Structural Biology - BIO315

Master SV - Spring Semester
Lecture 4

Matteo Dal Peraro
matteo.dalperaro@epfl.ch
AAB 048, phone: 31681

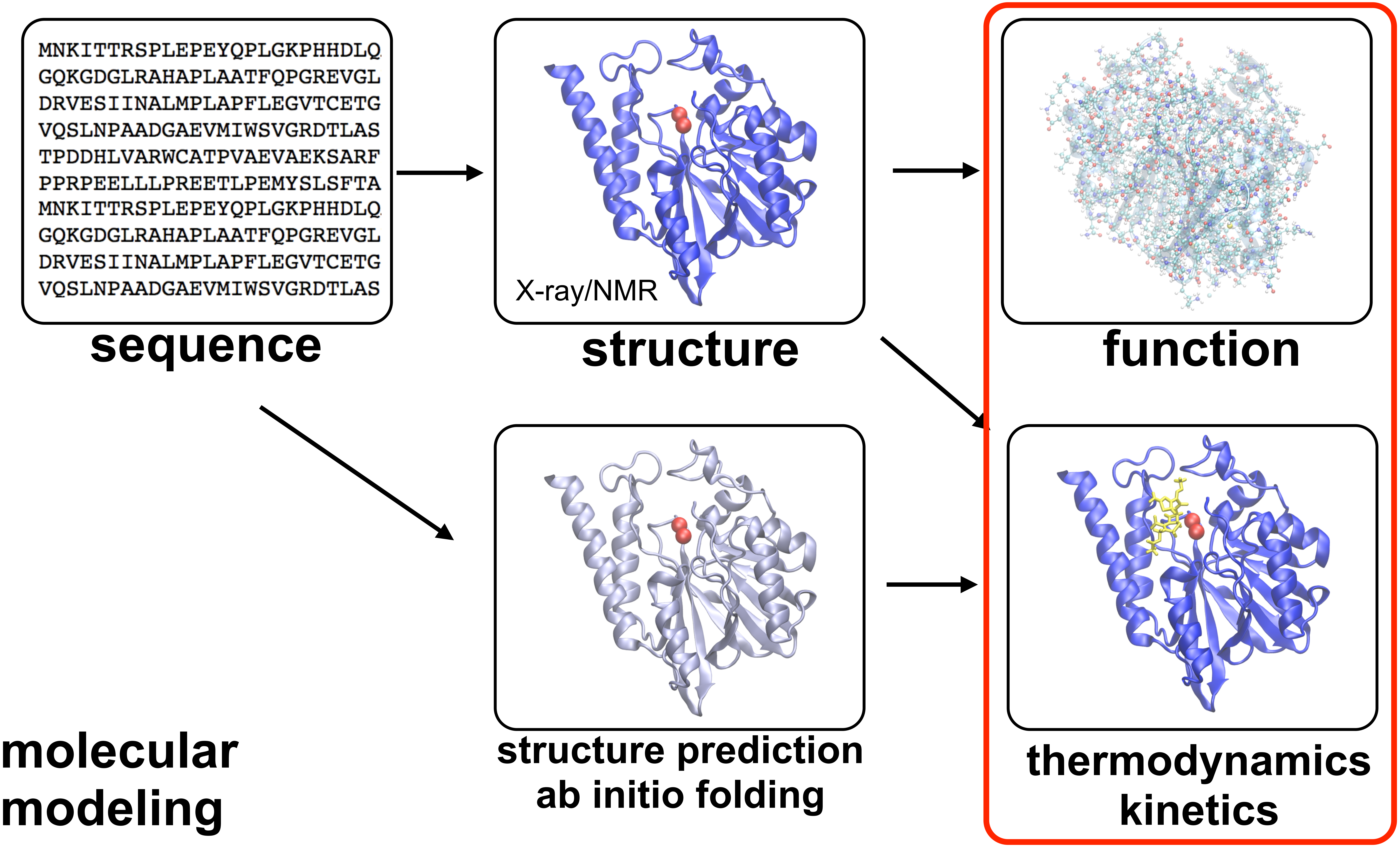
Outline of lecture 4

- introduction to **molecular mechanics** (MM)

- empirical **force fields** for biomolecules
  - ‣some history
  - ‣bonded and non-bonded interactions
  - ‣empirical models for water solvation
  - ‣non-bonded computation

➡ <u>Lab session</u>: week 7, small molecule parametrization (ligands)

Paradigm in Structural Biology



```
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
TPDDHLVARWCATPVAEVAEKSARF
PPRPEELLLPREETLPEMYSLSFTA
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
```

sequence

X-ray/NMR

structure

function

structure prediction
ab initio folding

thermodynamics
kinetics

molecular modeling

- **knowledge-based: structural databases**
- **first principles:** $i\hbar \frac{\partial}{\partial t}\Psi(\mathbf{R},t) = \hat{H}\Psi(\mathbf{R},t)$

Different approaches

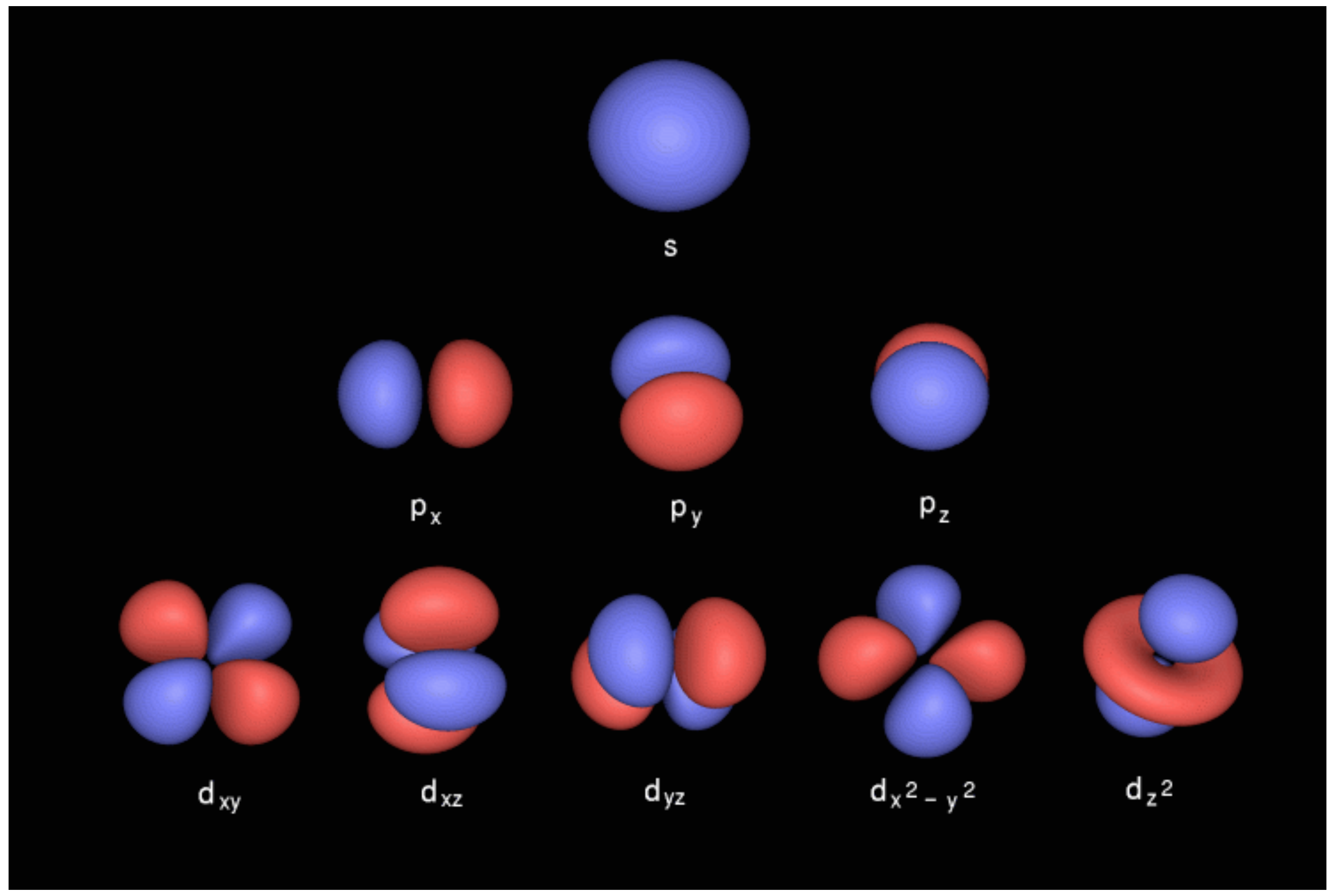
- **sequence-based methods**: structural data are not always available, so they are often used and can be valuable (see also AF lecture)
- **structure knowledge-based methods** can give suitable models based on templates or ML, but don’t describe the physics of biomolecules
- **first principles methods** are based on the laws of physics and aim at having a predictive power
  - they suffer though by limitations that affect **accuracy** (correct description of molecular interactions) and **feasibility** (proper sampling of the conformational space)

First principles modeling

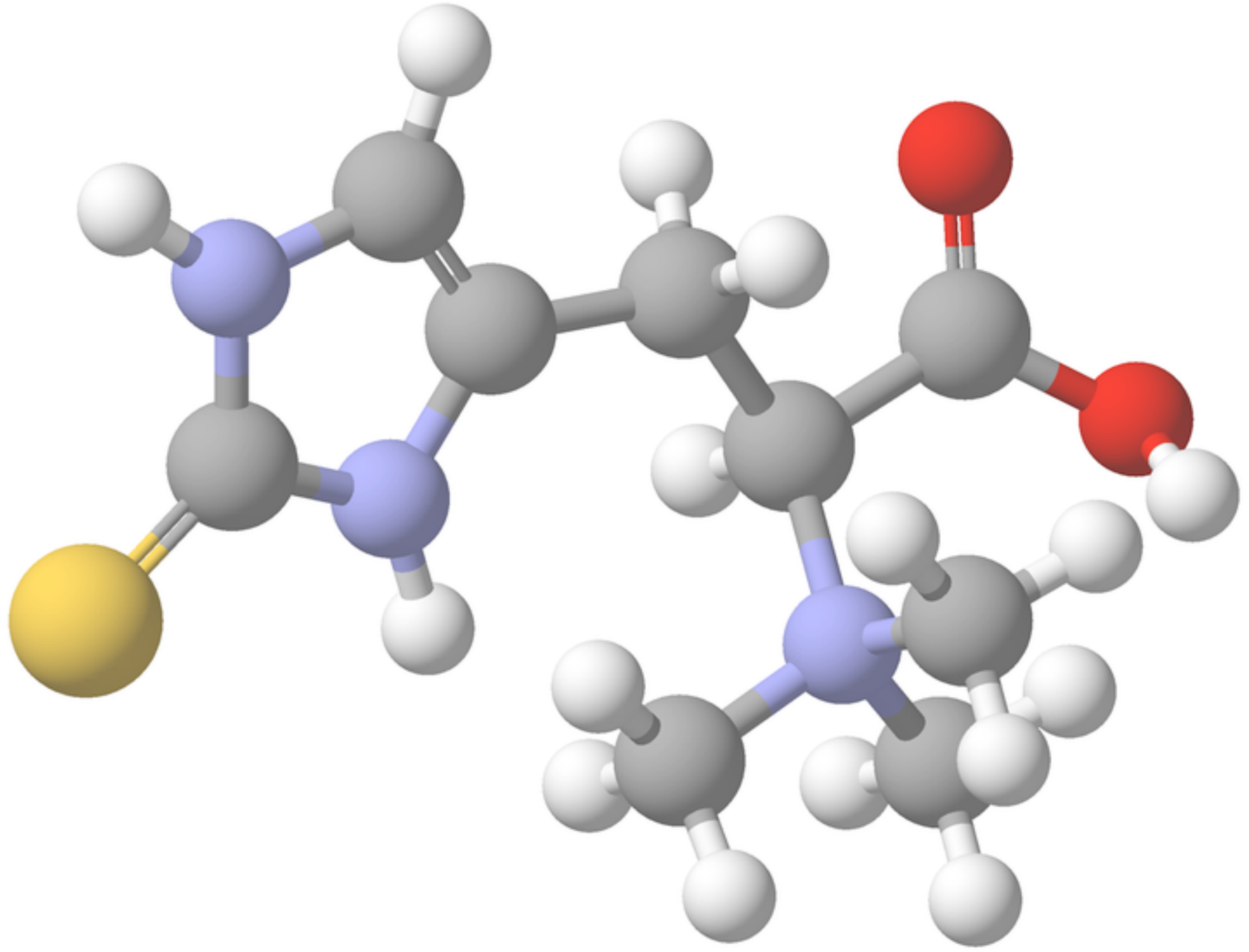
- **dynamics**: how biomolecules move in space and time
- **thermodynamics**: you can compute quantities that characterize the system (*e.g.* enthalpy, entropy, heat capacity, free energy differences, etc.)
- **energy contributions**: which atoms or which forces contribute the most to stability, or binding, etc.
- **conformational analysis**: what are the low-energy structures a protein can adopt (e.g. folding)
- **reactivity**: what are the mechanism and kinetics of chemical reactions (from quantum mechanics calculations)

