance R=2.5 A

small distance only possible due to shared proton!
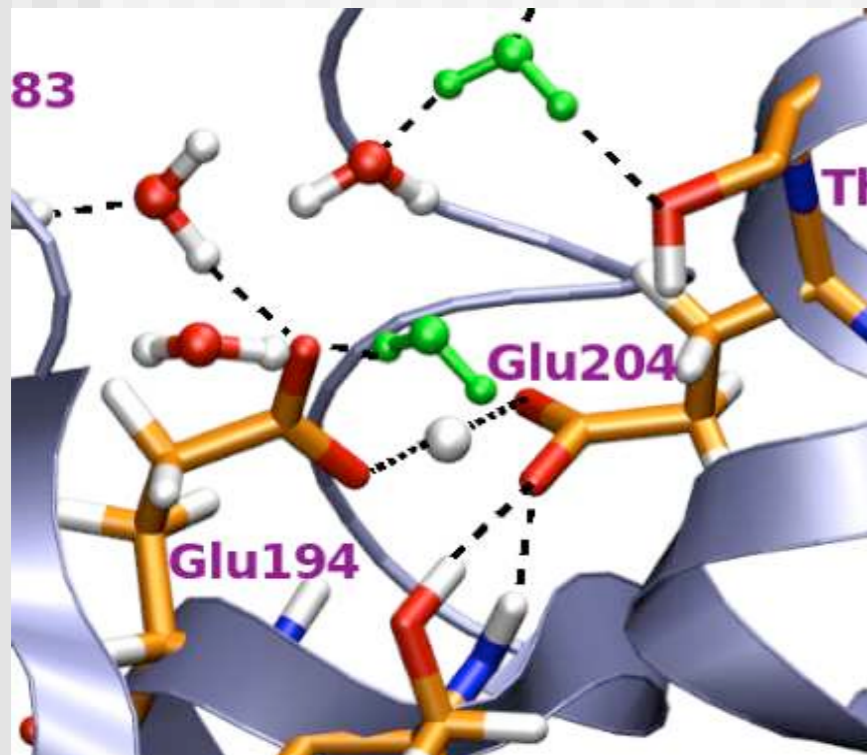
Phatak et al., PNAS 105, 19672

2) Continuum modes

shared proton leads to
continuum mode

IR spectra with only one Glu in
QM region: continuum missing

Phatak et al., PNAS 105, 19672

continuum mode

2) Continuum modes

83

Glu194

Glu

shared proton leads to continuum mode

IR spectra with only one Glu in QM region: continuum missing
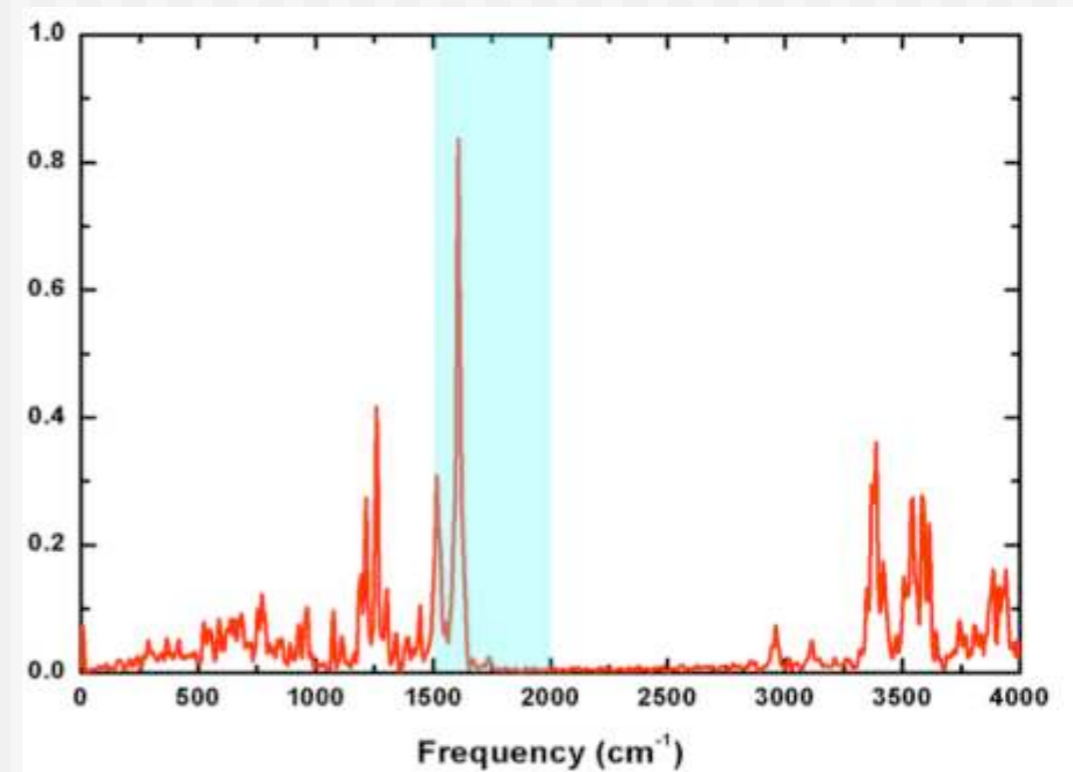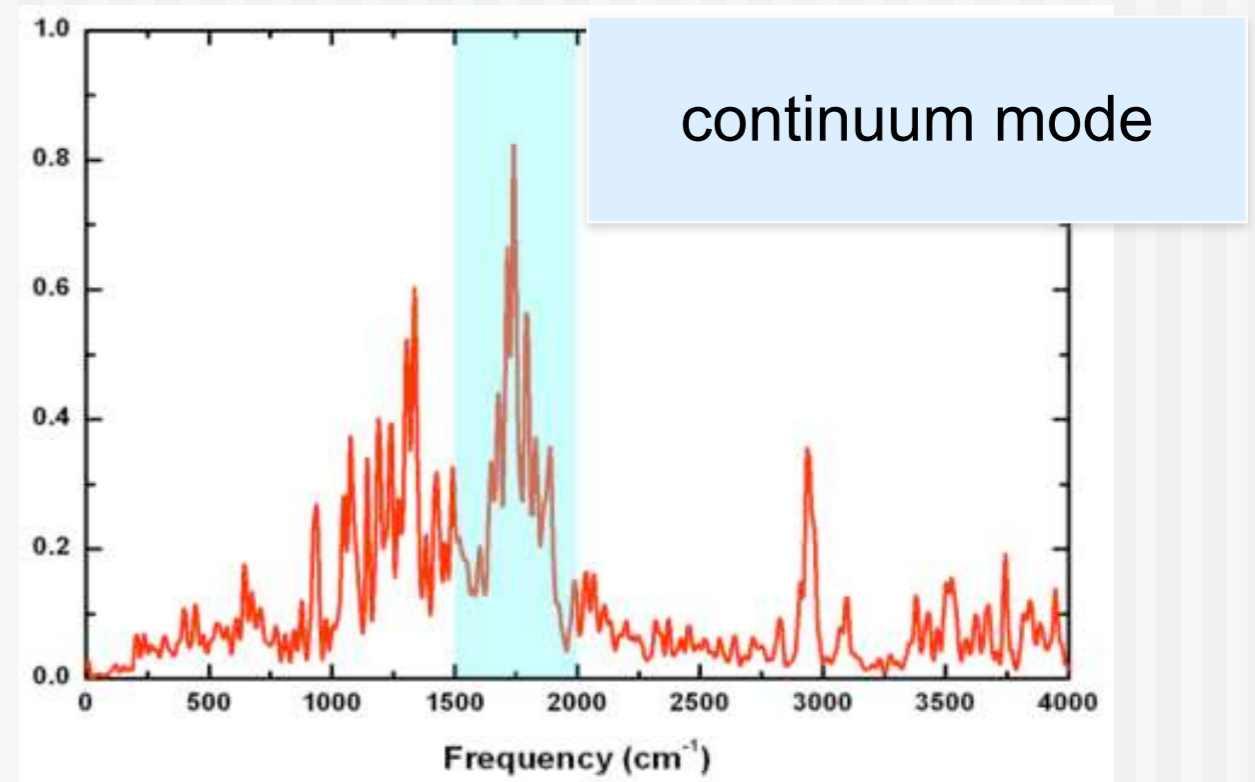
Phatak et al., PNAS 105, 19672
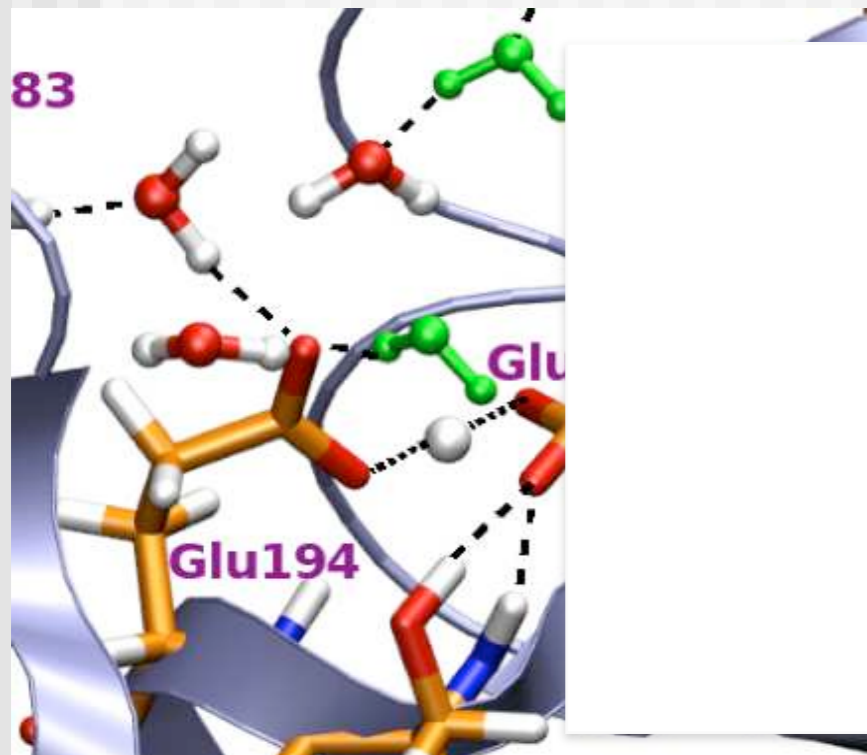
Further: Several mutants calculated
→good agreement with experiment

Frequency (cm^{-1})

Frequency (cm^{-1})
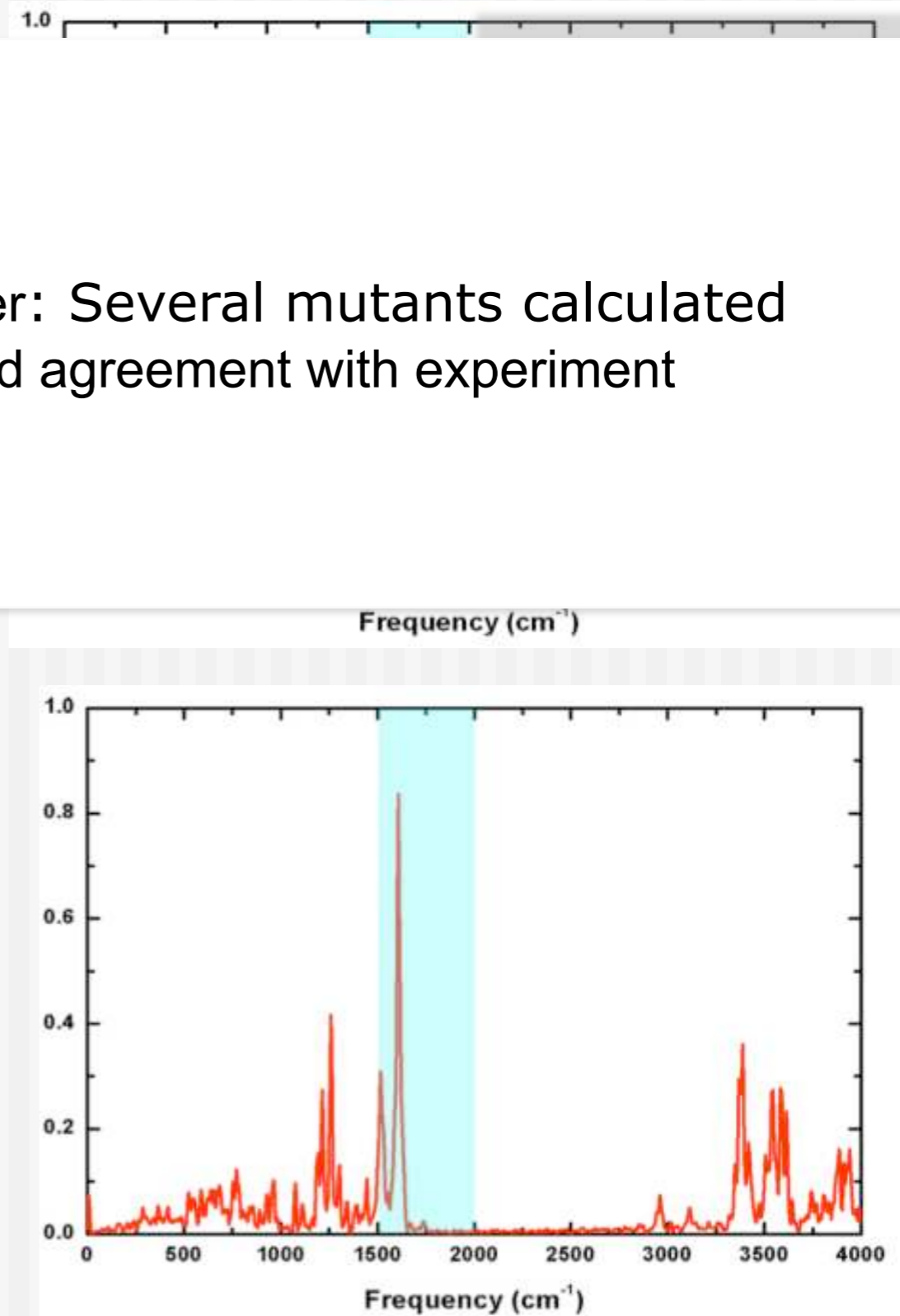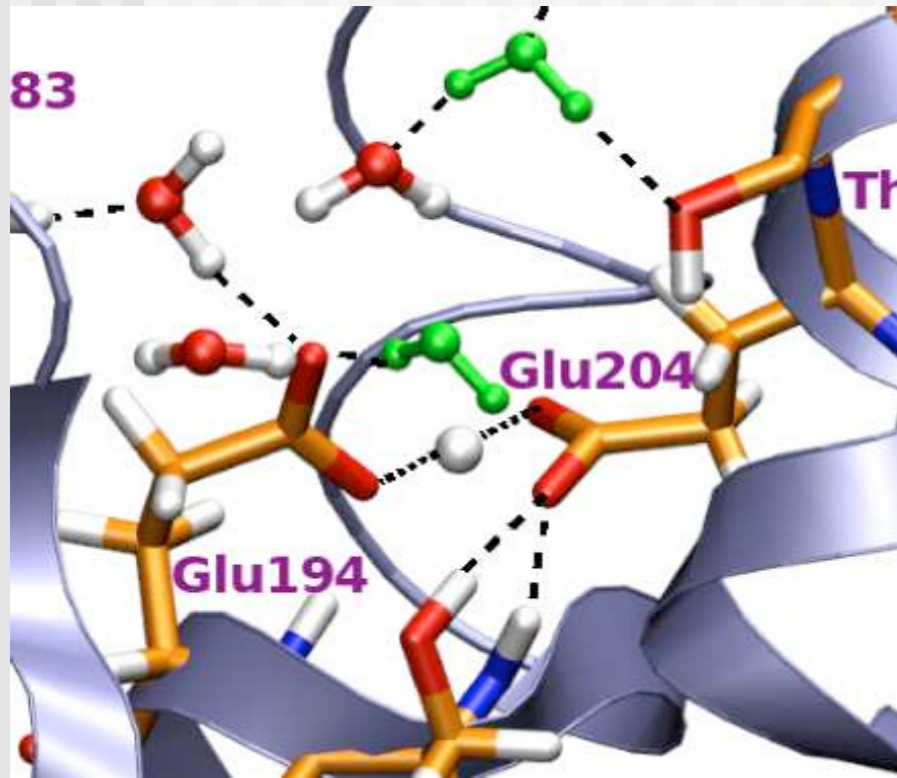
3) C=O stretch
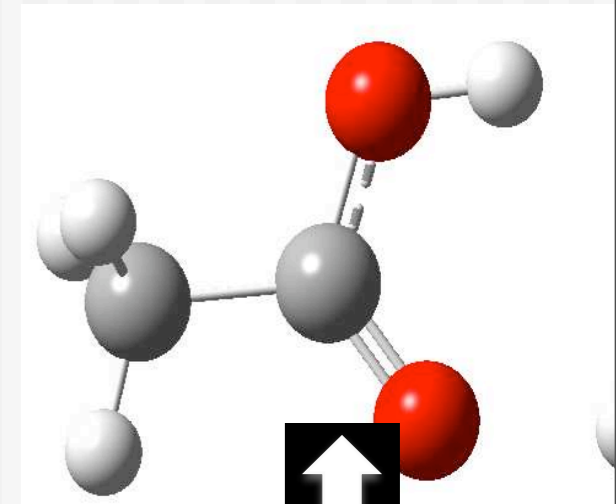
what about C=O?

1700-1750 cm^{-1}

3) C=O stretch
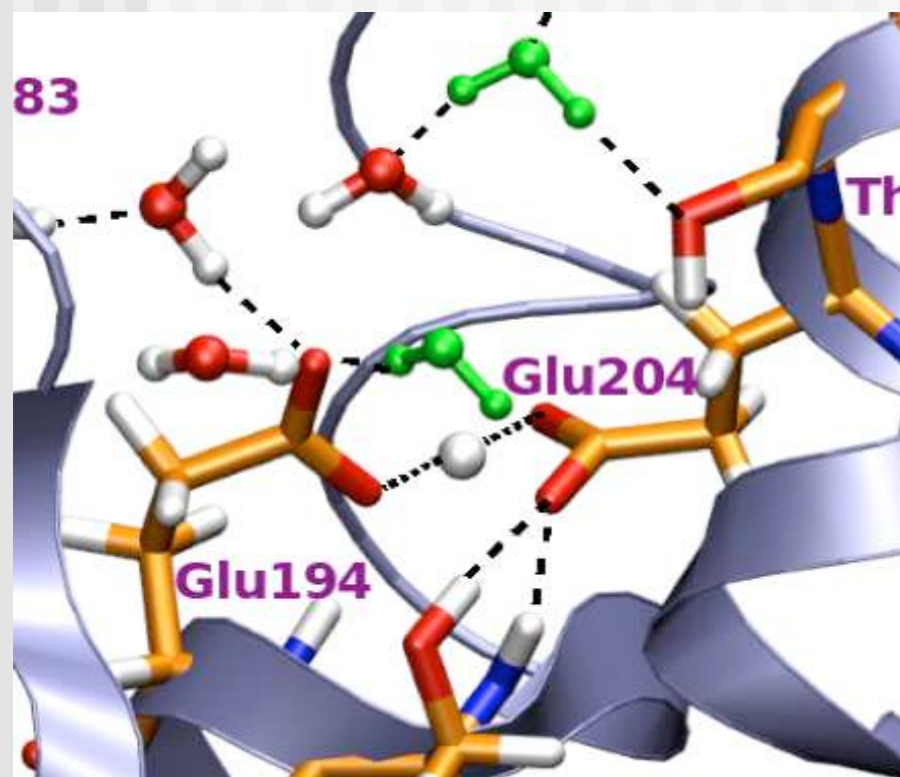
what about C=O?
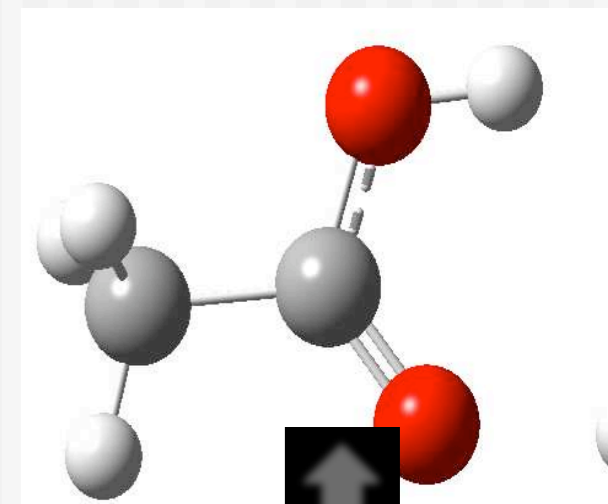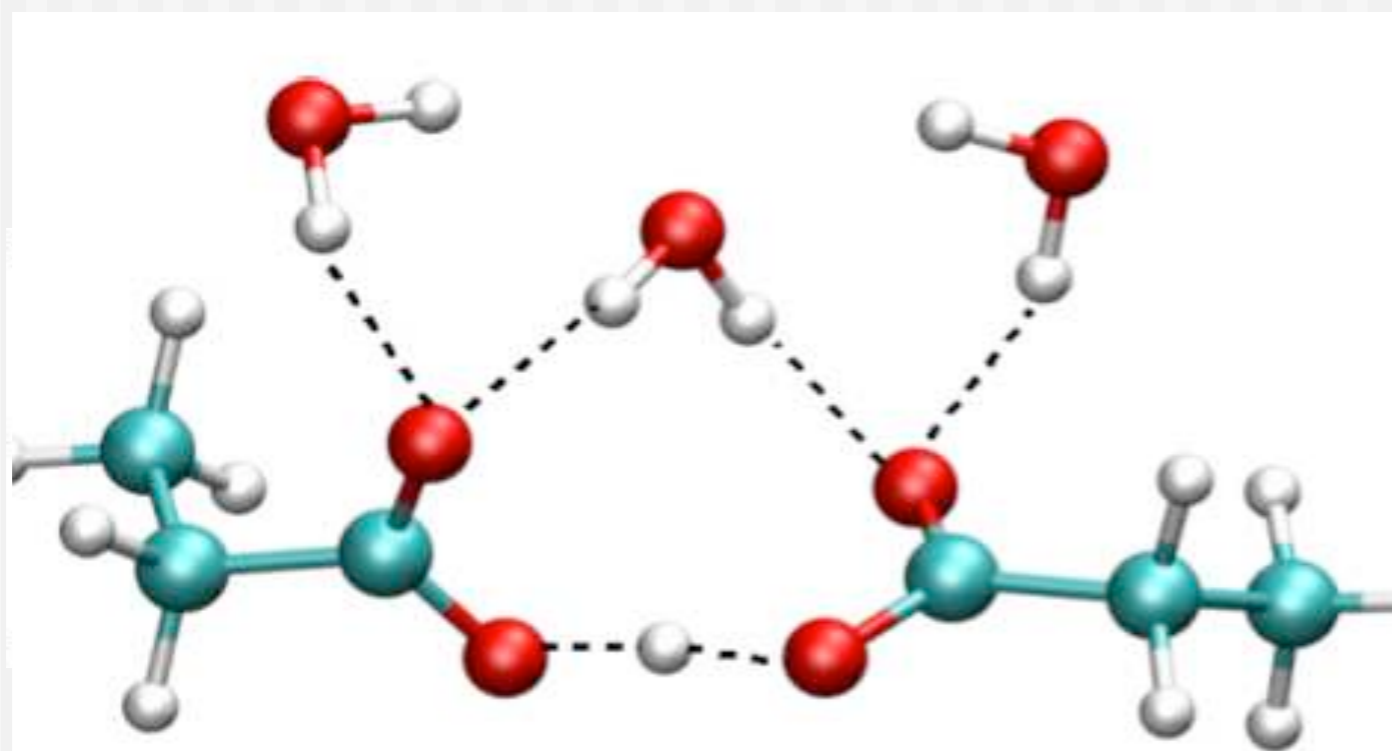
1700-1750 cm^{-1}