Levels of resolution in modeling

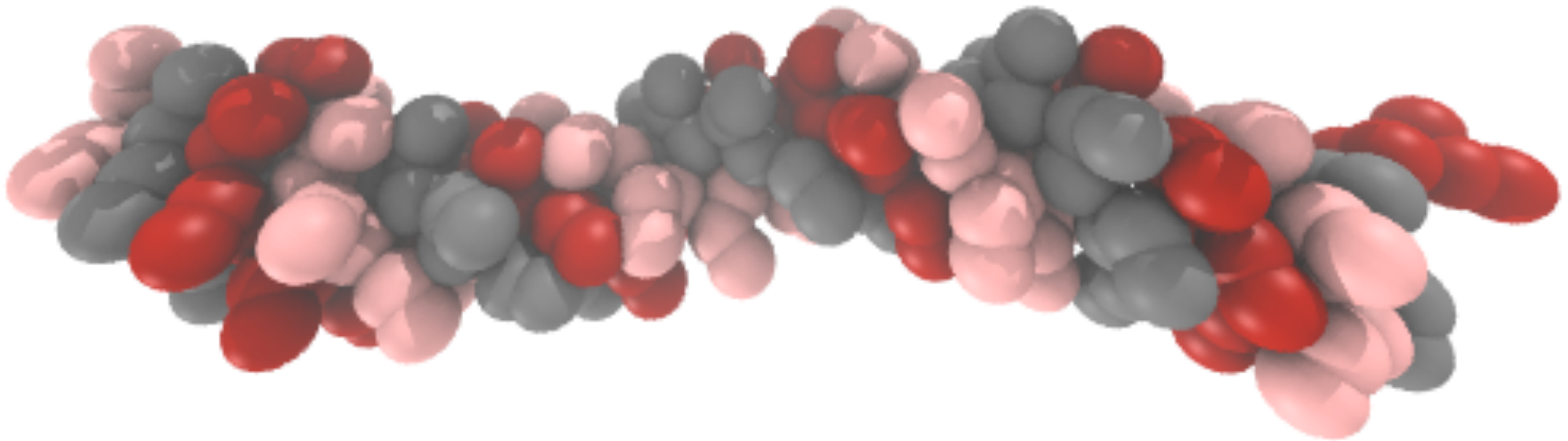
- electrons



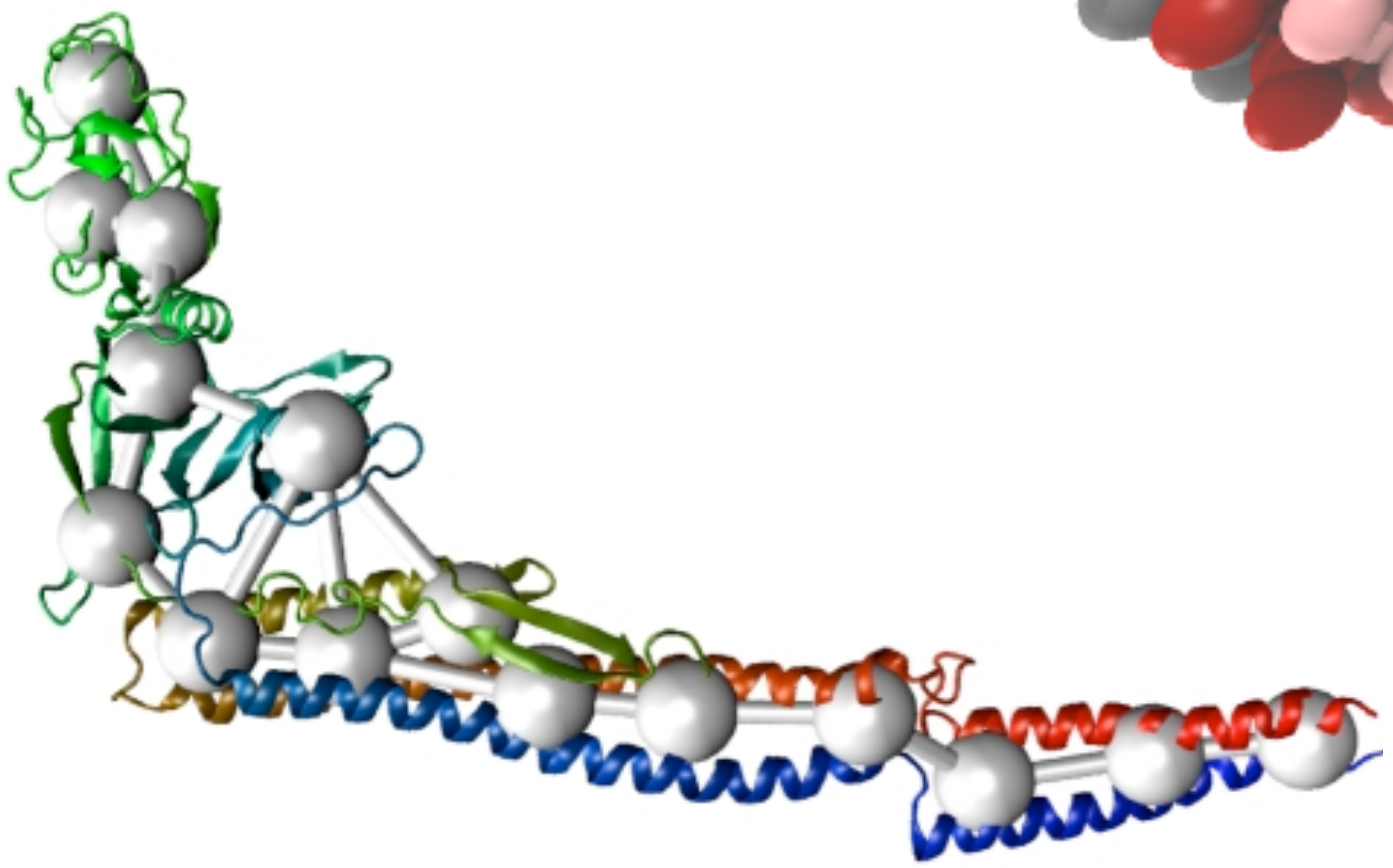
- atoms



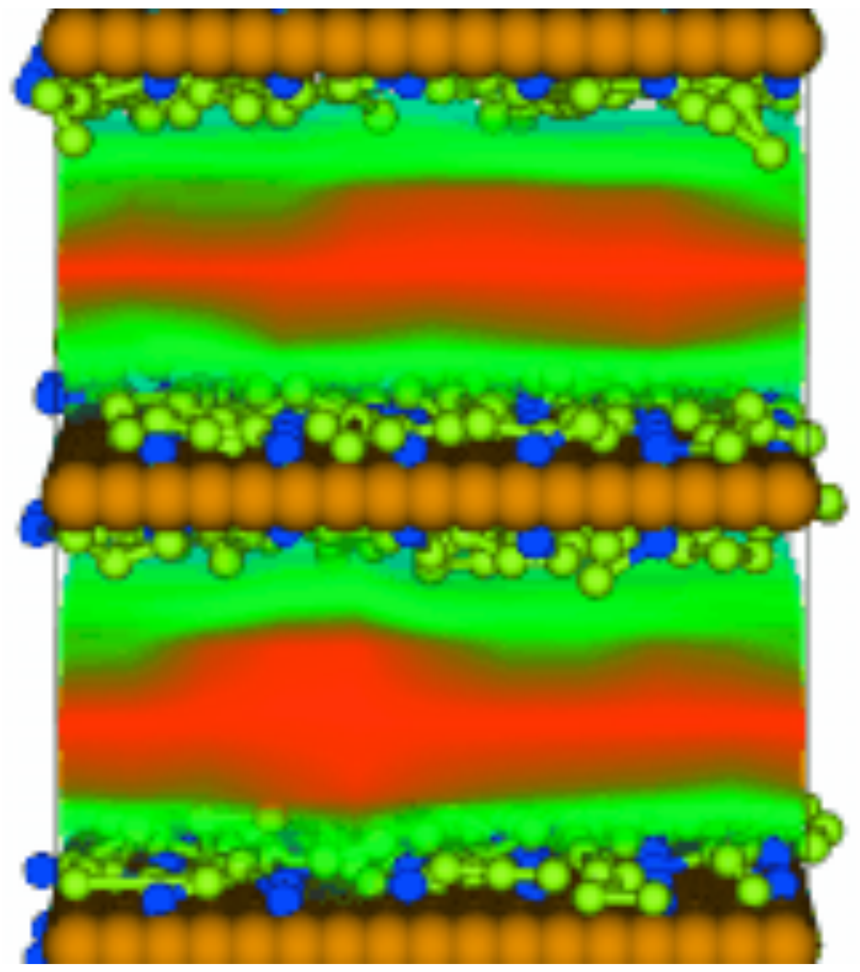
- amino-acids



- domains



- mesoscopic to continuum



What we need for modeling at the molecular mechanics (MM) level

For a molecular simulation or modeling one needs:

1. a **representation** of the biomolecules at a certain level of resolution (i.e. initial conditions)
2. a functional form for the **potential** energy for molecular mechanics (MM)
3. a search algorithm or optimizer/minimizer (**minimization** can be used to find favorable regions in the conformational space; **sampling** techniques to compute dynamics and thermodynamic quantities)

1. Structural initial conditions

```
                                           X         Y         Z      occ    B
ATOM         1   N    GLU       4       28.492    3.212   23.465   1.00 70.88
ATOM         2   CA   GLU       4       27.552    4.354   23.629   1.00 69.99
ATOM         3   C    GLU       4       26.545    4.432   22.489   0.00 67.56
ATOM         4   O    GLU       4       26.915    4.250   21.328   0.00 68.09
ATOM         5   CB   GLU       4       28.326    5.683   23.680   0.00 72.34
ATOM         6   CG   GLU       4       27.447    6.910   23.973   0.00 75.98
ATOM         7   CD   GLU       4       28.123    8.247   23.659   0.00 78.43
ATOM         8   OE1  GLU       4       29.375    8.299   23.604   0.00 79.32
ATOM         9   OE2  GLU       4       27.393    9.251   23.468   0.00 79.58
ATOM        10   N    ARG       5       25.274    4.610   22.852   1.00 63.77
ATOM        11   CA   ARG       5       24.179    4.807   21.907   1.00 59.83
ATOM        12   C    ARG       5       23.411    3.698   21.219   1.00 56.20
ATOM        13   O    ARG       5       23.987    2.808   20.596   1.00 57.33
ATOM        14   CB   ARG       5       24.604    5.784   20.812   1.00 60.86
ATOM        15   CG   ARG       5       23.926    7.127   20.866   1.00 61.89
ATOM        16   CD   ARG       5       24.295    7.944   19.647   1.00 62.21
```

2. Some historical milestones of MM

- **1920s**: quantum theory development (BO approx.)
- **1946**: first attempts of MM (Wertheimer et al.)
- **1956**: dynamics of hard spheres (Alder et al.)
- **1960s**: force fields development (Lifson at Weizmann, Scheraga at Cornell, Allinger at UGeorgia)
- **early 1970s**: first MD of water (Rahman & Stillinger)
- **late 1970s**: biomolecular force fields (Karplus, Kollman, Van Gunsteren et al.)
- **1990s**: large development of parallel codes and HPC (high performance computing) architectures (Schulten)

Quantum mechanical foundation

- quantum mechanics (**QM**) describes the energy of a molecule in terms of a total wavefunction describing the location and motion of nuclei and electrons

$$H\Psi(\mathbf{r},\mathbf{R}) = E\Psi(\mathbf{r},\mathbf{R})$$

- non-tractable even for simple systems (e.g. He)
- **Born-Oppenheimer approximation** permits to decouple electrons and nuclei degrees of freedom

$$\Psi(\mathbf{r},\mathbf{R}) = \phi(\mathbf{r})\bullet\chi(\mathbf{R})$$

(protons are 1836 times heavier than electrons).

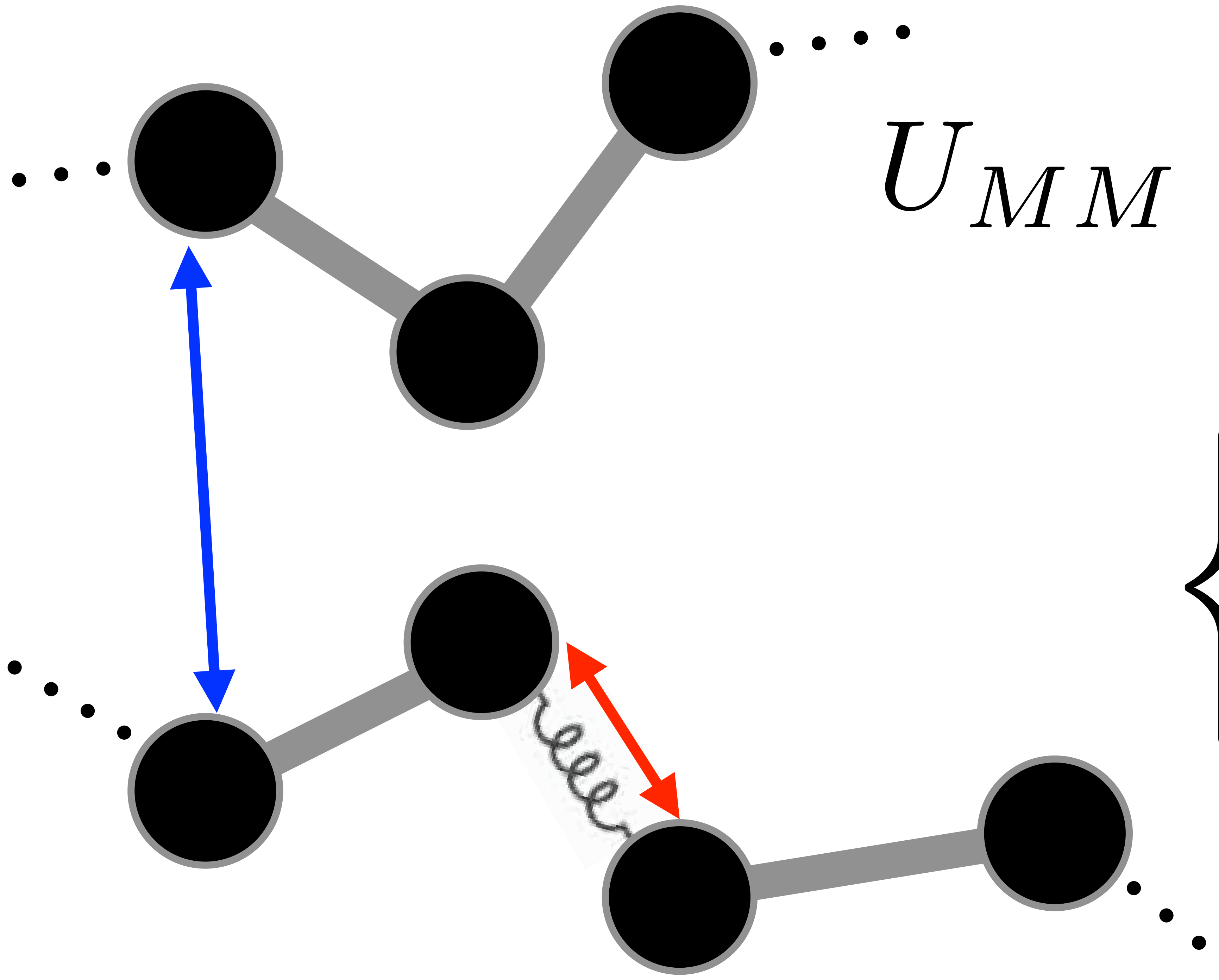
- still very expensive to compute energies of proteins and DNA using quantum methods

Molecular mechanics principles

- BO approximation permits to write the potential energy of a system as a function of the **nuclear** positions only (unlike in QM, the potential is empirically evaluated)
- **thermodynamic hypothesis**: native structures of biomolecules are in a global minimum of free energy
- **additivity**: total energy potential can be divided in simpler local and non-local contributions
- **transferability**: potential derived from simple molecules can be used and generalized for larges complexes and different environments

Molecular mechanics potentials

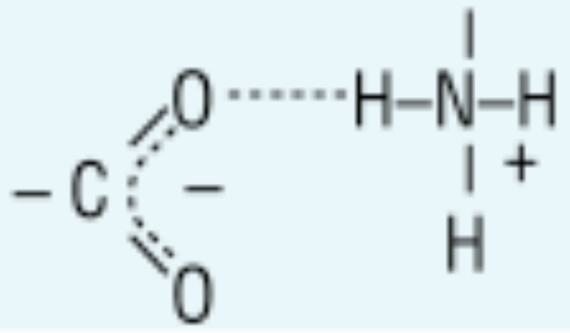
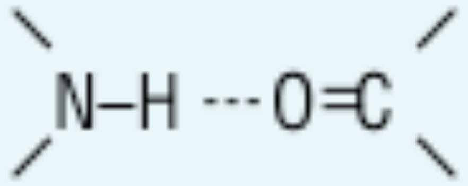
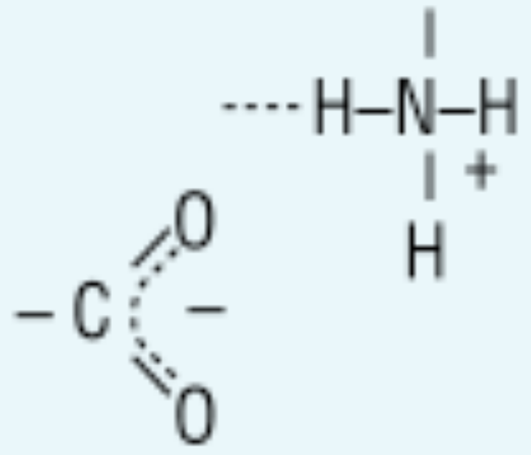
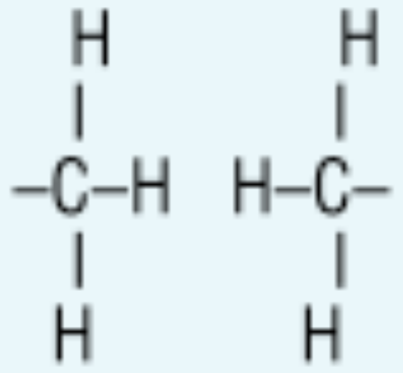
- **molecular mechanics** (MM) potential energy gives minimum-energy conformation of a molecule
- based on **physics**, but uses simplified “ball-and-spring” models (**classical** physics, *Newton equation*), which mask the quantum nature (*Schrodinger equation*)
- are **empirical**, i.e. calibrated to describe the quantum nature of chemical bonds and short-range interactions



$$U_{MM} = \boxed{U_{bonded}} + \boxed{U_{non-bonded}}$$

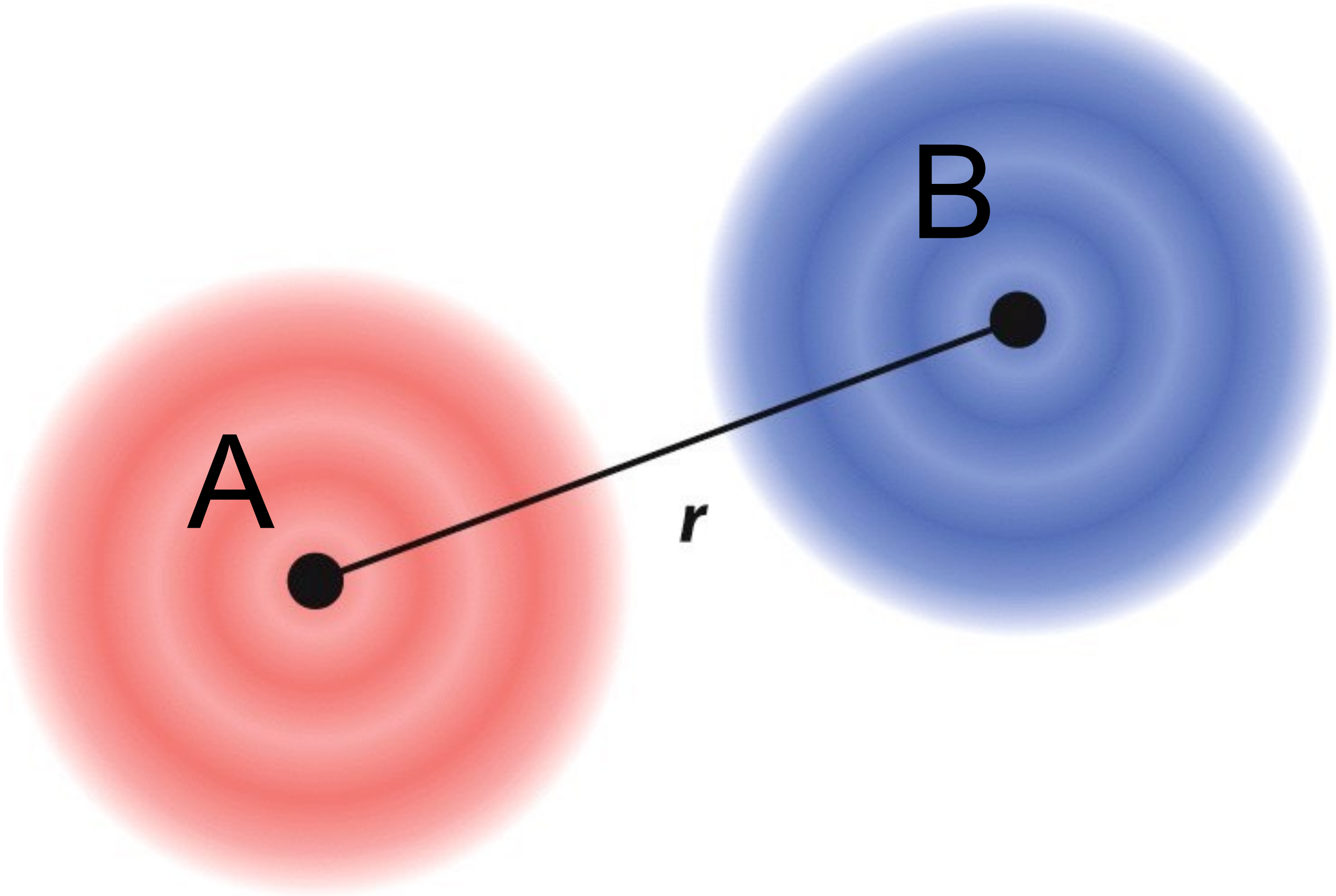
$$\left\{ \begin{aligned} U_{bonded} &= U_{bond} + U_{angle} + U_{torsion} \\ U_{non-bonded} &= U_{electrostatics} + U_{VdW} \end{aligned} \right.$$