DFT (PBEPBE) 6-311++g(2d,2p)	Mode
1618.65	ν_{COOH}
1739.93	ν_{OH}

3) C=O stretch
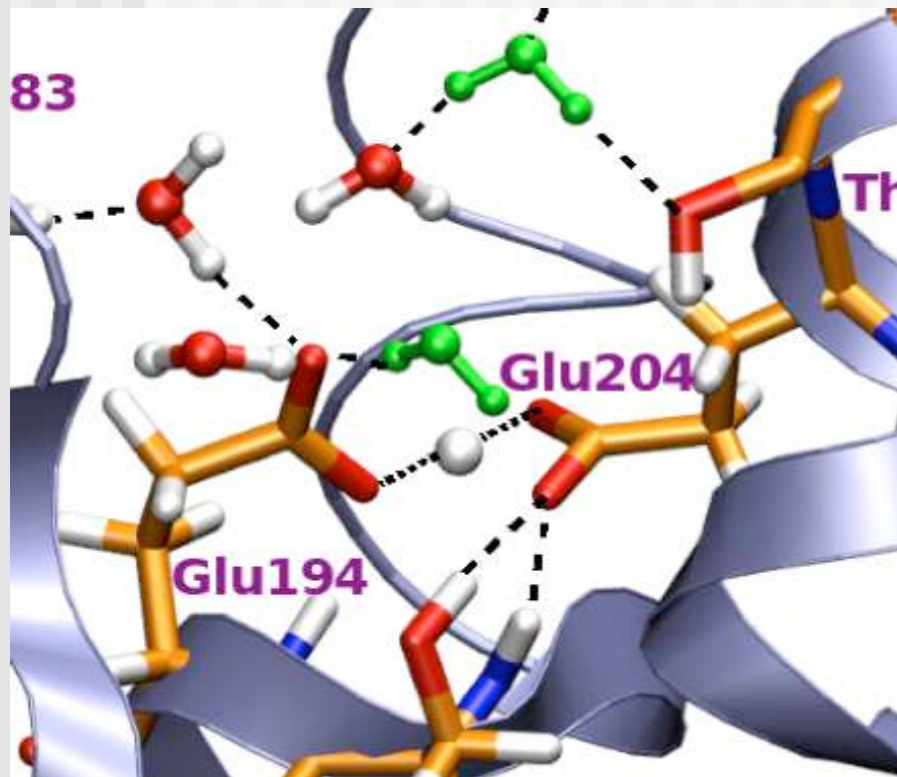
what about C=O?

➔ C=O is below 1690cm^{-1} due to strong H-bond

can not be detected experimentally due to strong absorption of backbone!
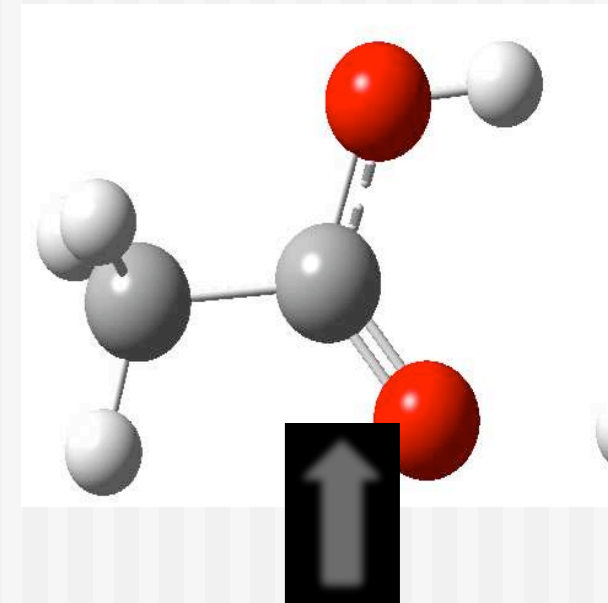
1700-1750 cm^{-1}
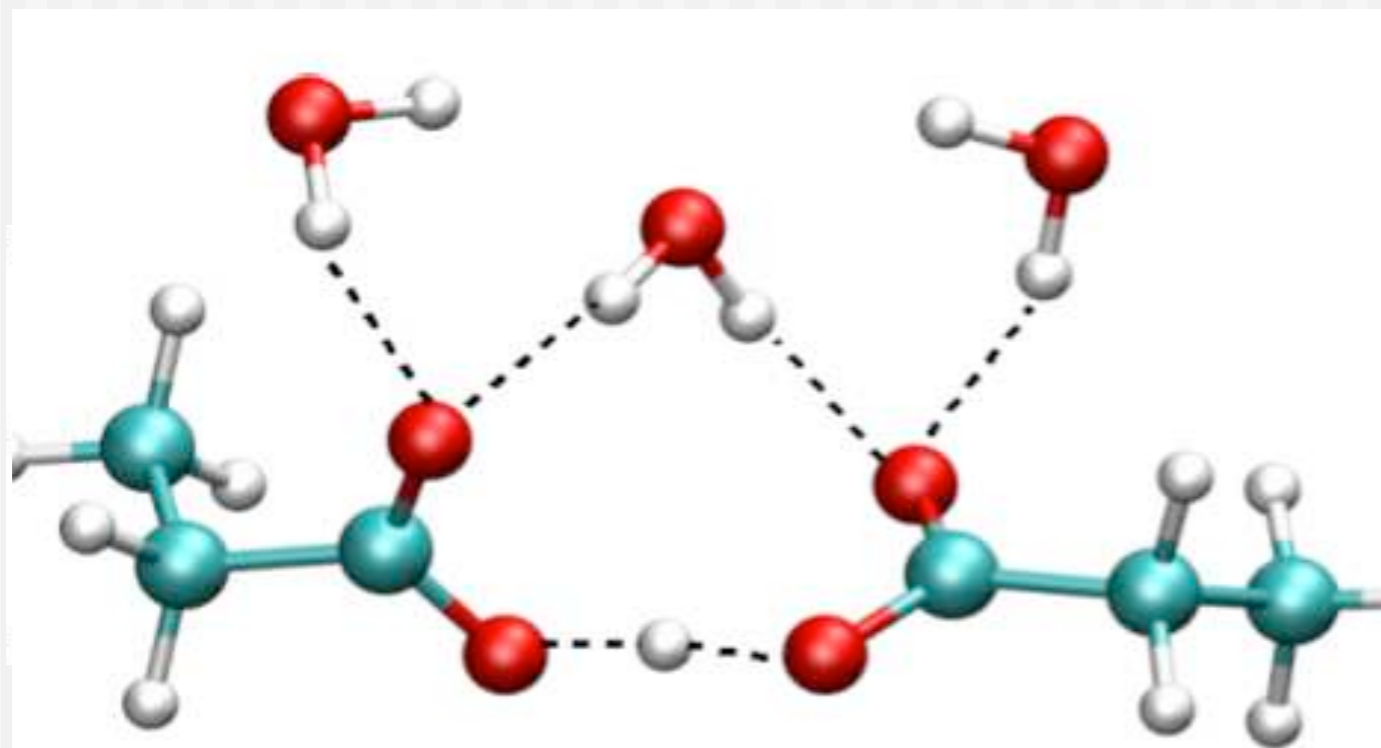
DFT (PBEPBE) 6-311++g(2d,2p)	Mode
1618.65	ν_{COOH}
1739.93	ν_{OH}

QM/MM MD on PRG

Downward movement of ARG82 occurs after 1st PT:
  leads to proton release

Clemens et al., to be published

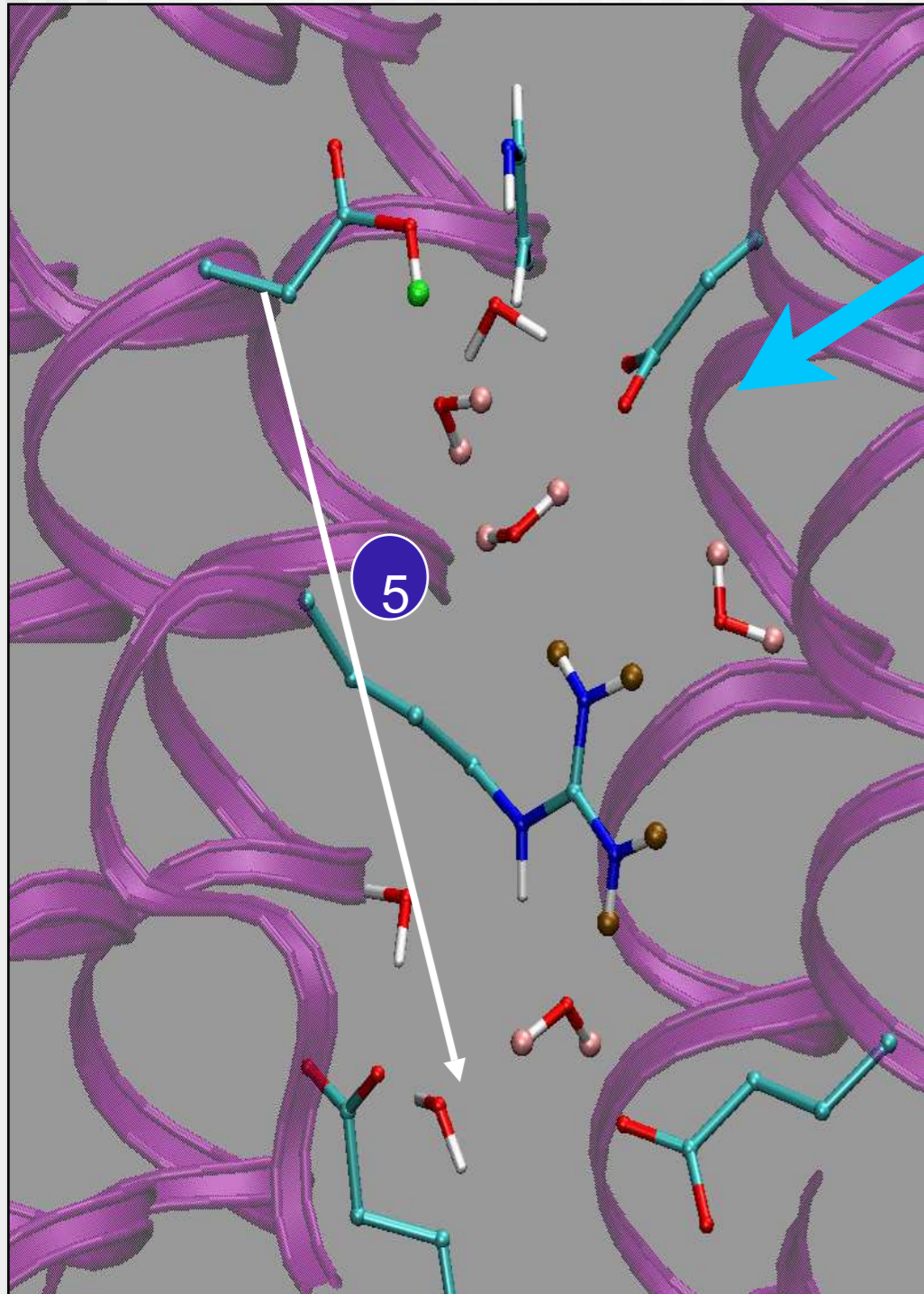Structural model of proton release group: proton shared by Glu's

1) consistent with crystal structures

2) strong H-bond shift C=O below 1700cm^{-1}: not measurable

3) continuum mode is NOT protonated water cluster,
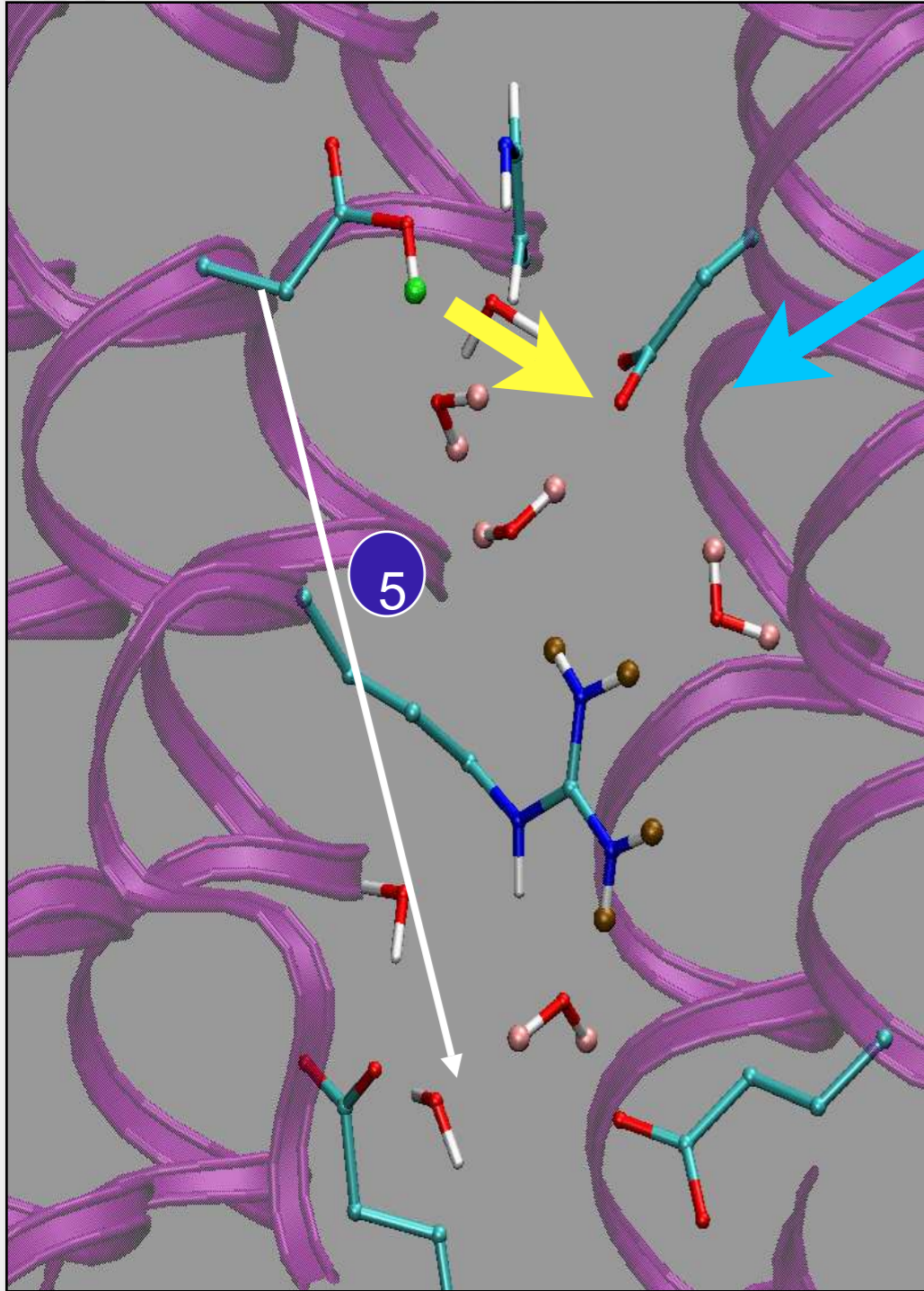
  strong H-bonds (LBHB's) do as well