Molecular interactions in biomolecules (bonded and non-bonded)

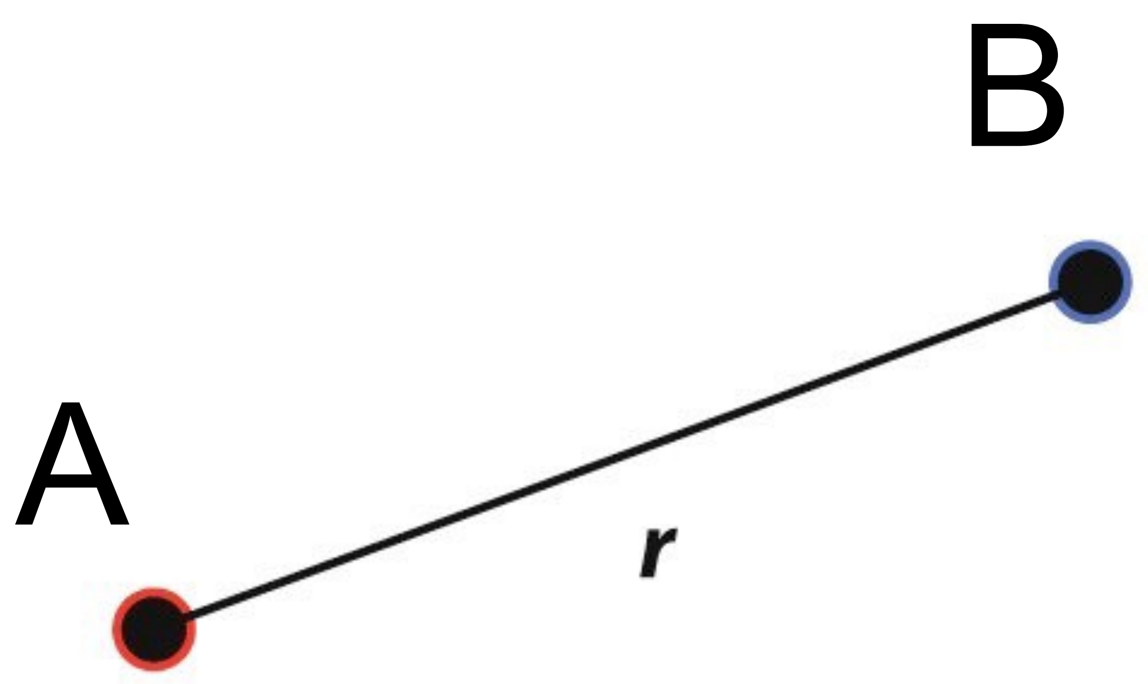
Interaction	Example	Distance dependence	Typical distance	Free energy (bond dissociation enthalpies for the covalent bonds)
Covalent bond	$-C_{\alpha}-C-$	-	1.5 Å	356 kJ/mole (610 kJ/mole for a C=C bond)
Disulfide bond	–Cys–S–S–Cys–	-	2.2 Å	167 kJ/mole
Salt bridge		Donor (here N), and acceptor (here O) atoms <3.5 Å	2.8 Å	12.5–17 kJ/mole; may be as high as 30 kJ/mole for fully or partially buried salt bridges (see text), less if the salt bridge is external
Hydrogen bond		Donor (here N), and acceptor (here O) atoms <3.5 Å	3.0 Å	2–6 kJ/mole in water; 12.5–21 kJ/mole if either donor or acceptor is charged
Long-range electrostatic interaction		Depends on dielectric constant of medium. Screened by water. 1/r dependence	Variable	Depends on distance and environment. Can be very strong in nonpolar region but very weak in water
Van der Waals interaction		Short range. Falls off rapidly beyond 4 Å separation. $1/r^6$ dependence	3.5 Å	4 kJ/mole (4–17 in protein interior) depending on the size of the group (for comparison, the average thermal energy of molecules at room temperature is 2.5 kJ/mole)

k_BT(300 K) = 0.6 kcal/mole - 2.5kJ/mole - 4.1 pN nm - 4.1 * 10^{-21} J
(1 kcal/mol = 4.184 kJ/mole)

Potential energy for biomolecules



bond potential

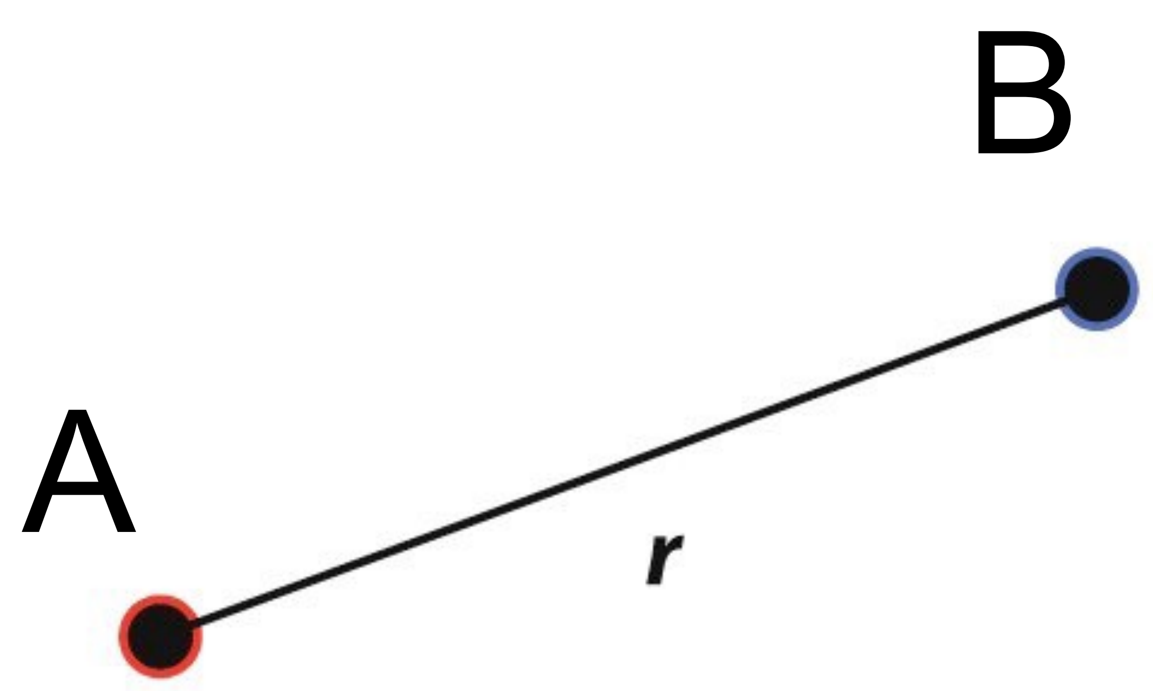


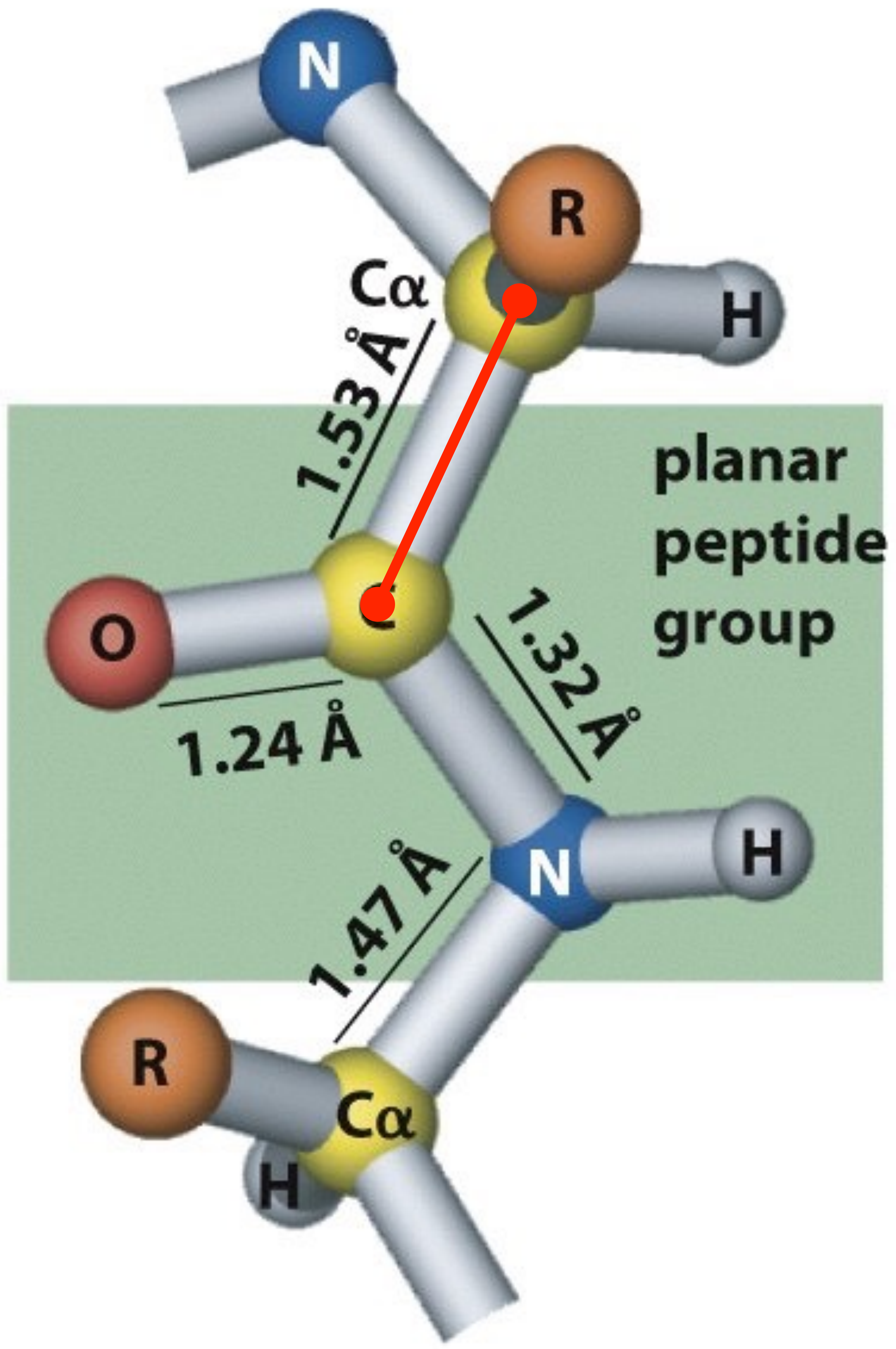
- nuclei and electrons are described by **quantum mechanics** using the *Schrodinger equation*

$$H\Psi(\mathbf{r},\mathbf{R}) = E\Psi(\mathbf{r},\mathbf{R})$$

- simplified “ball-and-spring” models following the laws of **classical physics** can be used to “mask” the quantum nature and define a **energy potential** U(**R**)

Potential energy for a chemical bond





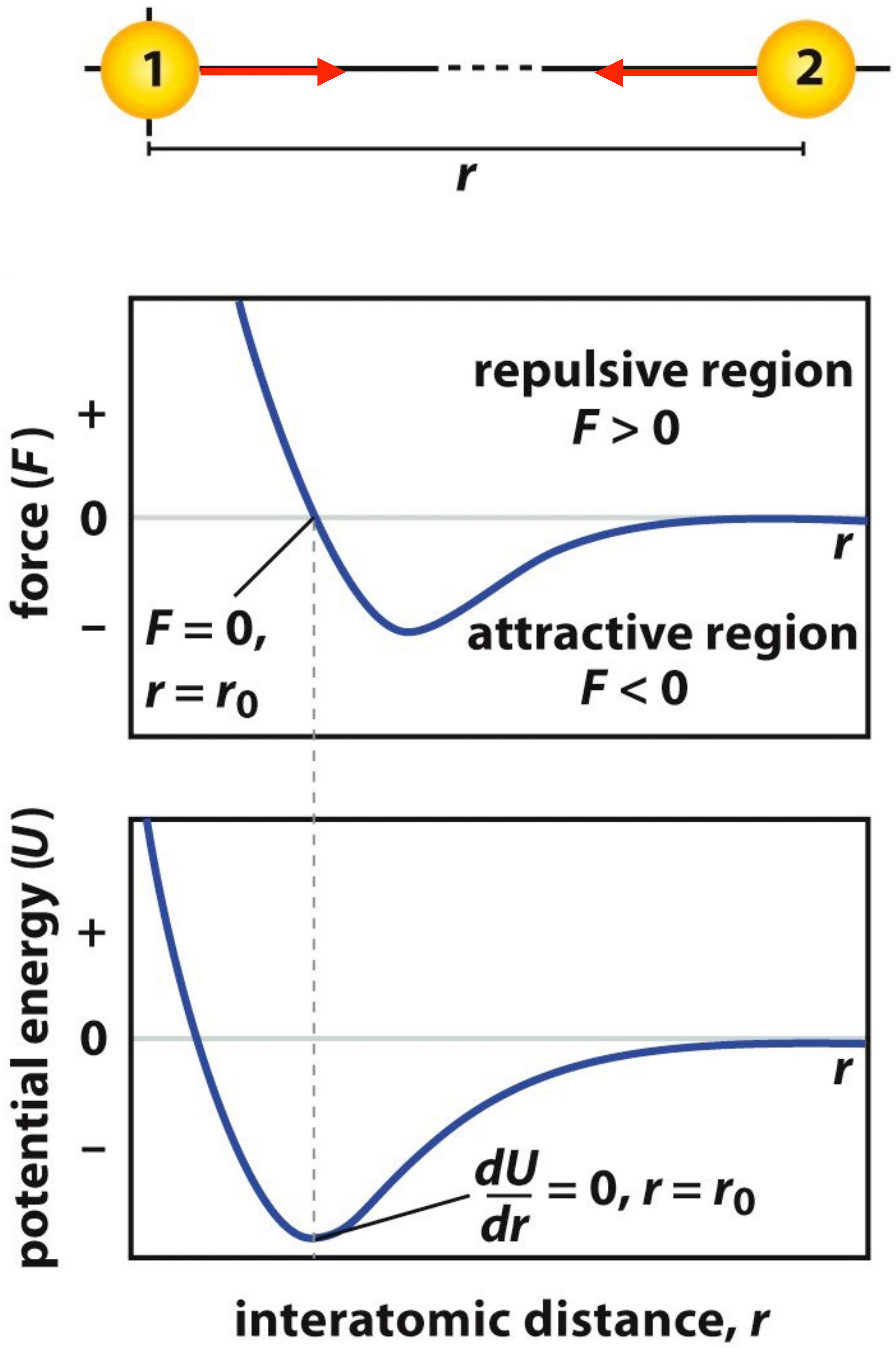
how such potential would look like?

➡ at very large distance r ?

➡ at short distance?

➡ at equilibrium distance, r_0 ?