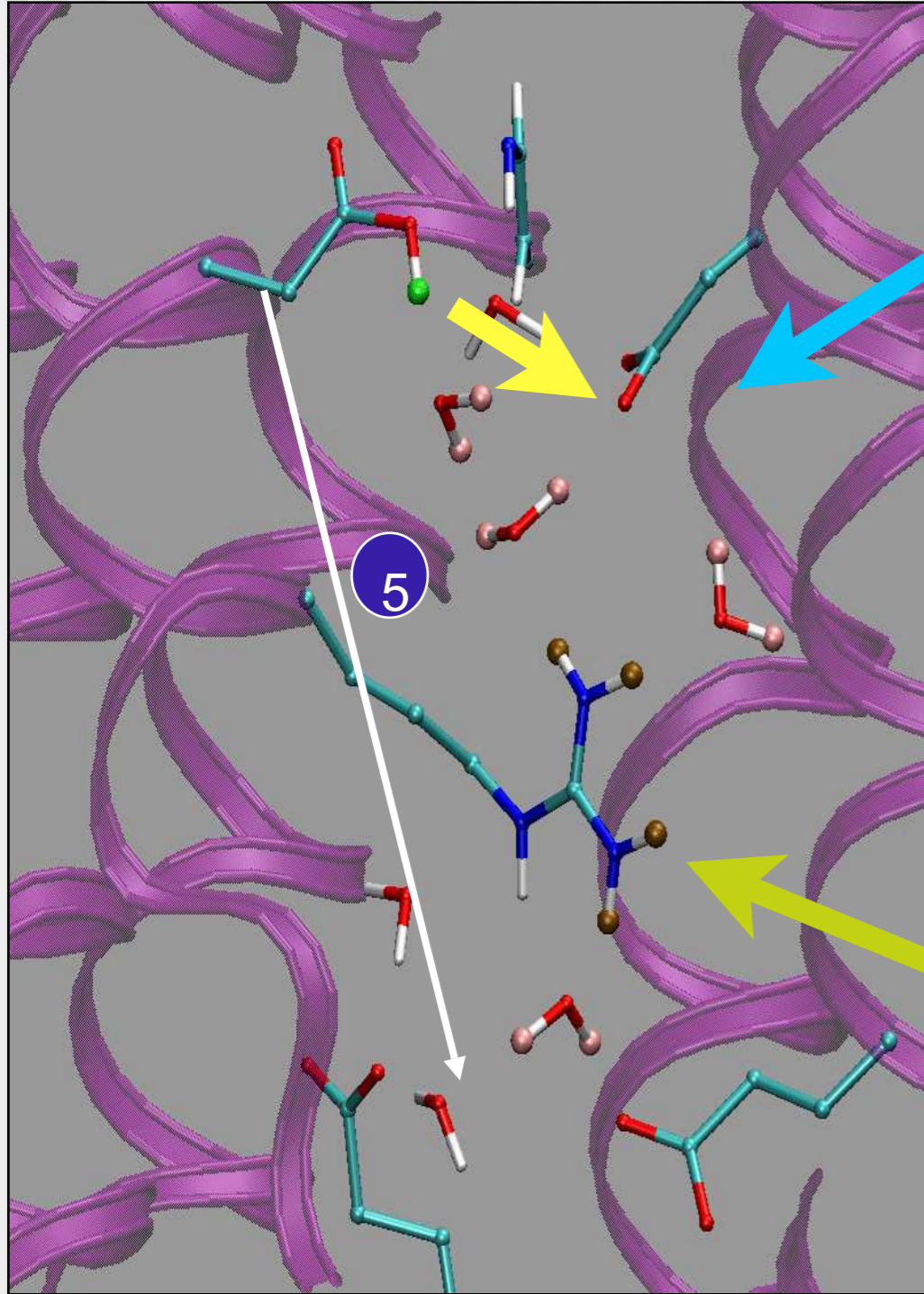Phatak et al., PNAS 105, 19672
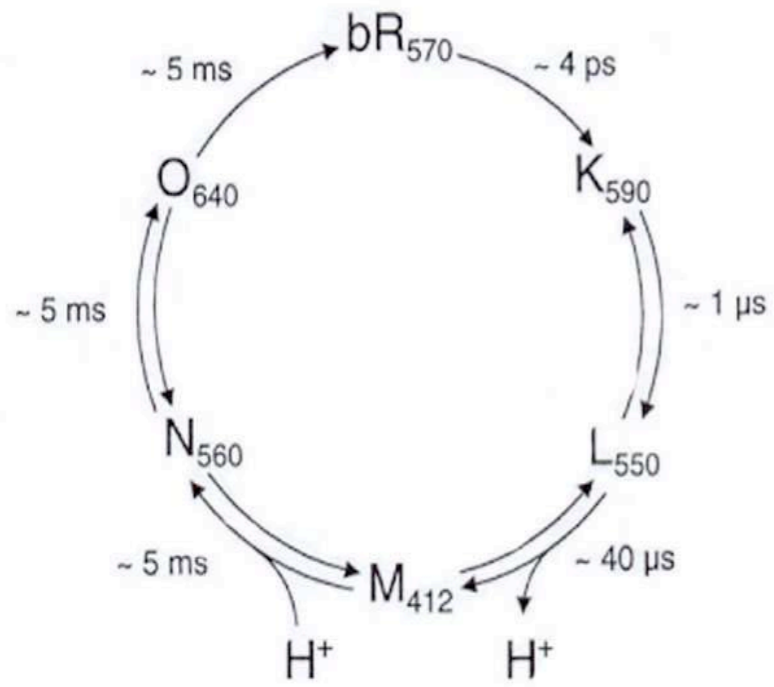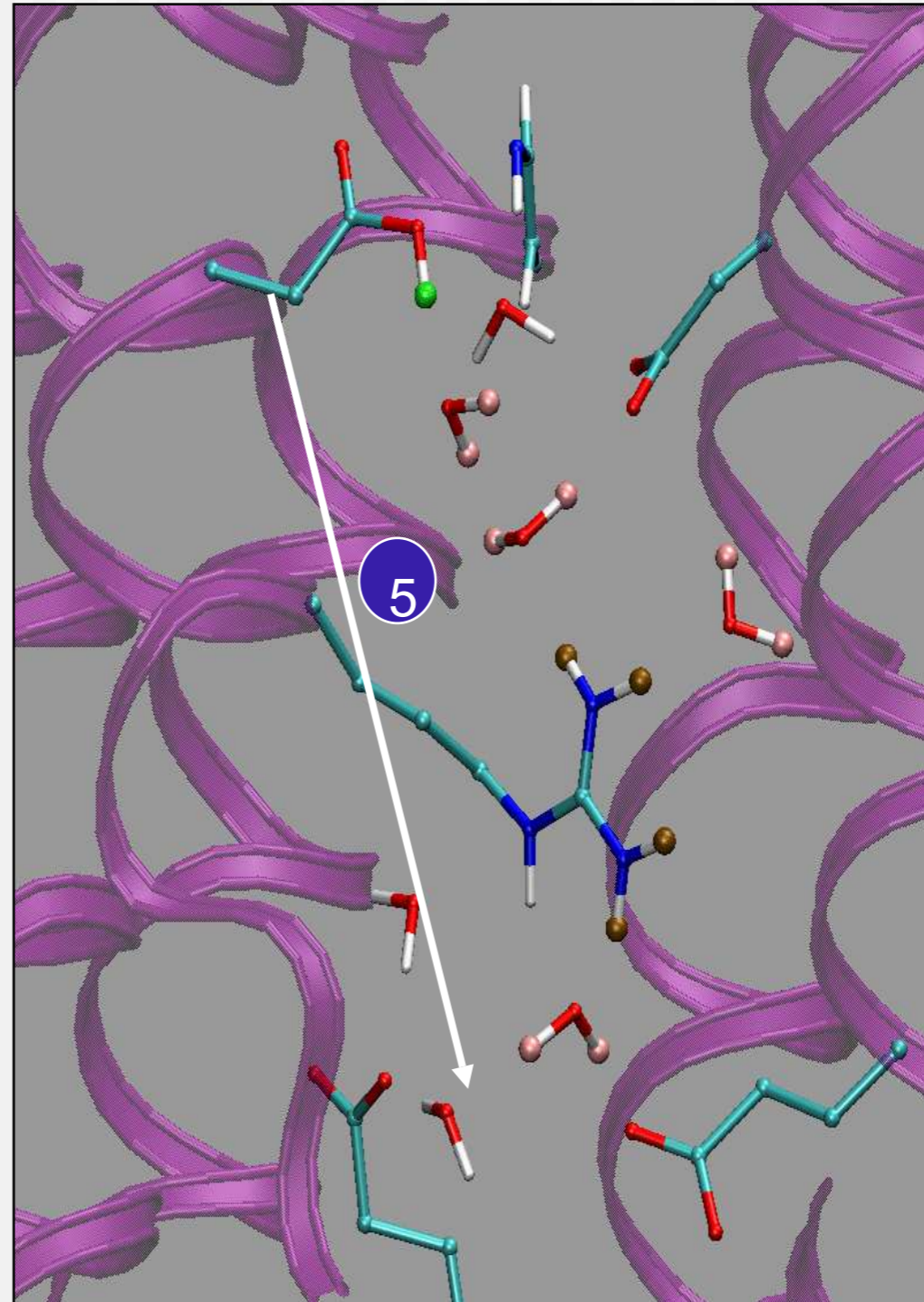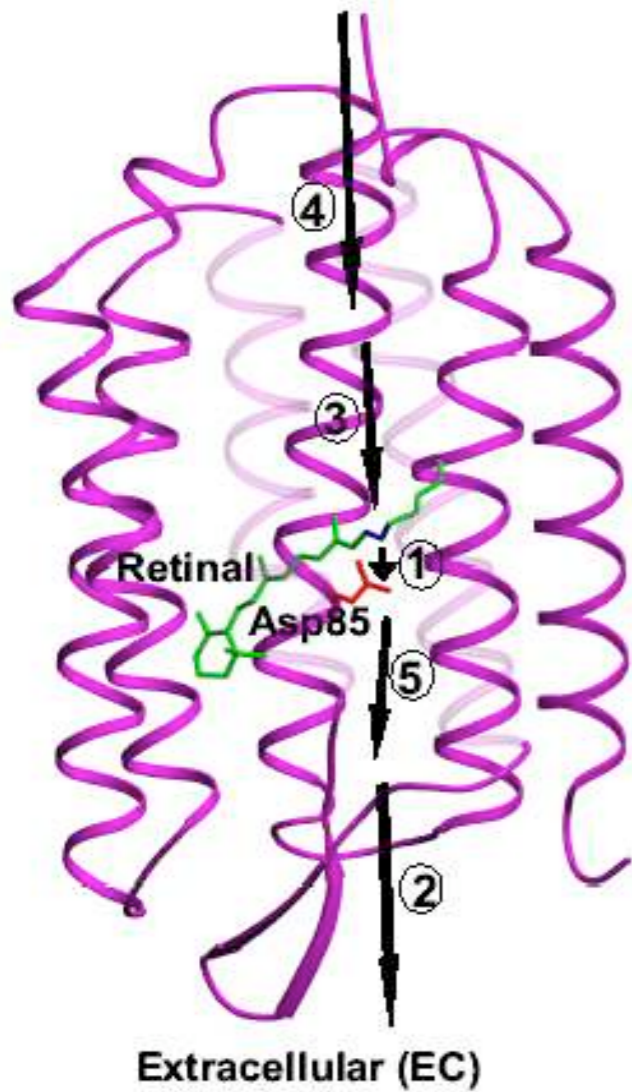
no wt X-ray structure, open questions:

1. number of waters inactive site?

1. number of waters inactive site?

2. is there an {O} intermediate
   with Asp212 protonated?

	O	{O}
C=O (cm^{-1})	1752	1712
UVvis (eV)	1.94	1.98

1. number of waters inactive site?

2. is there an {O} intermediate
   with Asp212 protonated?

	O	{O}
C=O (cm^{-1})	1752	1712
UVvis (eV)	1.94	1.98

3. movement of Arg82 side-chain?
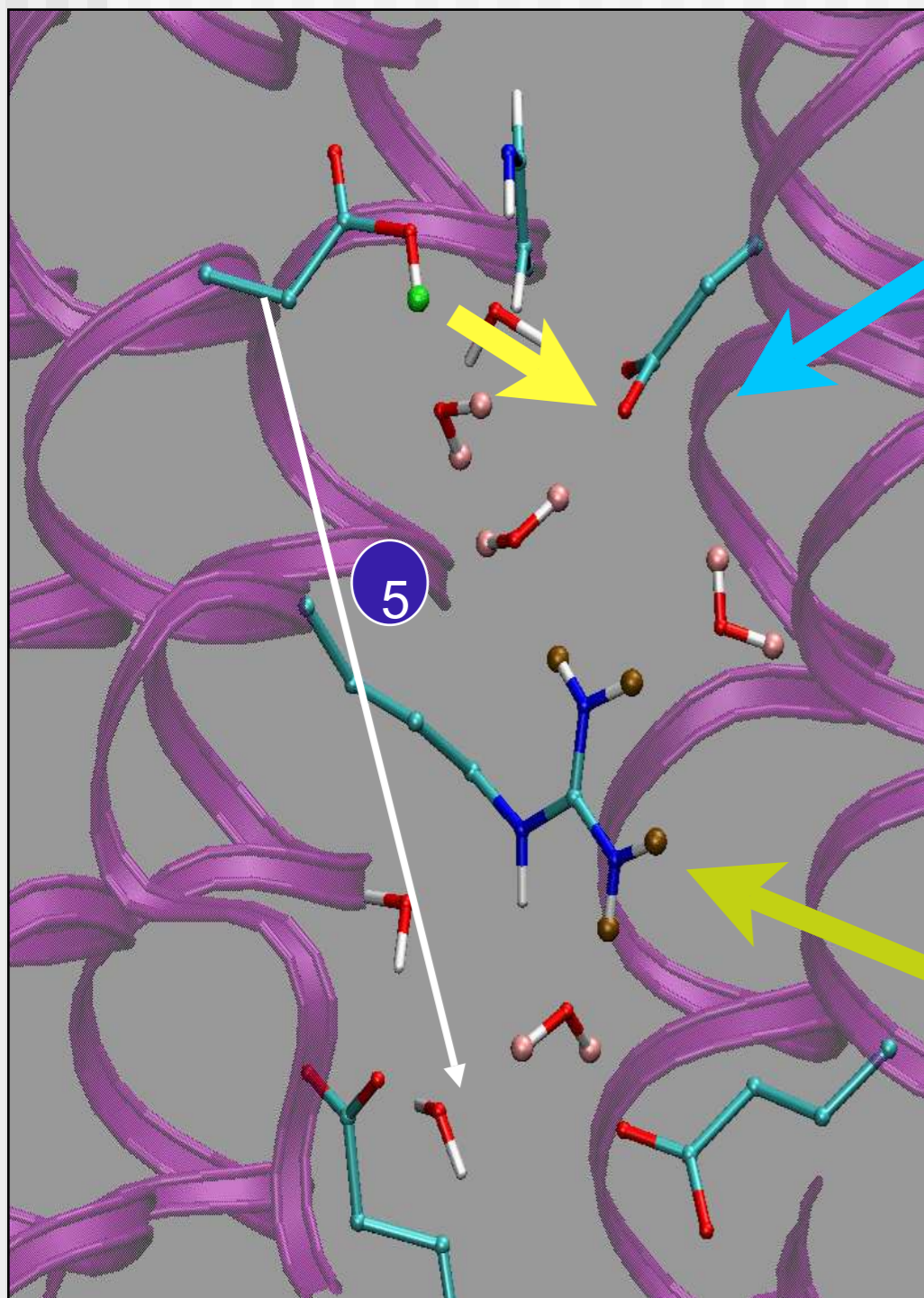
0. **use acid blue strucure**!
1. number of waters inactive site: **3**