Potential energy and forces



- potential energy is the energy stored in the system
- a simple case is when U depends on distance ***r*** only
- *U* can have complex forms, from which forces on each atom are derived:

$$F(r) = -\frac{dU(r)}{dr}$$

$$U(r) = -\int_{\infty}^{r} F(\widetilde{r})d\widetilde{r}$$

Covalent bond potential energy function

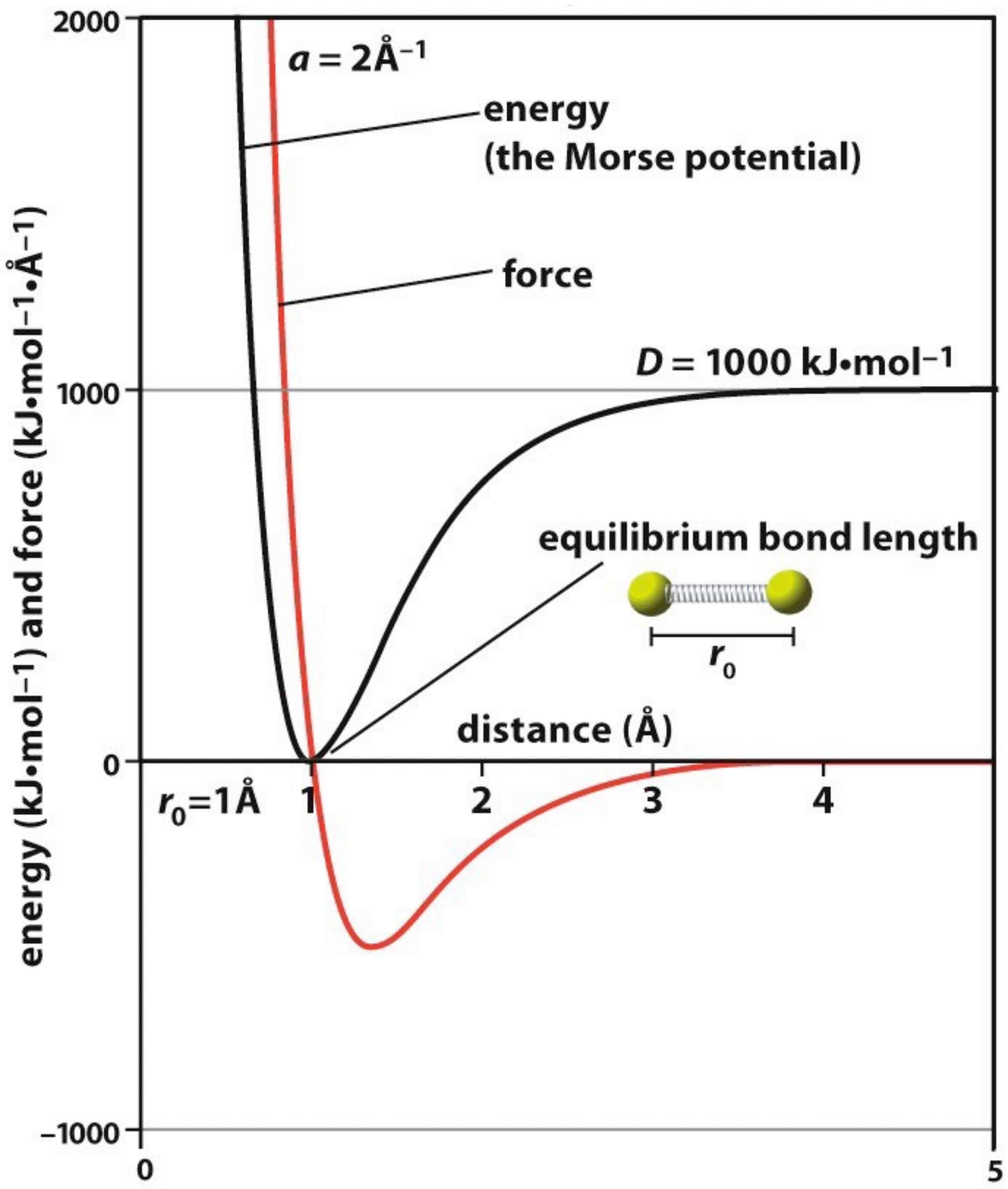
$$U_{bond}(r) = D[1 - e^{-a(r-r_0)}]^2$$

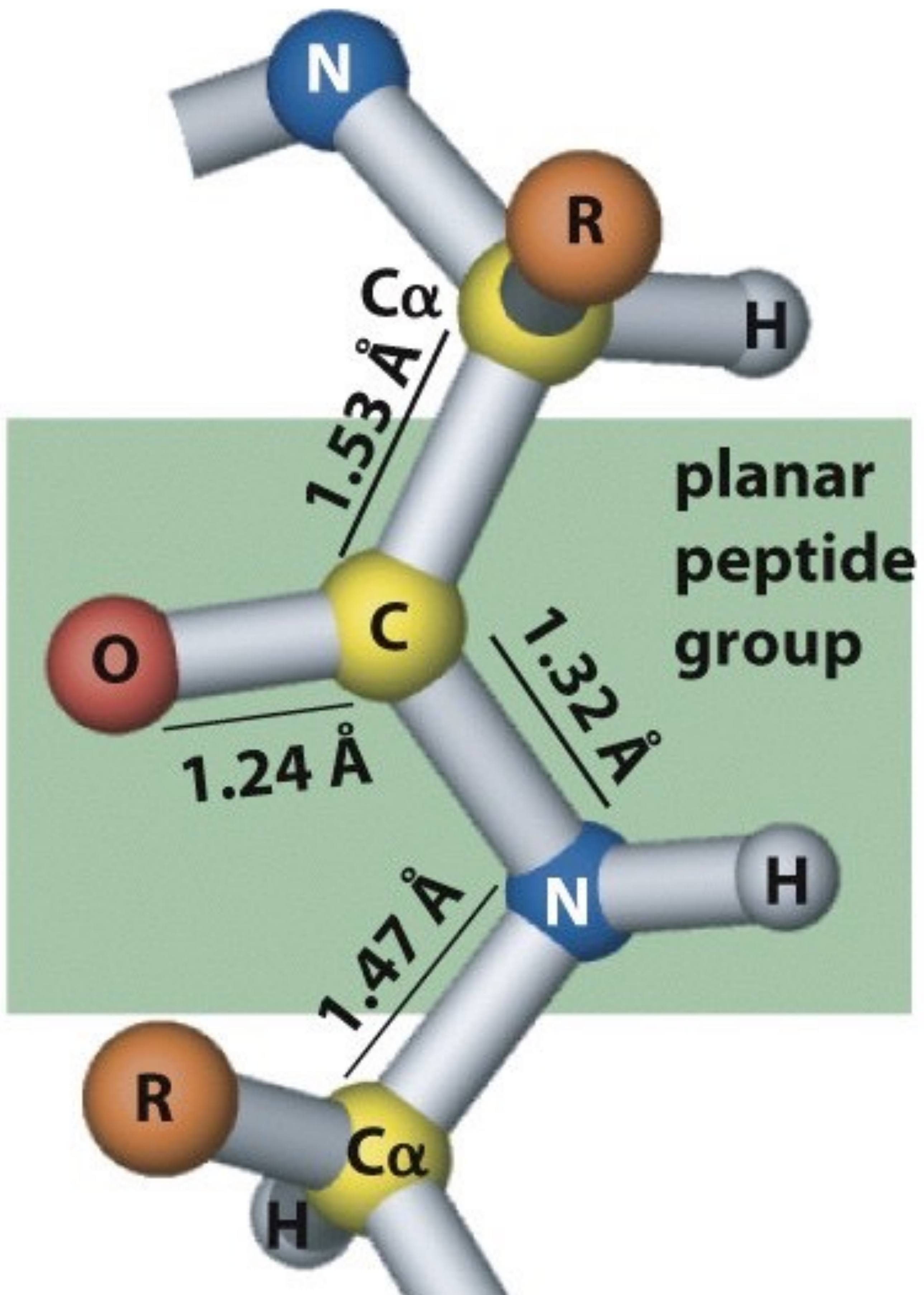
Morse potential

D: dissociation energy

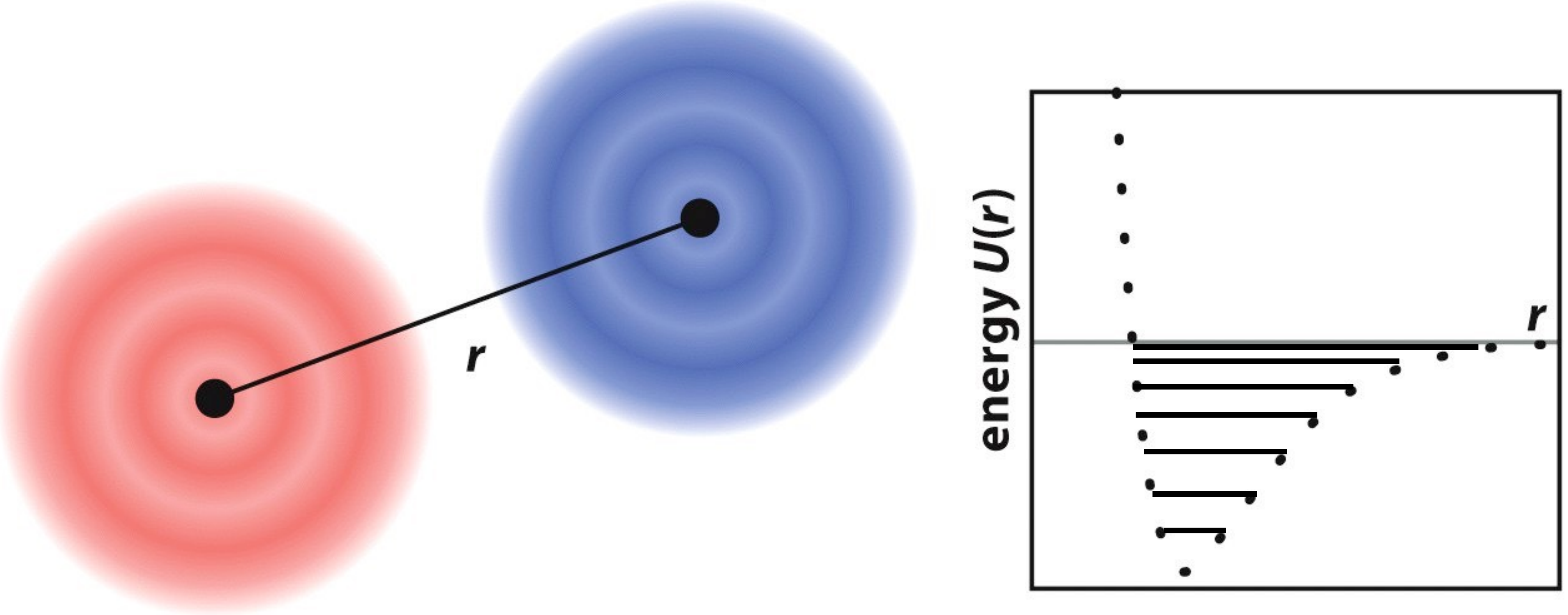
r₀: equilibrium bond distance

a: force constant





Empirical potential energy functions

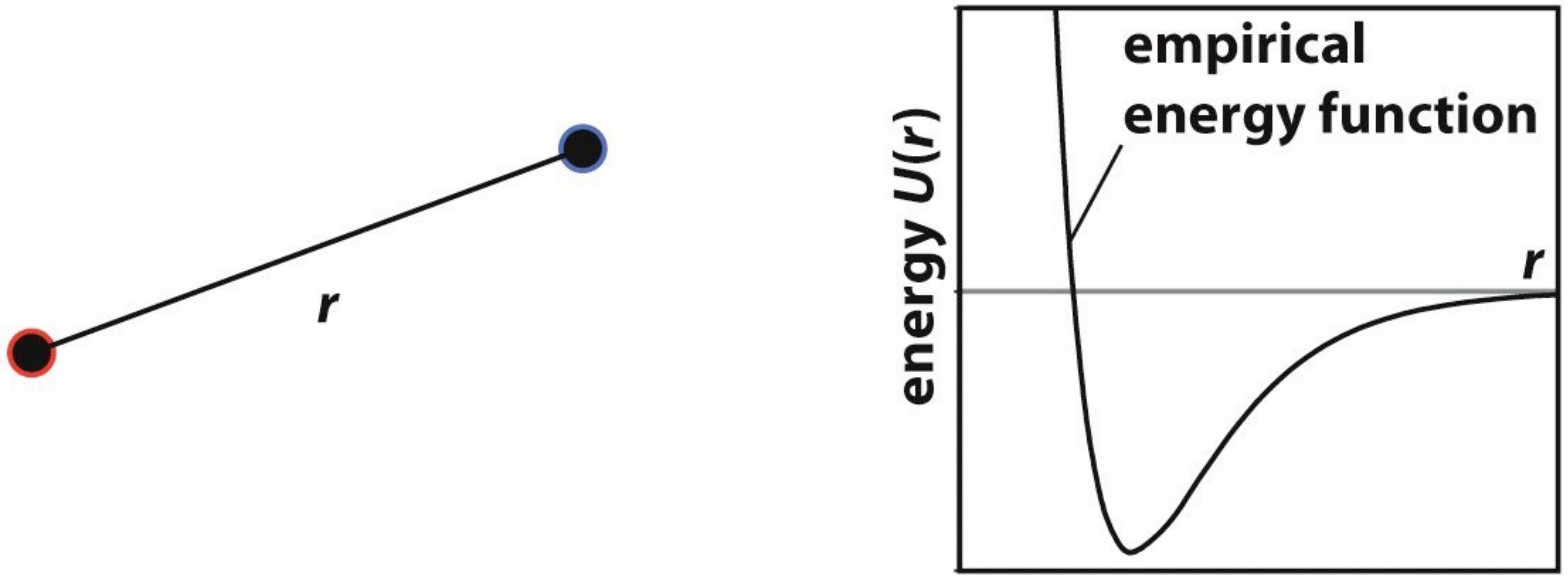


- the **quantum mechanical** nature of matter predicts discrete energy ladders

$$\Delta E = h\nu$$

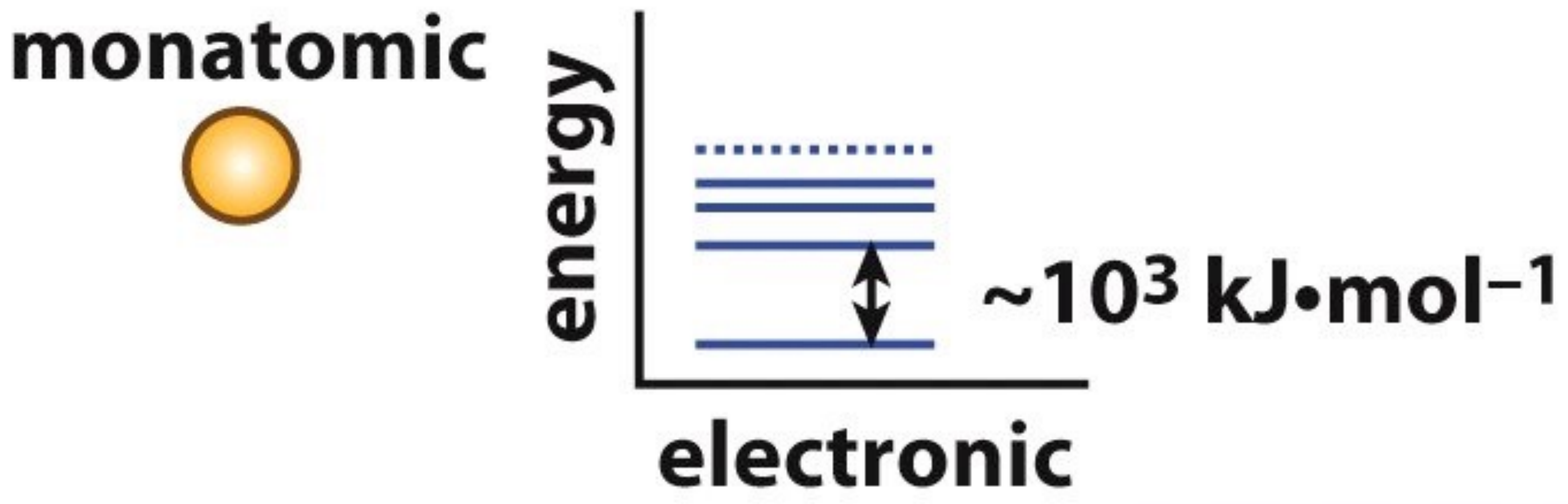
h: Planck's constant;

ν: wavenumber (cm^{-1})



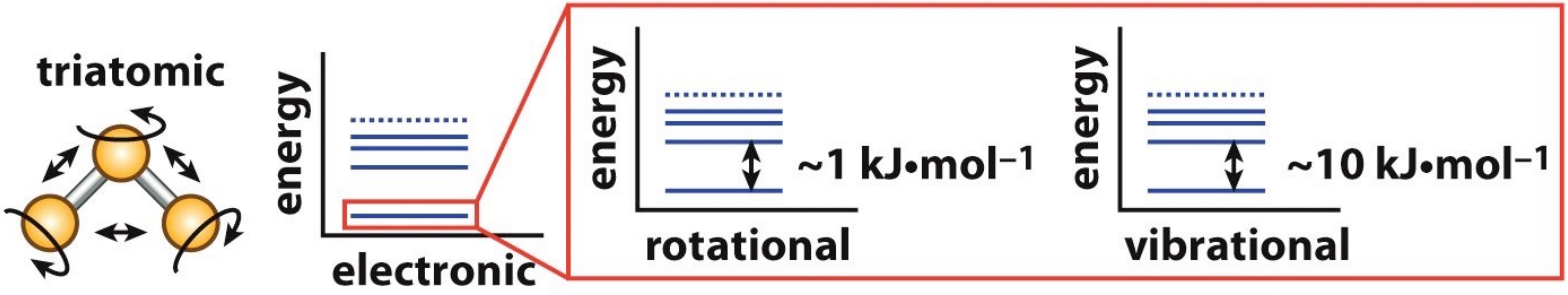
- but **empirical** continuum models describe chemical bonds with good accuracy (at least close to equilibrium)