2. is there an {O} intermediate
   with Asp212 protonated? **YES**

	O	**{O}**
C=O (cm^{-1})	**1752**	**1712**
UVvis (eV)	**1.94**	**2.00**

3. movement of Arg82 side-chain? **DOWN**

Phatak et al., JACS 2009, in print
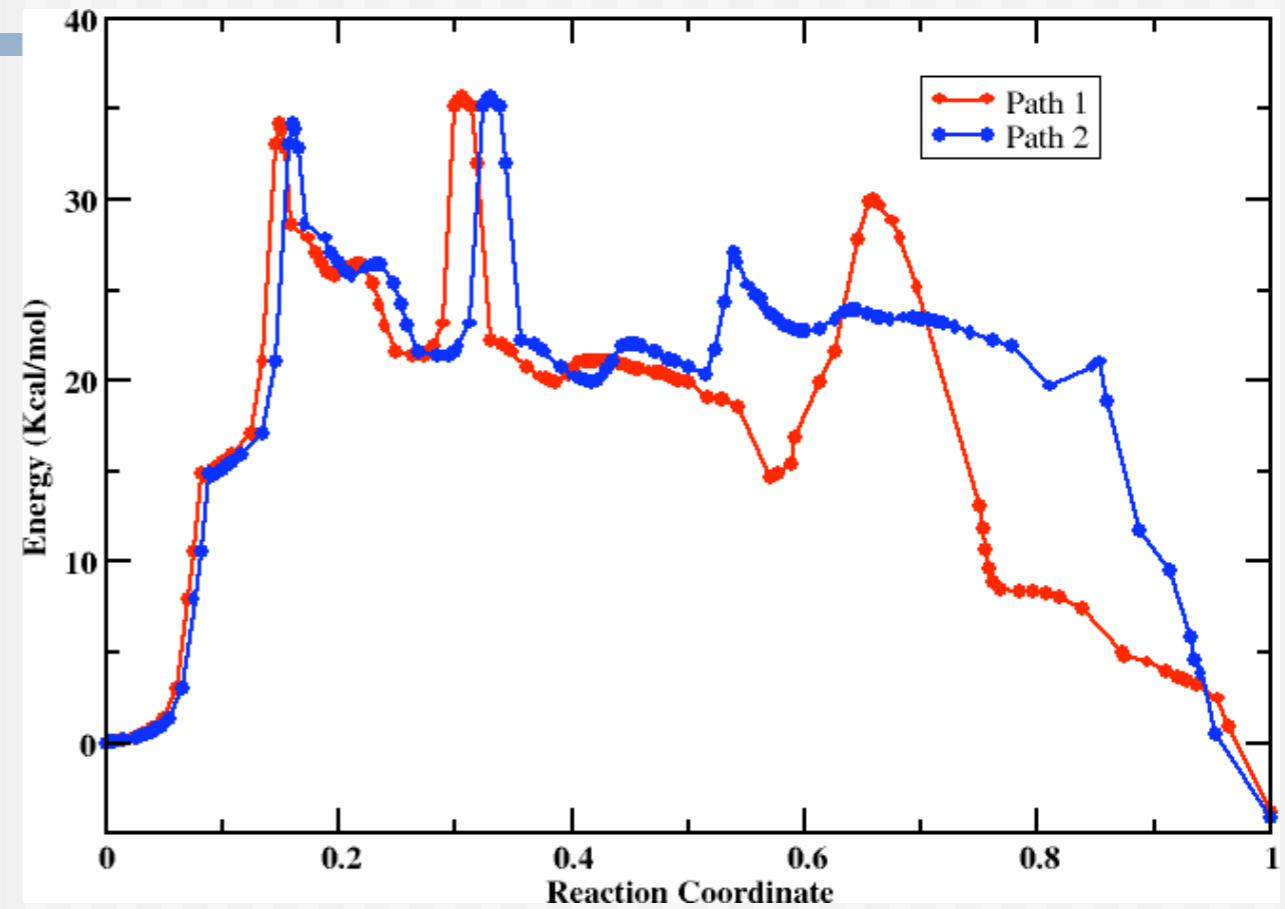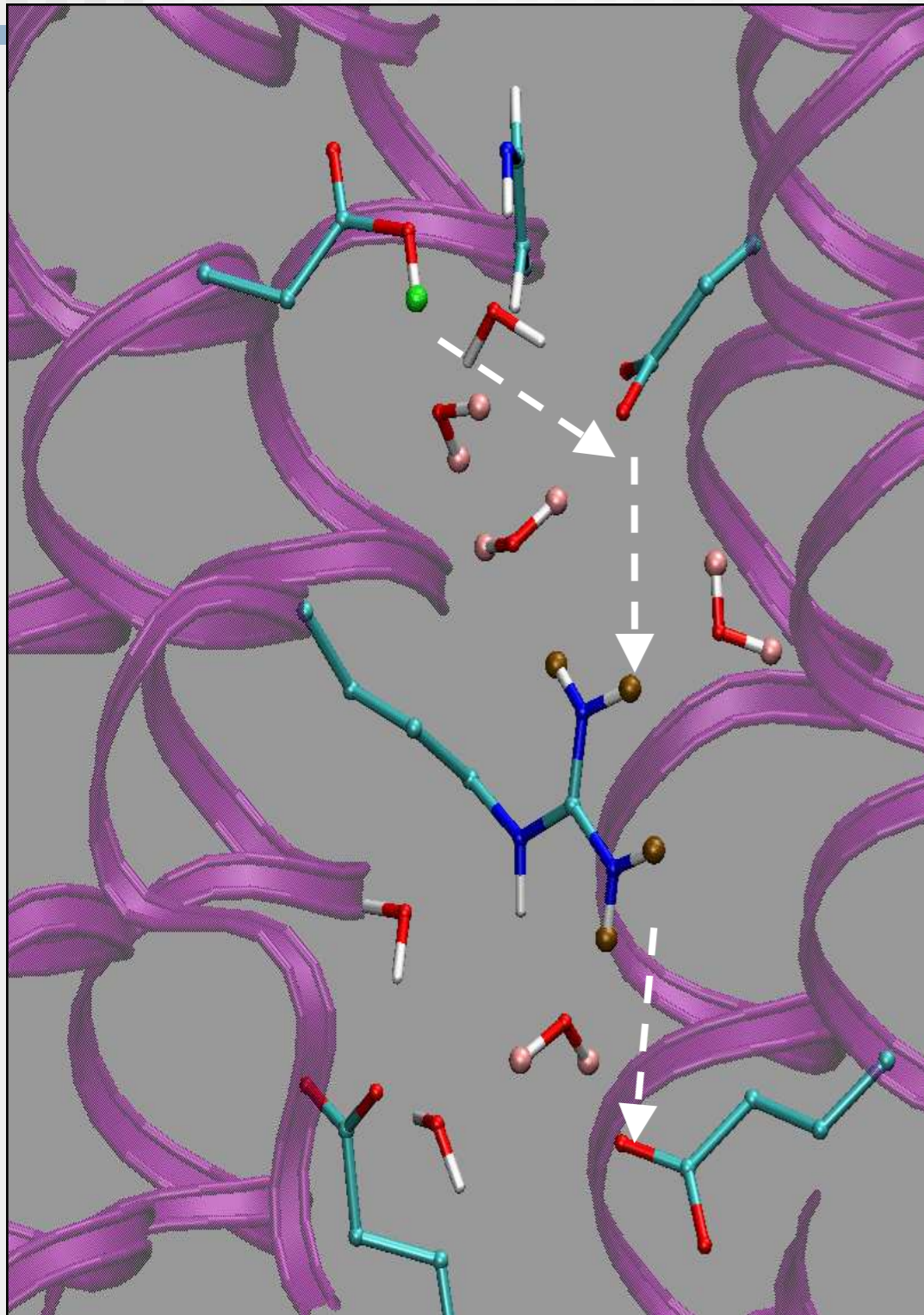
O➜ bR: PT step 5
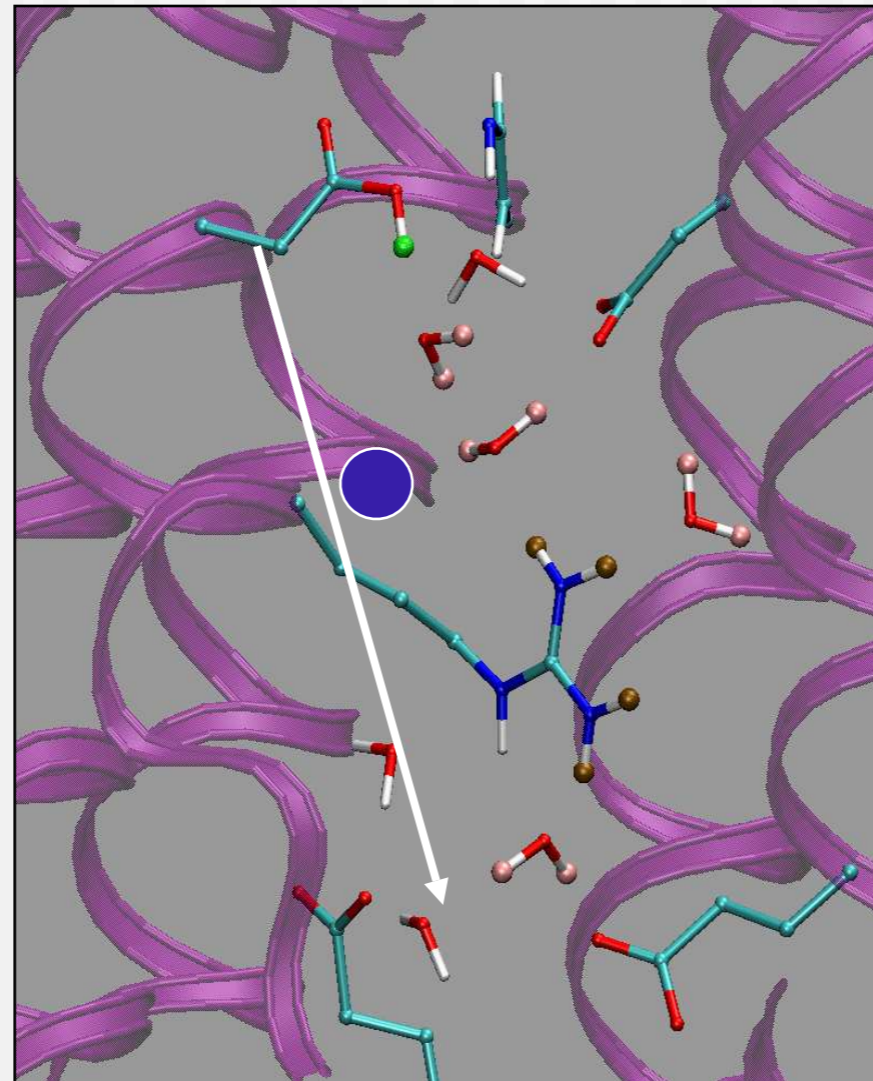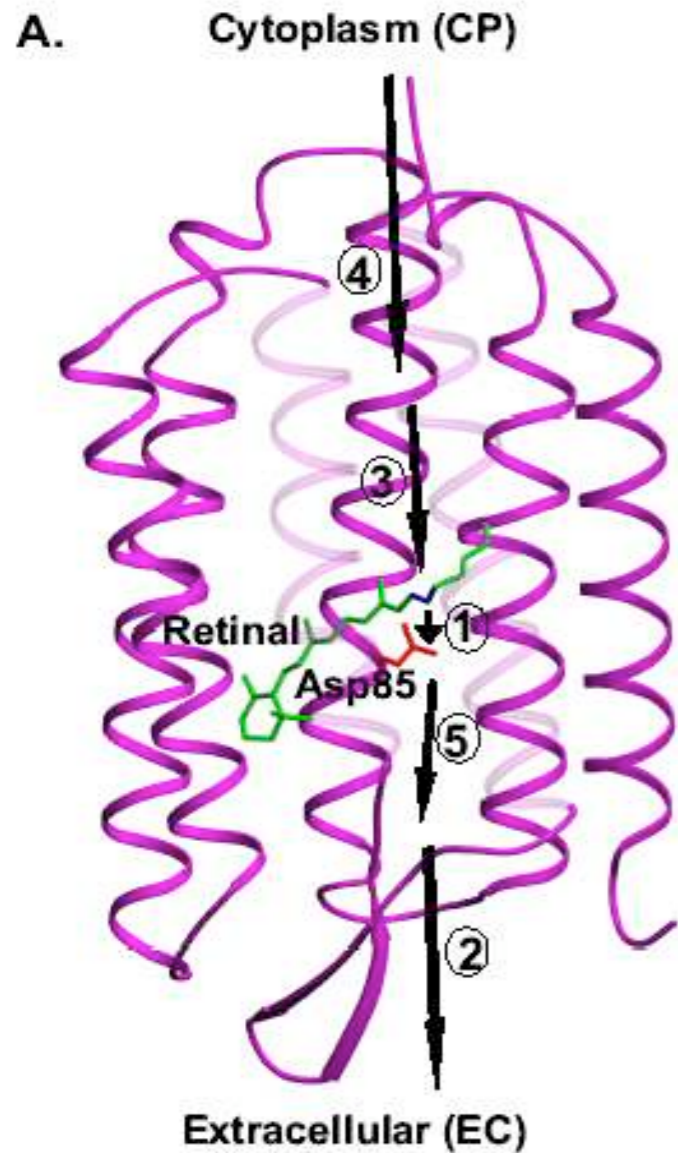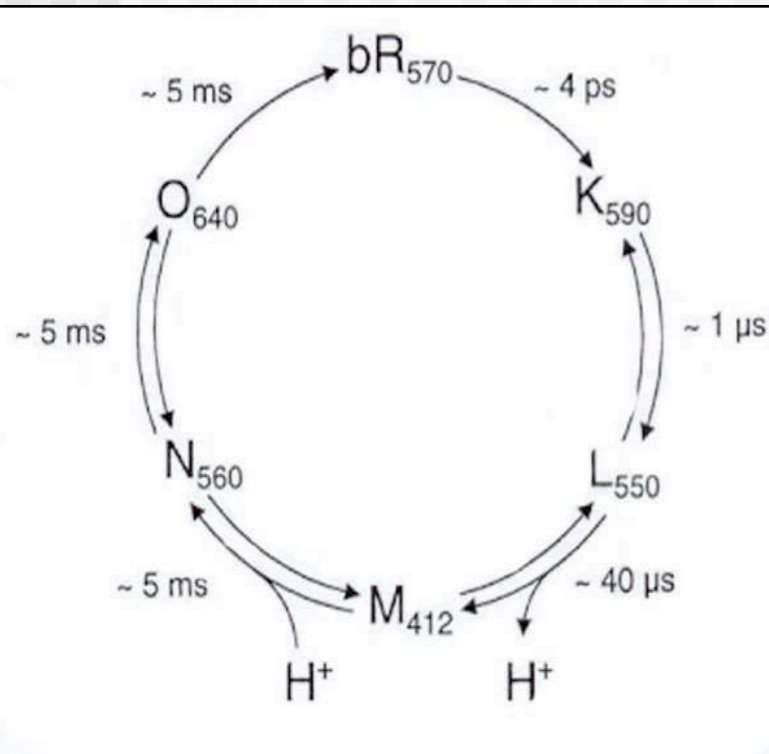
initial and end states resolved:

calculate PT pathways using:
- reation path techniques
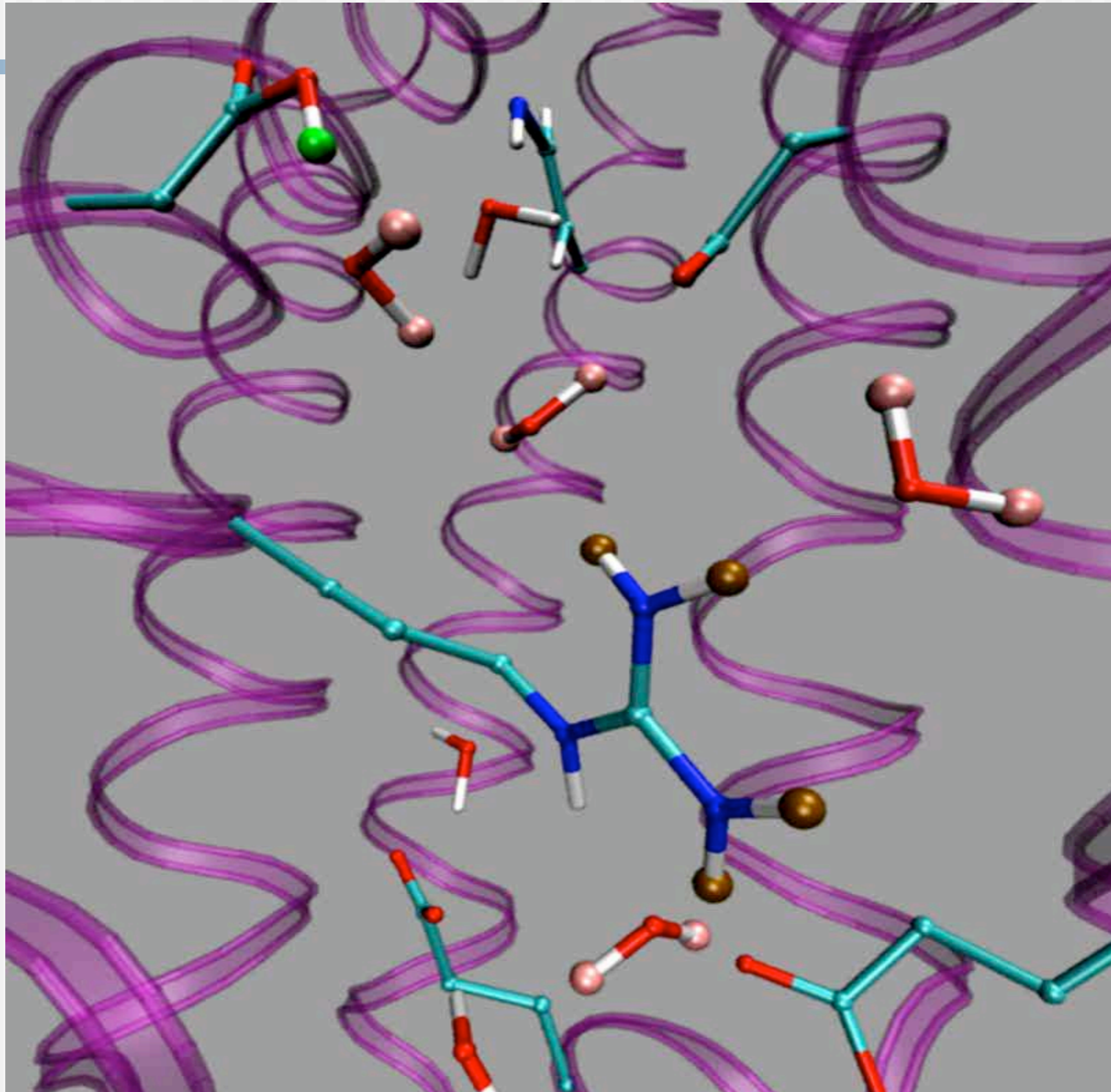- umbrella sampling

Reaction path calculations for PT

Model	Path	Member of proton wire	Activation energy (kcal/mol)	Reaction Energy (kcal/mol)
gsbr	Path1	Arg82 & waters	**35.6**	**-3.8**
gsbr	Path 2	Asp212, Arg82 & waters	**35.6**	**-4.2**