Internal modes of biomolecules



- excited electronic states are accessible at a high energy cost (~10^3 kcal/mol)

- but biomolecules have a finer ladder of E levels associated with internal vibrations and rotations: few kcal/mol ~ $k_B T=2.5$ *KJ/mol*



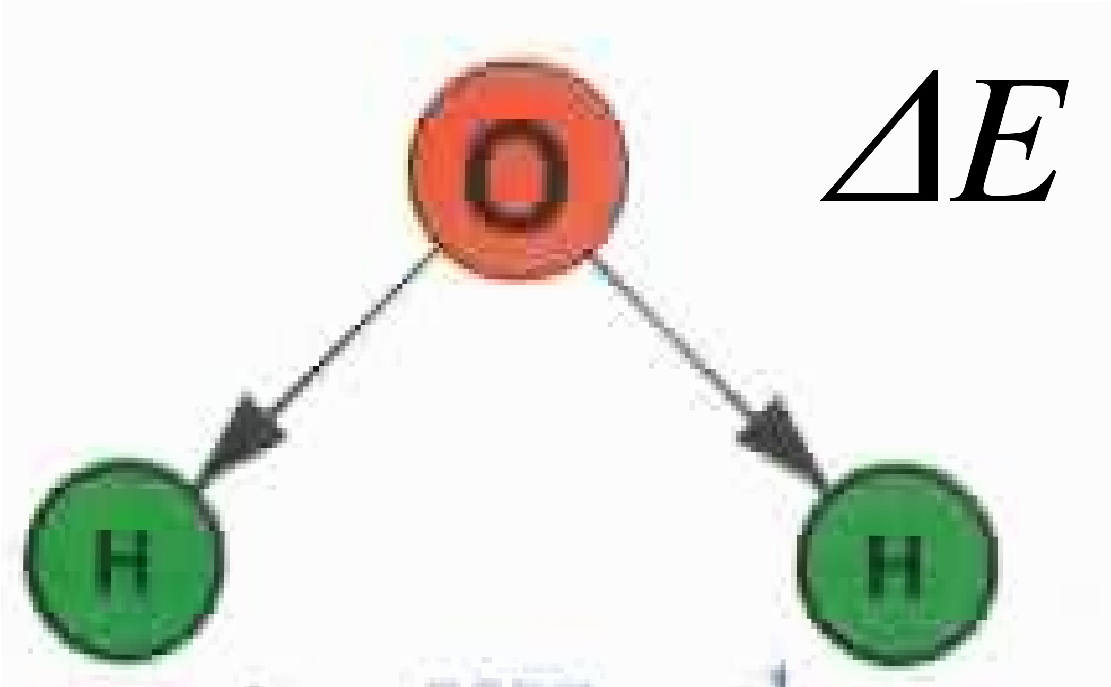
Vibrational modes

- vibrational spectra give information about **bonded** interactions (3N-6 modes, with N number of atoms)

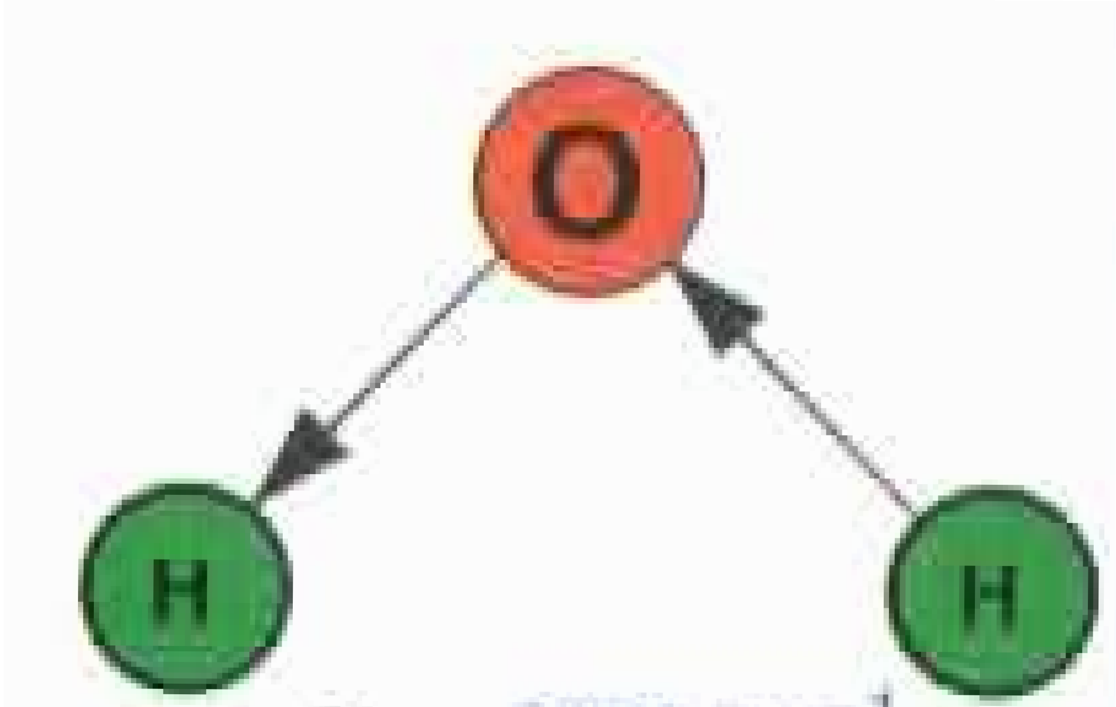
- derived from infrared spectroscopy (IR: 300-3000 cm^{-1}), Raman spectroscopy or QM calculations

- the higher the wavenumber the more difficult the deformation, but also depends on the environment and mass atoms

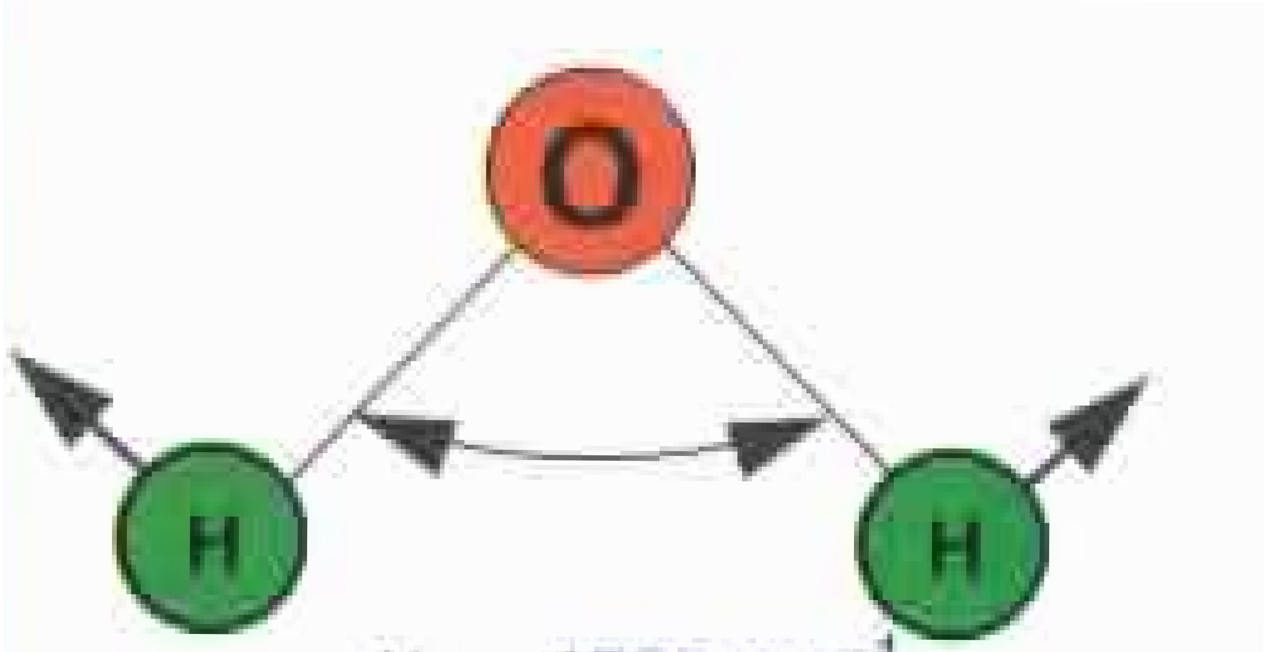
$$\Delta E = h\nu$$



$\nu_1 = 3657\ \mathrm{cm}^{-1}$
(symmetric stretch)



$\nu_2 = 3776\ \mathrm{cm}^{-1}$
(asymmetric stretch)



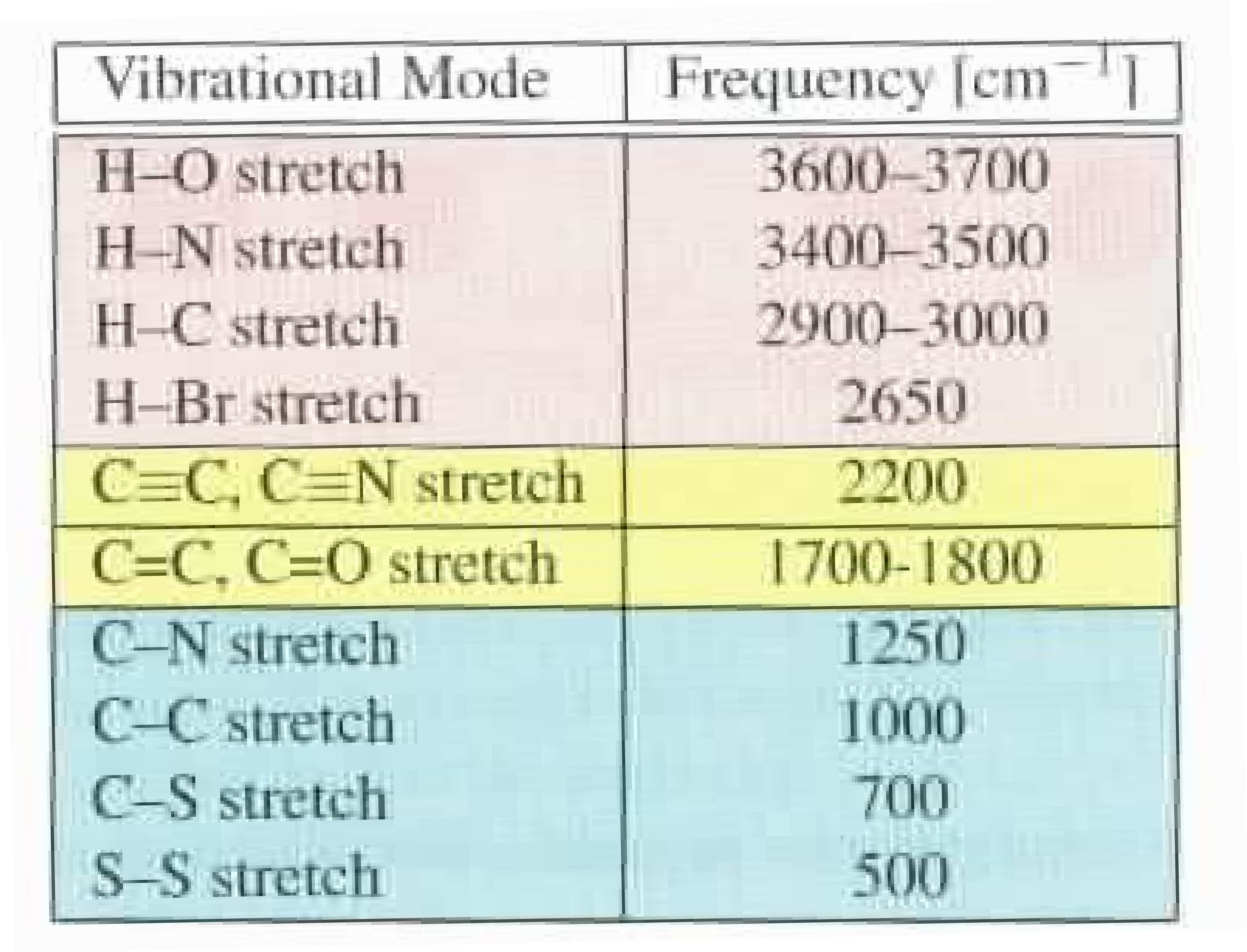
$\nu_3 = 1595\ \mathrm{cm}^{-1}$
(bend)