O➜ bR

Model 2:Arg82 down
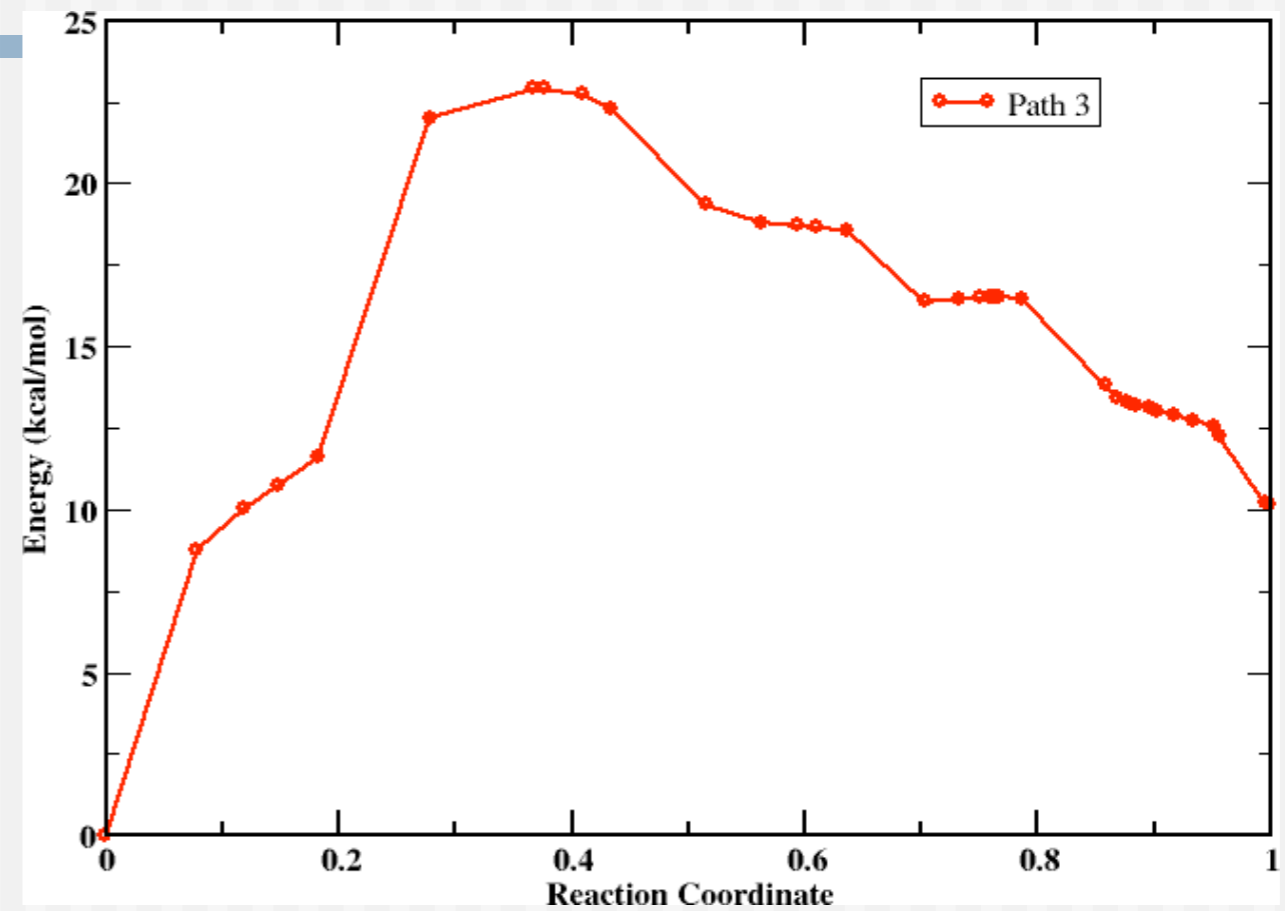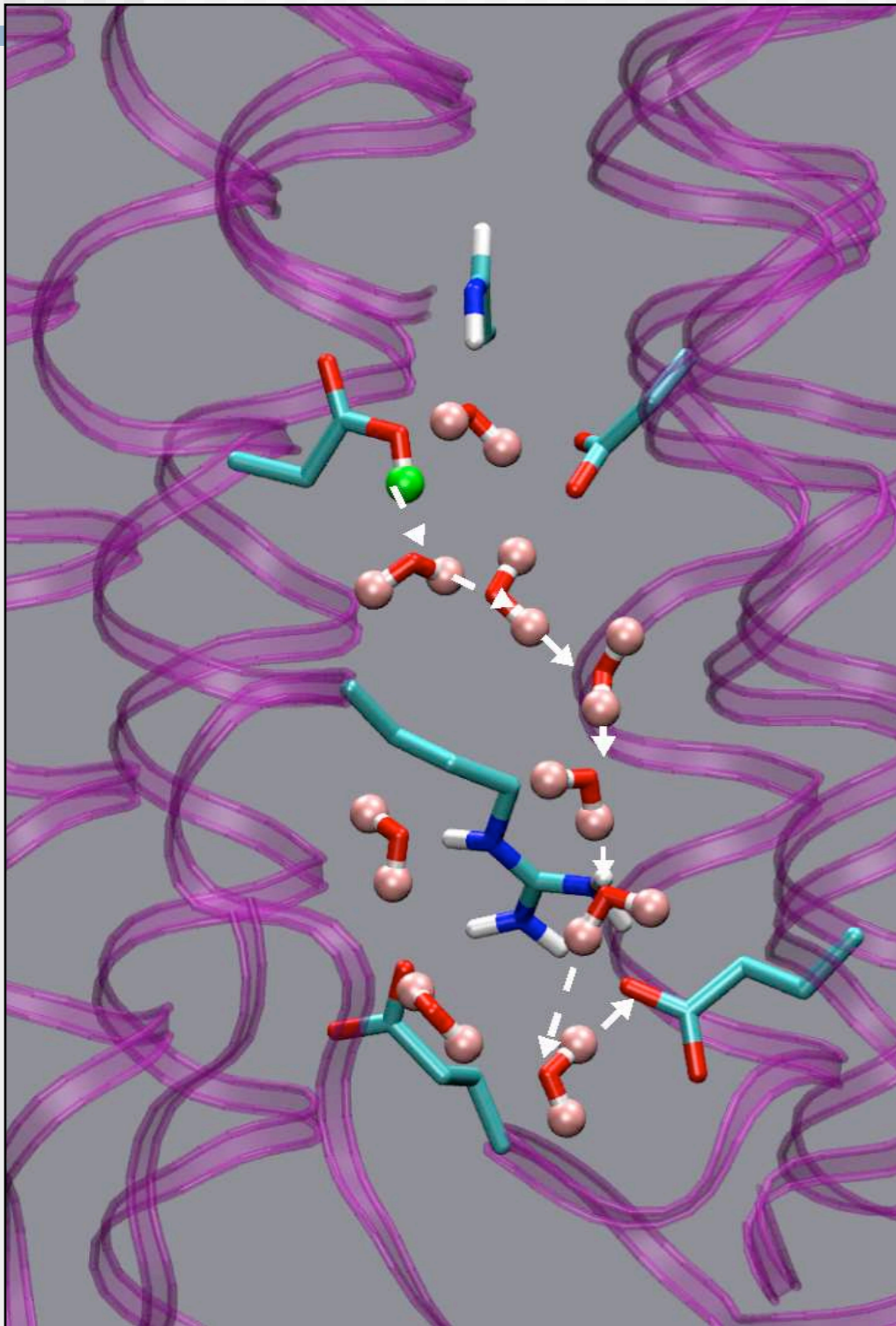
Model	Path	Member of proton wire	Activation energy (kcal/mol)	Reaction Energy (kcal/mol)
da-gsbr	Path3	waters	**23.0**	**10.1**

Proton transfer: other systems
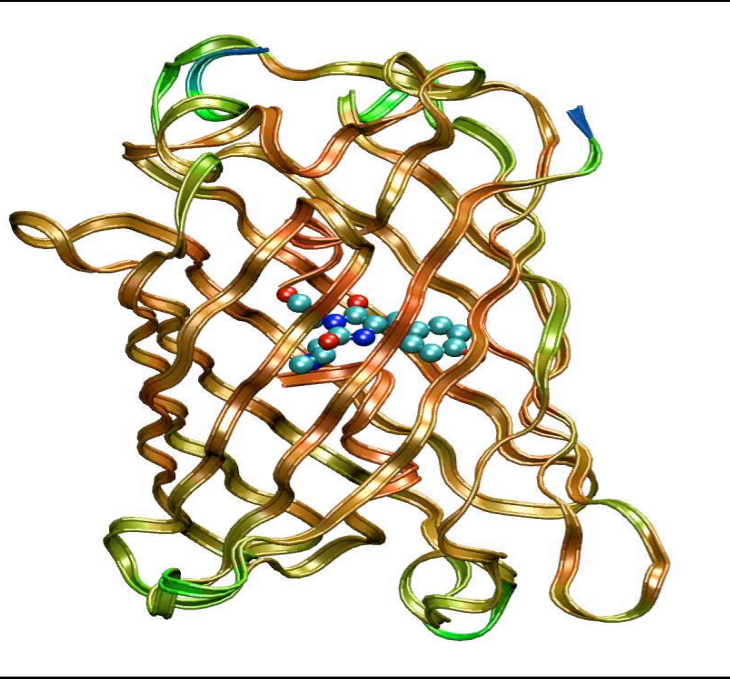
GFP

Proton transfer

Bacterial reaction center

Bacteriorhodopsin

Aquaporin

ADH, CA

GFP: excited states PT

DNA- Bases

Aquaporin
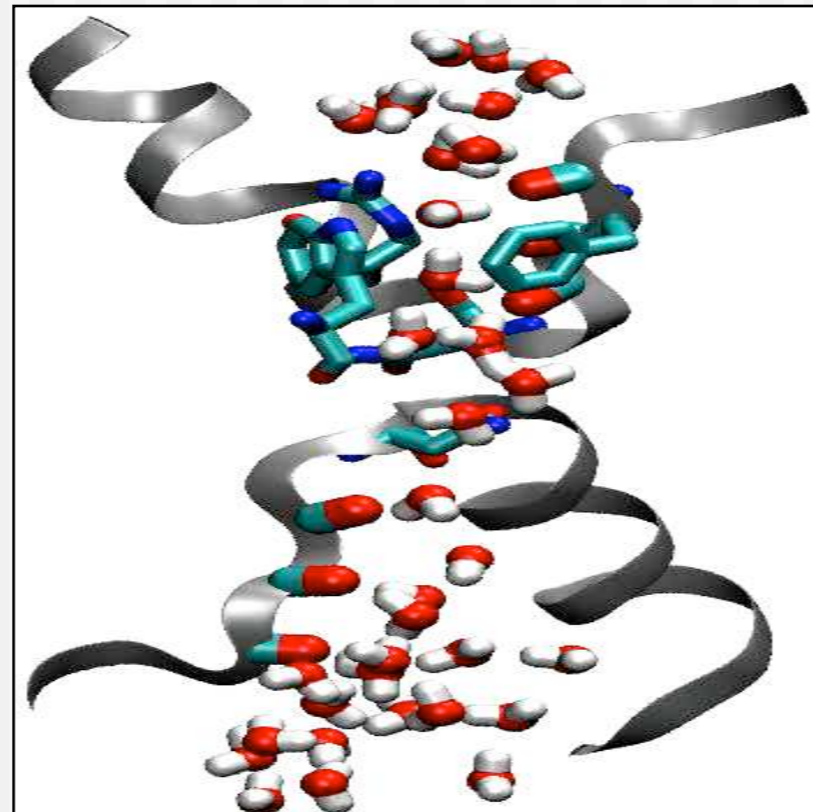
Photosynthetic Reaction Center