water vibrational modes

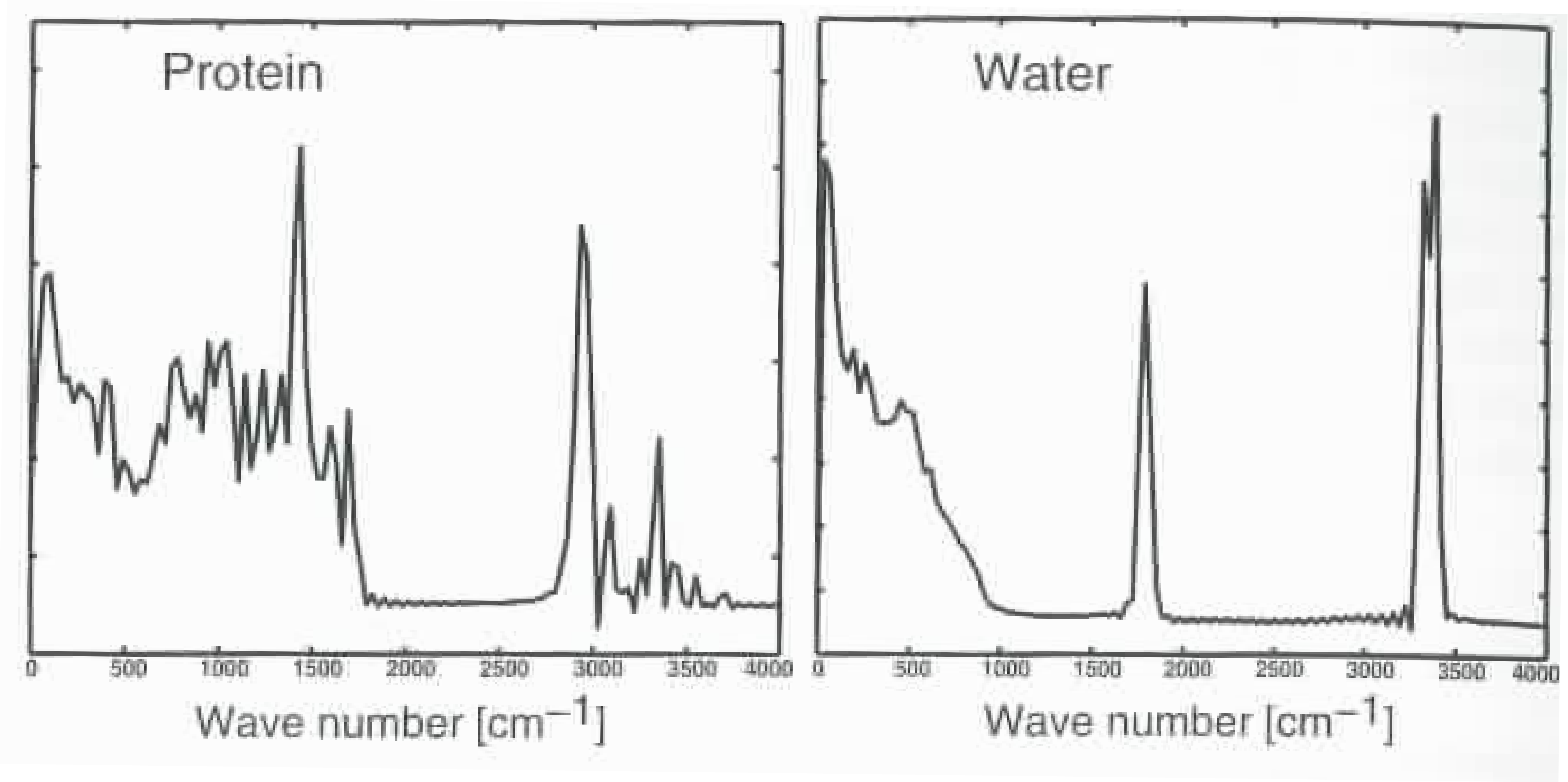
Vibrational modes for bonds

- frequencies used to derive **force constants** for potential energy models

- **stretching frequencies**

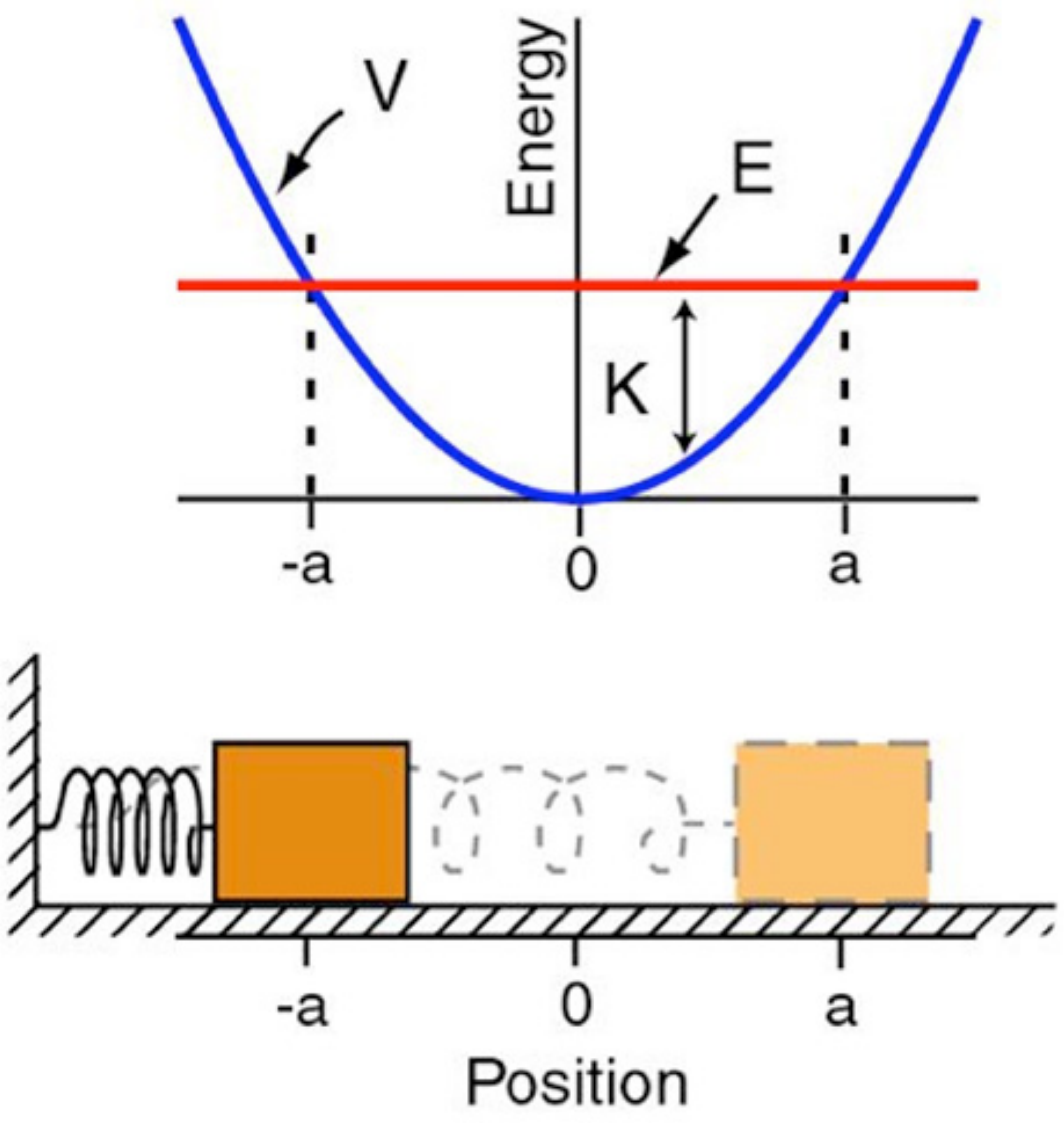


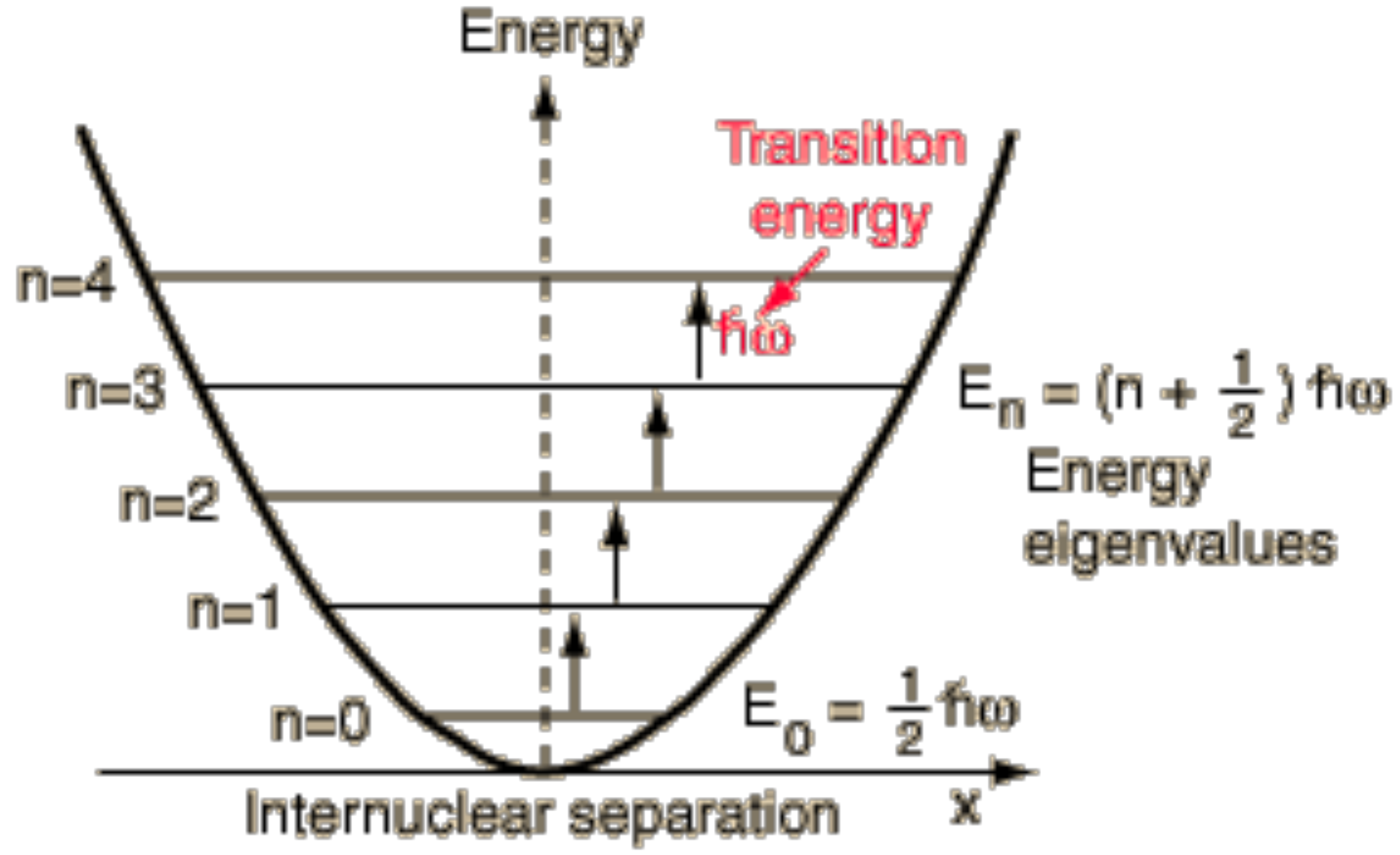
Vibrational Mode	Frequency [cm^{-1}]
H–O stretch	3600–3700
H–N stretch	3400–3500
H–C stretch	2900–3000
H–Br stretch	2650
C≡C, C≡N stretch	2200
C=C, C=O stretch	1700-1800
C–N stretch	1250
C–C stretch	1000
C–S stretch	700
S–S stretch	500



example of IR vibrational spectra from protein and water

Vibrational modes





- vibrational degrees of freedom can be modeled as harmonic oscillators, following **Hooke's law**

$$\boldsymbol{F} = -k\,\boldsymbol{x}$$

- the potential energy associated is:

$$U(\boldsymbol{x}) = k/2\ \boldsymbol{x}^2$$

- if one solves the Schrodinger equation with this potential

$$E_\upsilon = (n+1/2)\, h\nu\ ;\ n = 0, 1, 2\ \ldots$$

quantum number

$\nu = 1/2\pi\,(k/m)^{1/2}$ **wavenumber** [cm^{-1}]

$\omega = \sqrt{\dfrac{k}{m}} = 2\pi\,\nu$ **angular frequency**

Harmonic approx for covalent bonds

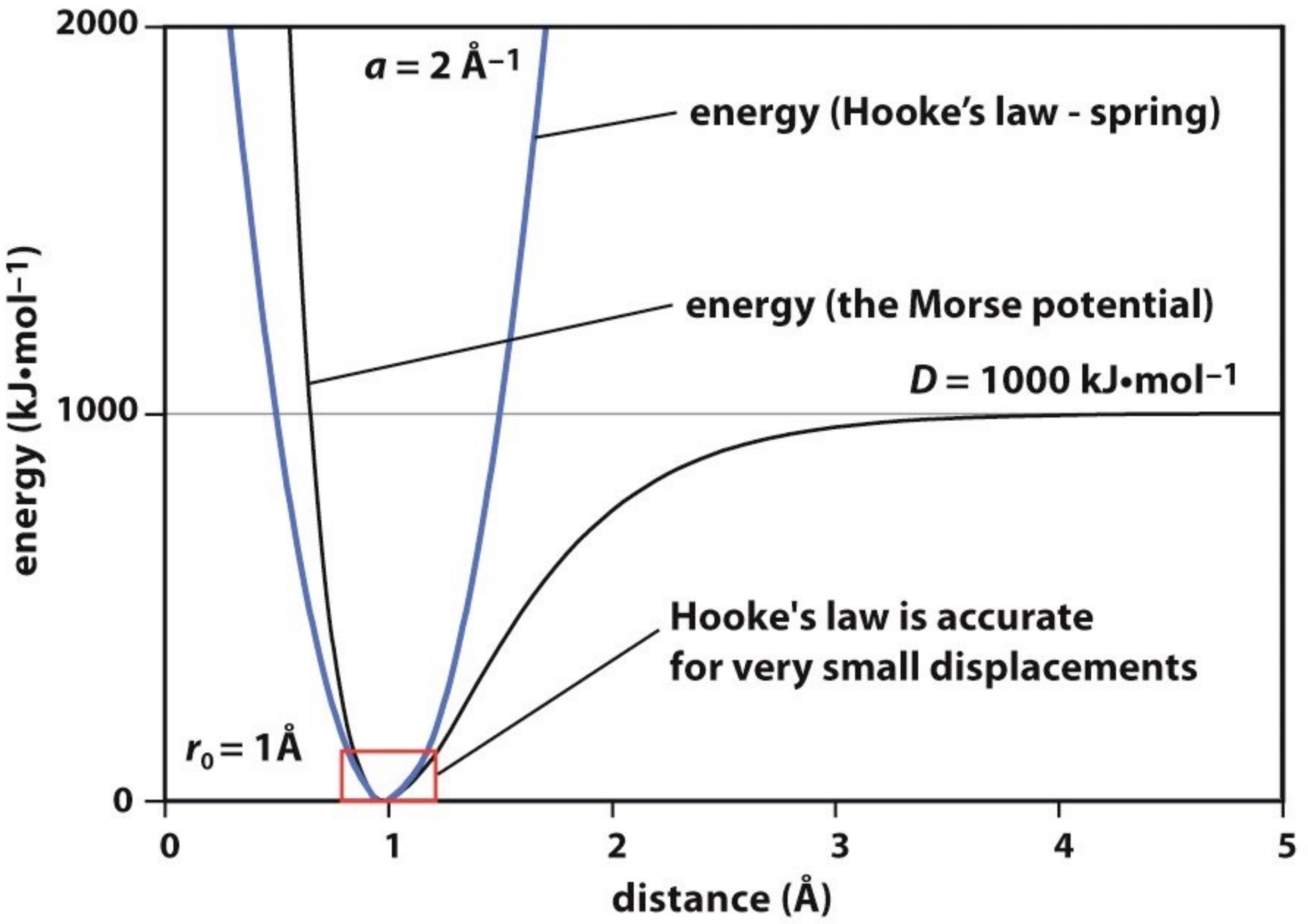
$$U_{bond}(r) = \frac{k_b}{2}(r - r_0)^2$$

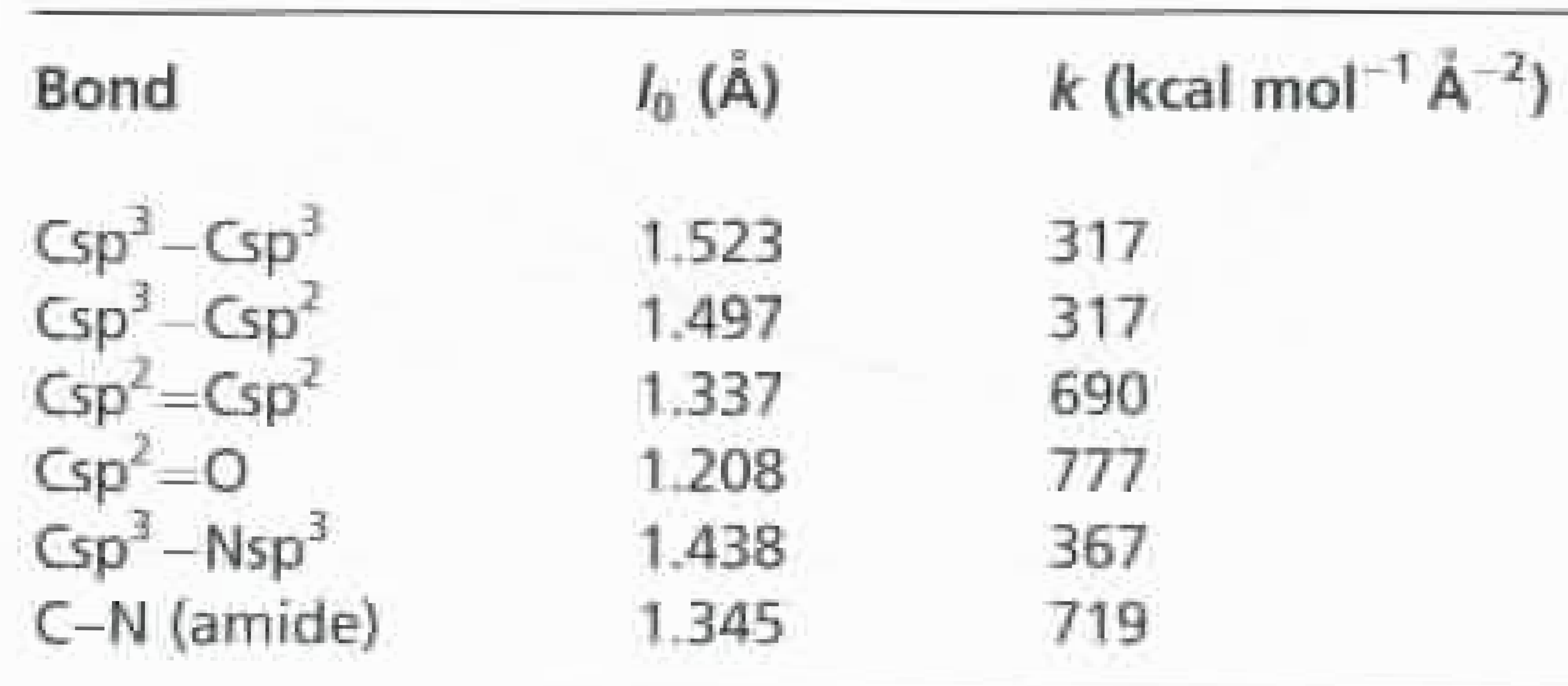
- equilibrium bond lengths are obtained from X-ray or QM calculations (r_0)
- deviation from reference value (r_0) i modeled as an **harmonic** potential
- k_b is the **force constant**

$$k_b = (2\pi\nu)^2\mu = \omega^2\mu$$

$$\mu = m_1 m_2/(m_1 + m_2)$$

- reasonable approximation for small deviation from r_0 (~0.1 Å), dissociation for larger values



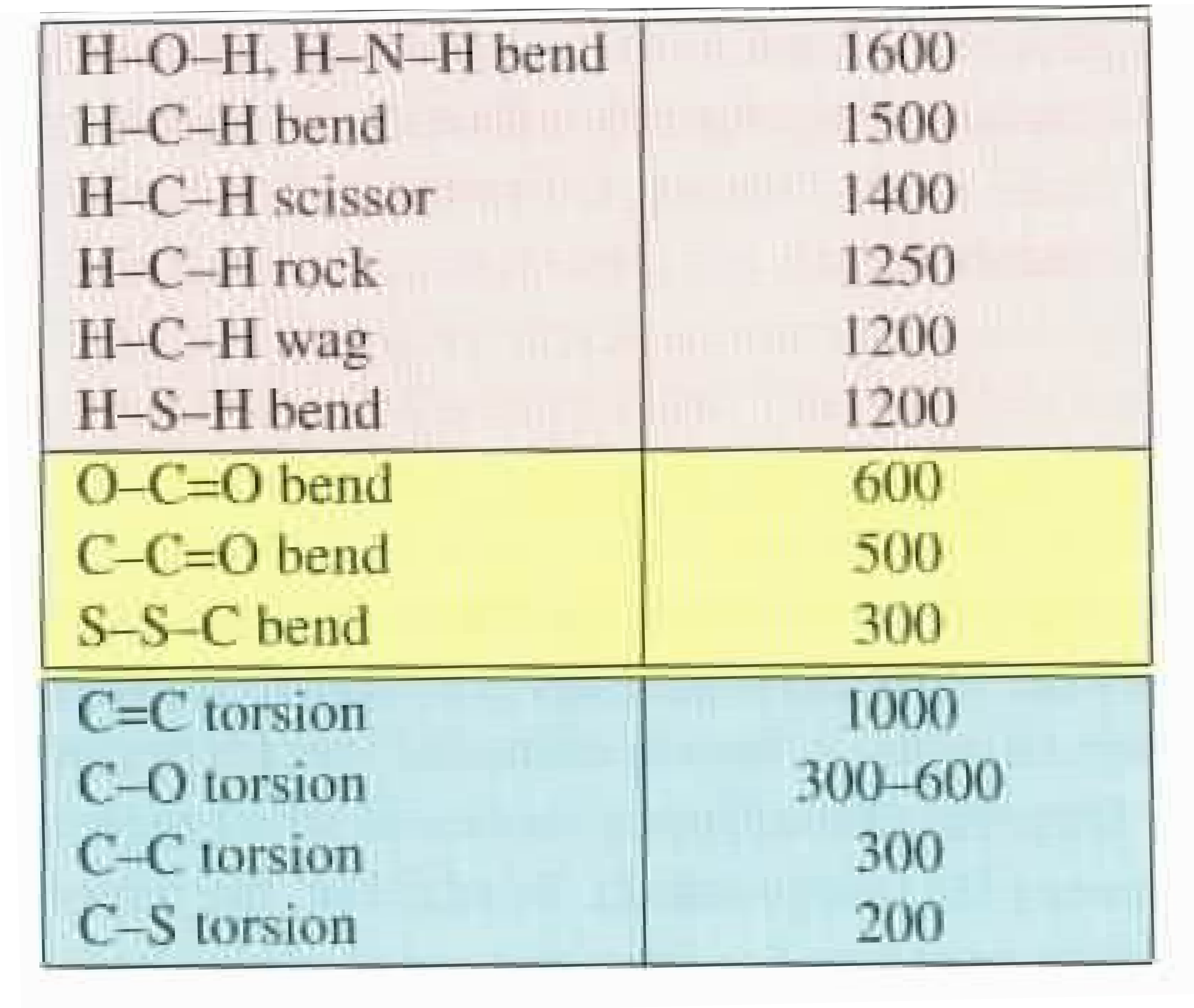


Bond	l_0 (Å)	k (kcal mol^{-1} Å^{-2})
Csp^3–Csp^3	1.523	317
Csp^3–Csp^2	1.497	317
Csp^2=Csp^2	1.337	690
Csp^2=O	1.208	777
Csp^3–Nsp^3	1.438	367
C–N (amide)	1.345	719

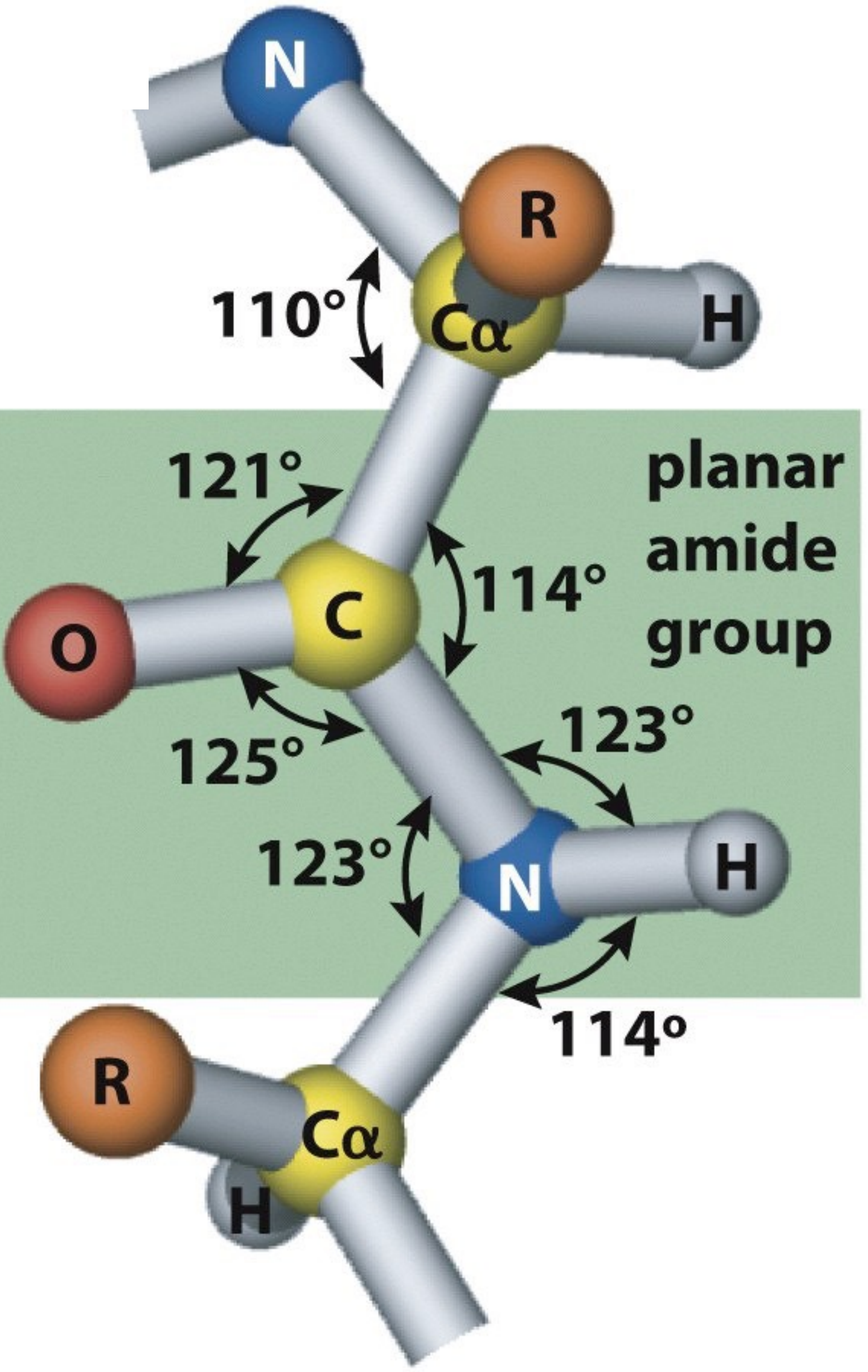
Bending potential energy functions

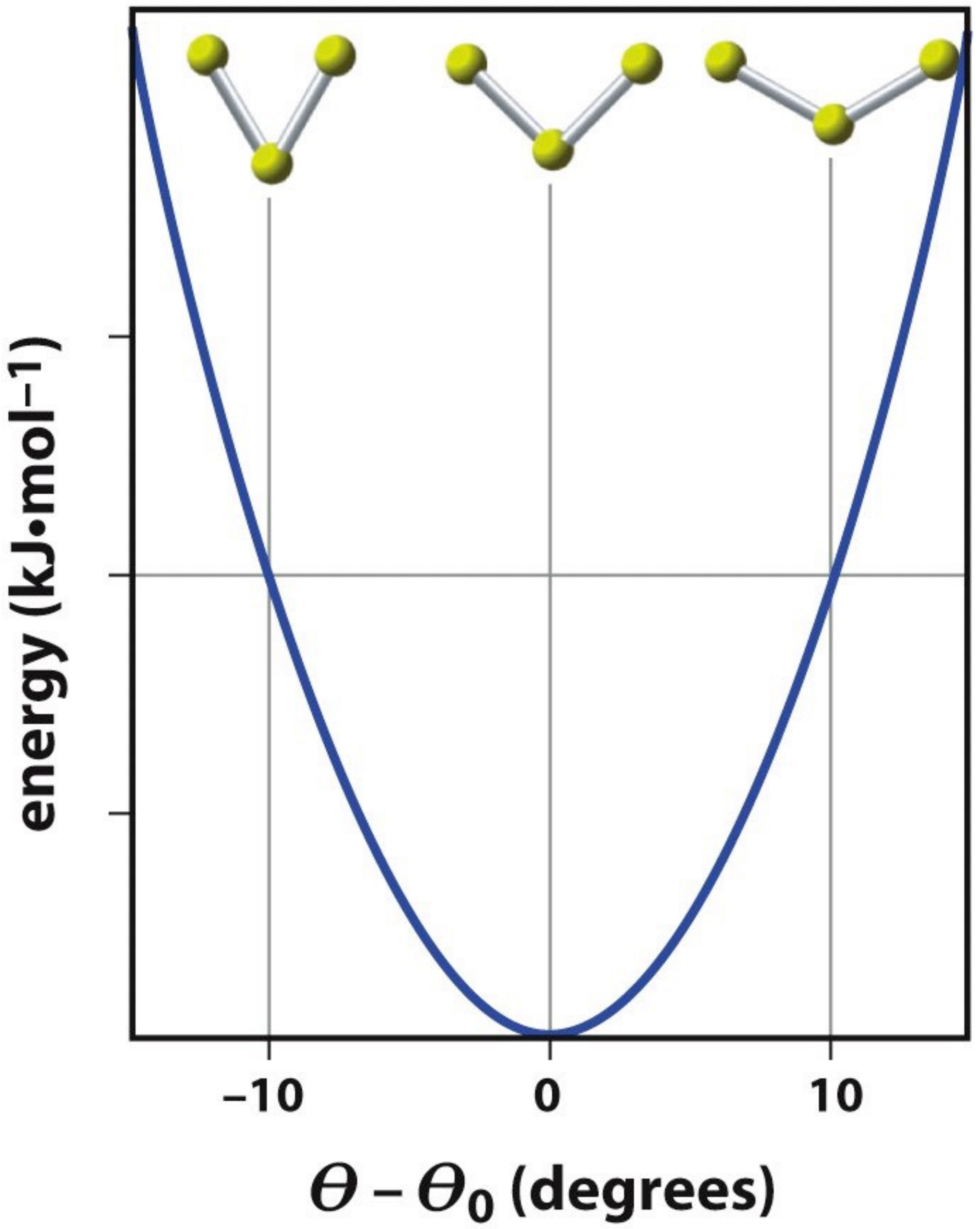
- angles depend on hybridization of electronic **orbitals**
- rule of thumb: sp:~180°; sp^2:~120°; sp^3:~109.5°

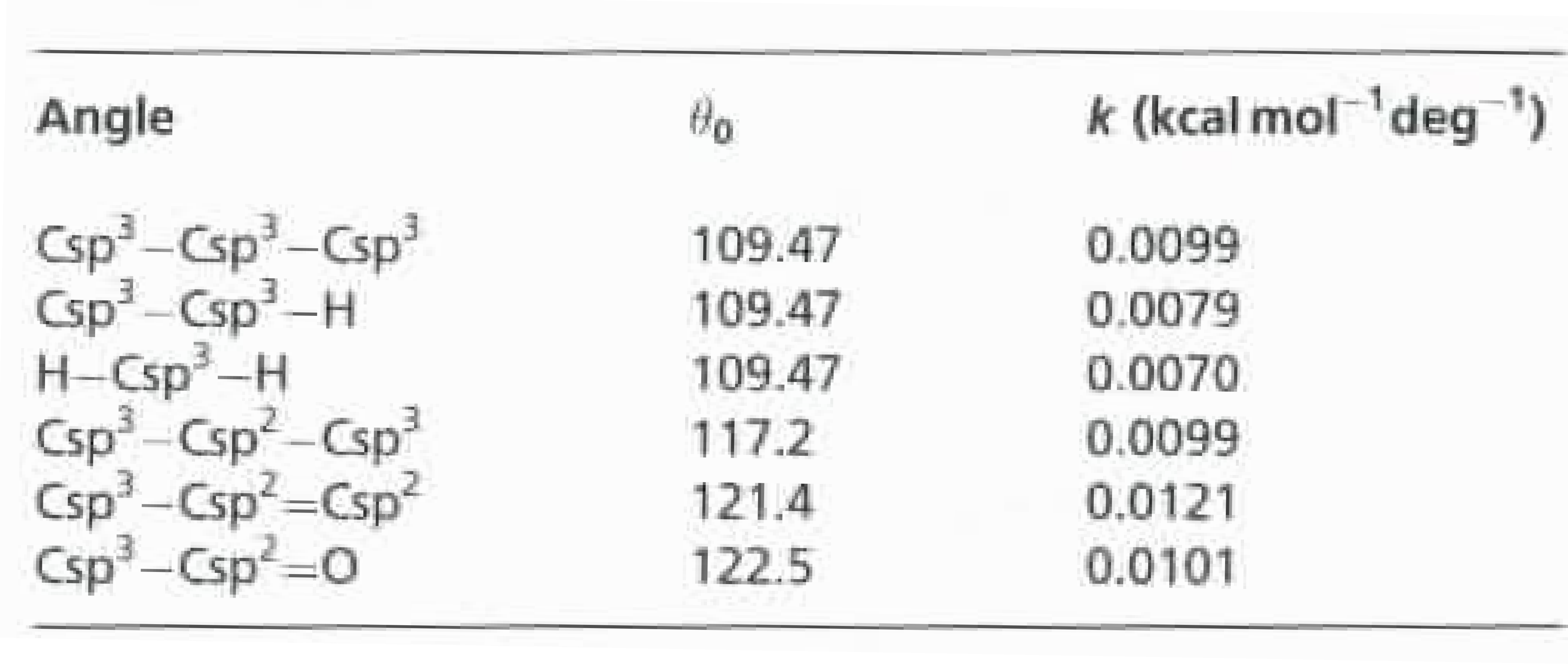
$$U_{angle}(\theta) = \frac{k_\theta}{2}(\theta - \theta_0)^2$$



H–O–H, H–N–H bend	1600
H–C–H bend	1500
H–C–H scissor	1400
H–C–H rock	1250
H–C–H wag	1200
H–S–H bend	1200
O–C=O bend	600
C–C=O bend	500
S–S–C bend	300
C=C torsion	1000
C–O torsion	300–600
C–C torsion	300
C–S torsion	200







Angle	θ_0	k (kcal mol^{-1}deg^{-1})
Csp^3–Csp^3–Csp^3	109.47	0.0099
Csp^3–Csp^3–H	109.47	0.0079
H–Csp^3–H	109.47	0.0070
Csp^3–Csp^2–Csp^3	117.2	0.0099
Csp^3–Csp^2=Csp^2	121.4	0.0121
Csp^3–Csp^2=O	122.5	0.0101

Cross stretch/bend terms

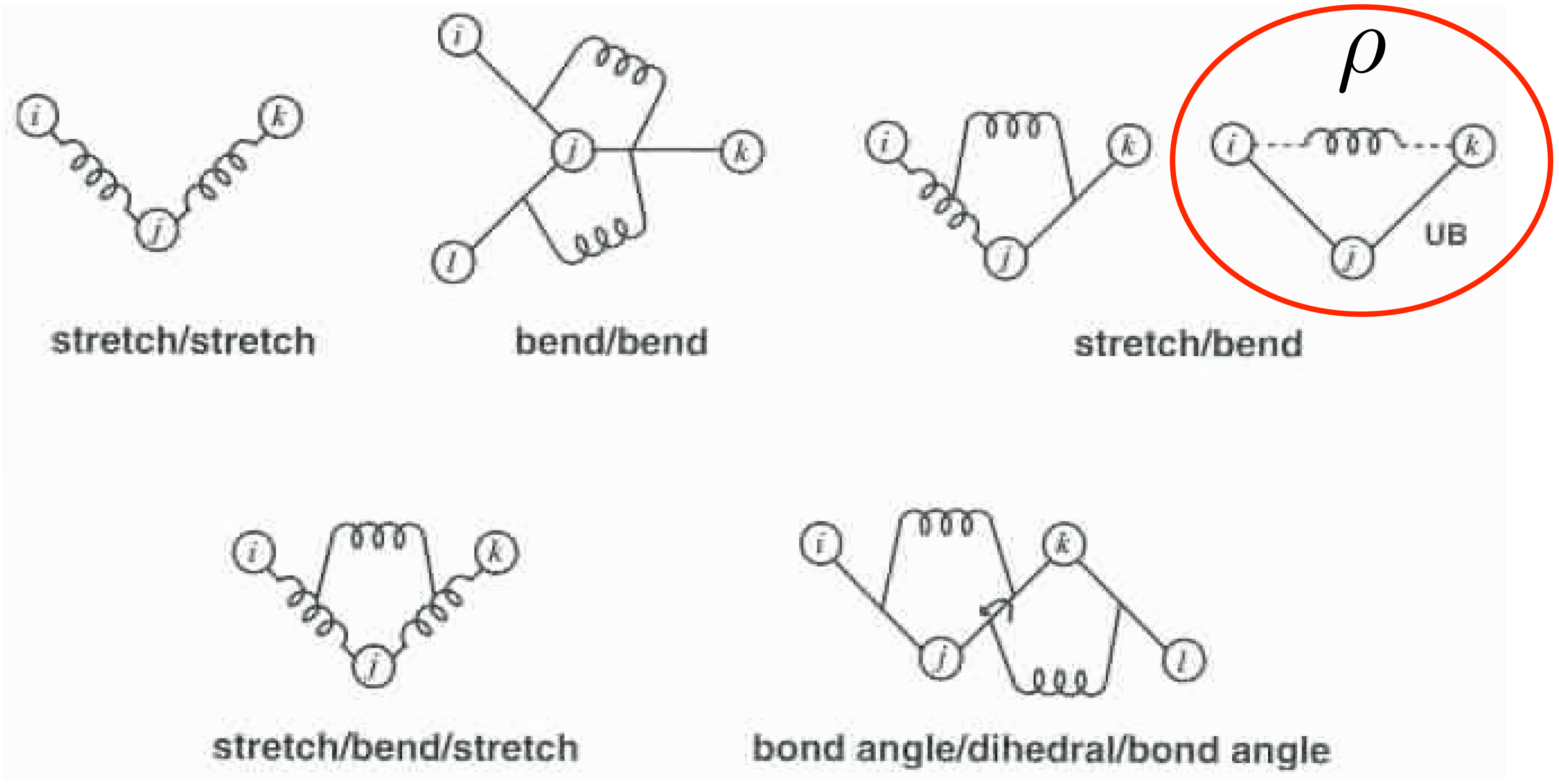
$$U_{cross}(r, r^{'}) = K_r(r - r_0)(r^{'} - r_0^{'})$$

$$U_{cross}(\theta, \theta^{'}) = K_\theta(\theta - \theta_0)(\theta^{'} - \theta_0^{'})$$

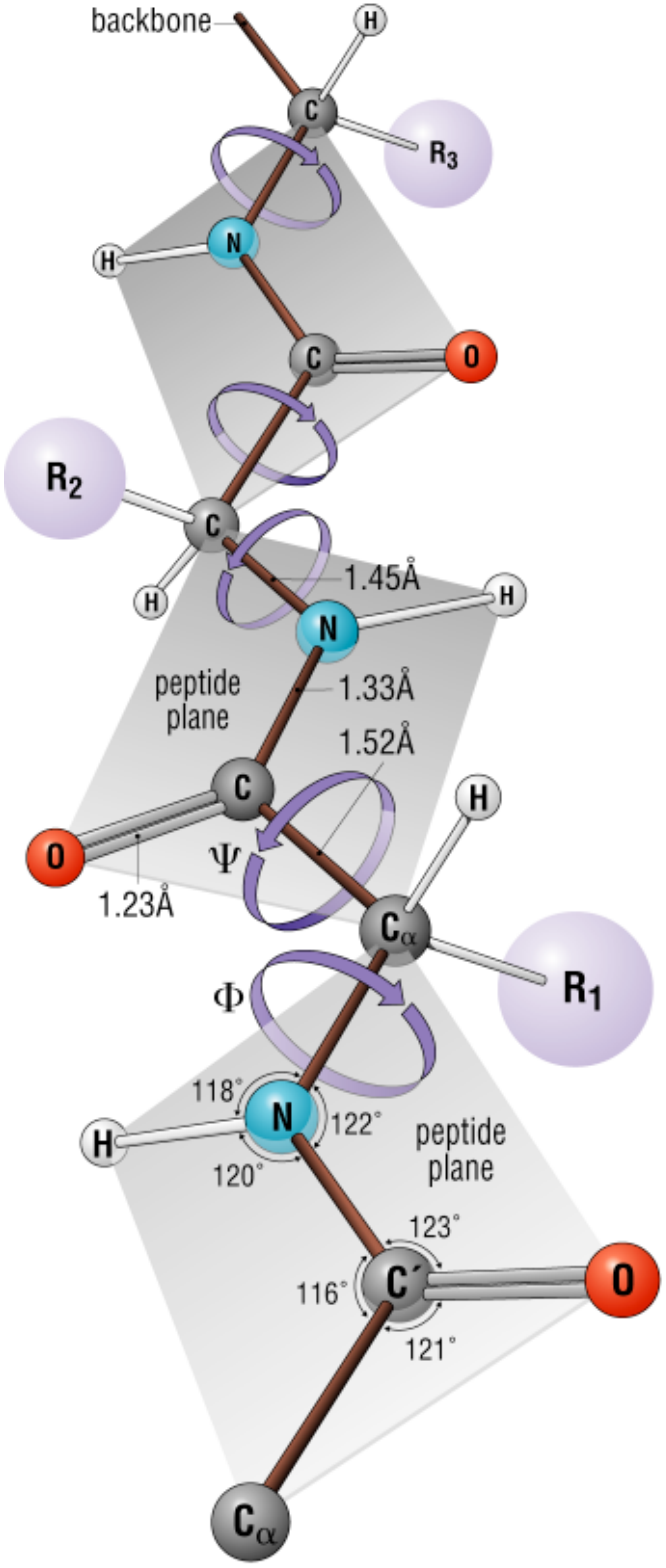
$$U_{cross}(r, \theta) = K_{r,\theta}(r - r_0)(\theta - \theta_0)$$

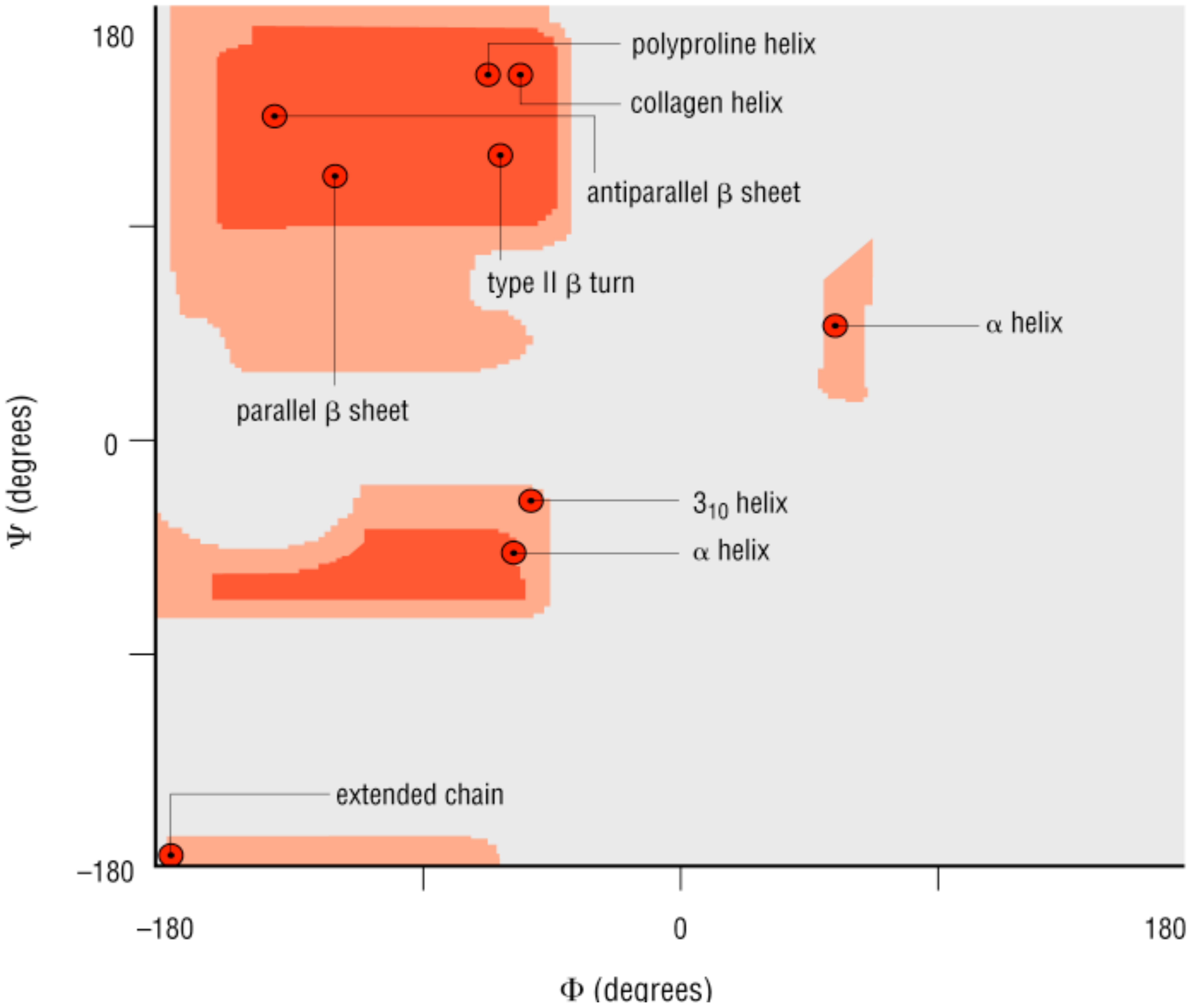
$$U_{UB}(\rho) = K_\rho(\rho - \rho_0)^2$$

Urey-Bradley term (CHARMM)



Secondary structure geometry





Ramachandran plot

Procheck: http://www.ebi.ac.uk/pdbsum/

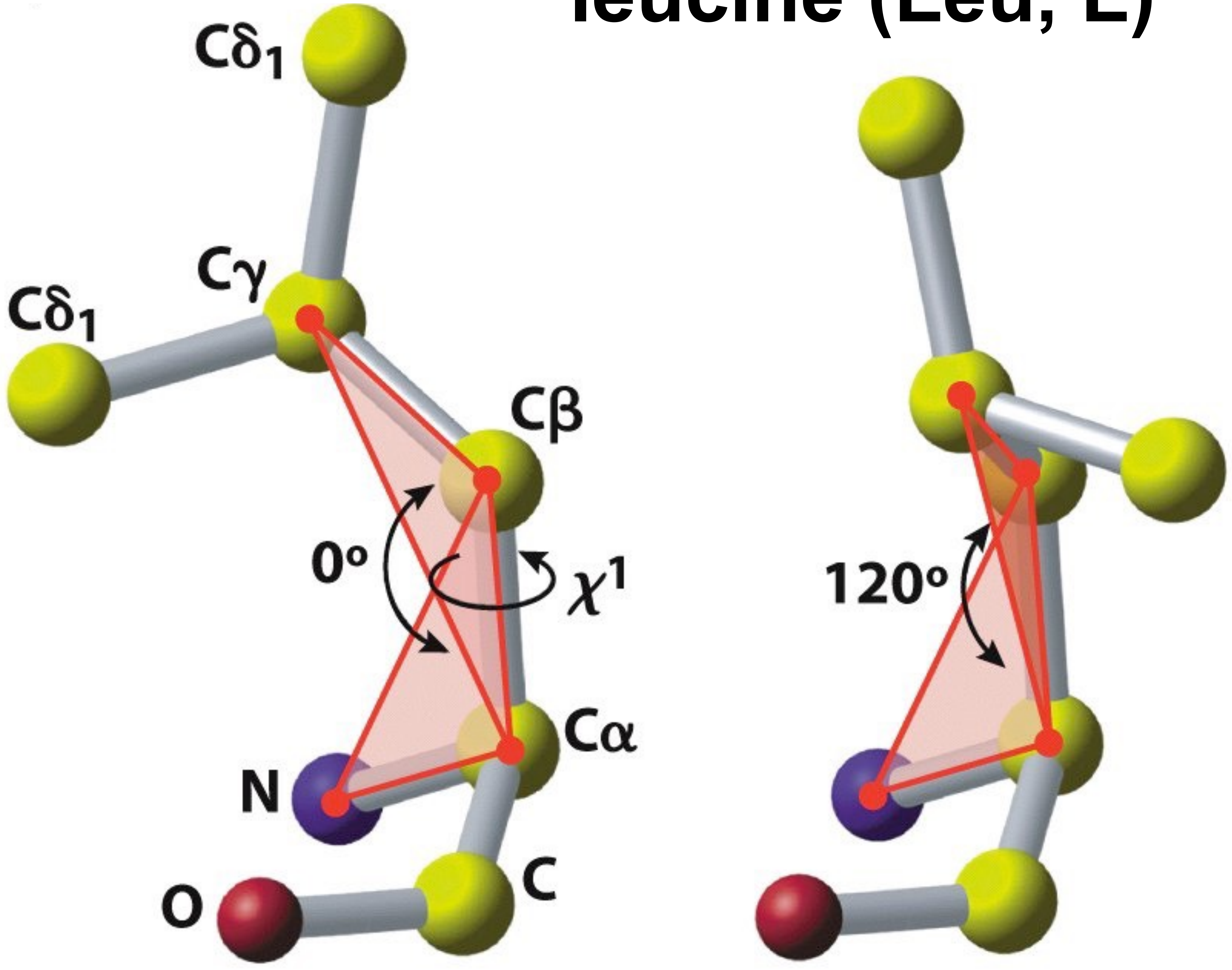
Torsional potential energy functions

- barriers for torsional rotations need a periodic potential term

$$U_{torsion}(\phi) = \sum_n \frac{k_{\phi,n}}{2}[1 + cos(n\phi - \delta)]$$

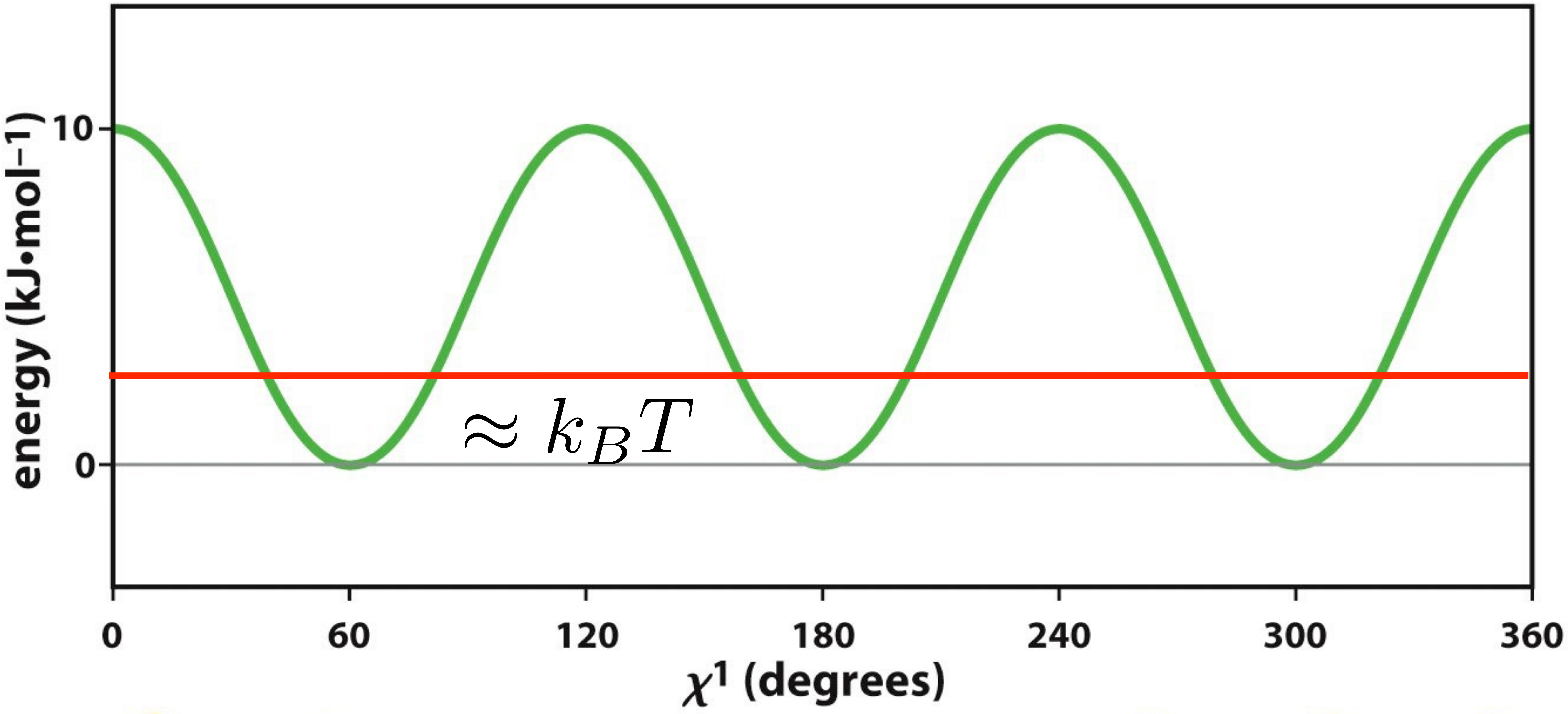
- n is the **periodicity**, k_ϕ is the barrier height, δ is called **phase** $(0, \pi)$
- Fourier series of torsional potentials (n=1,2,3,4)
- parameters are again derived from **spectroscopy** or **quantum mechanics** calculations

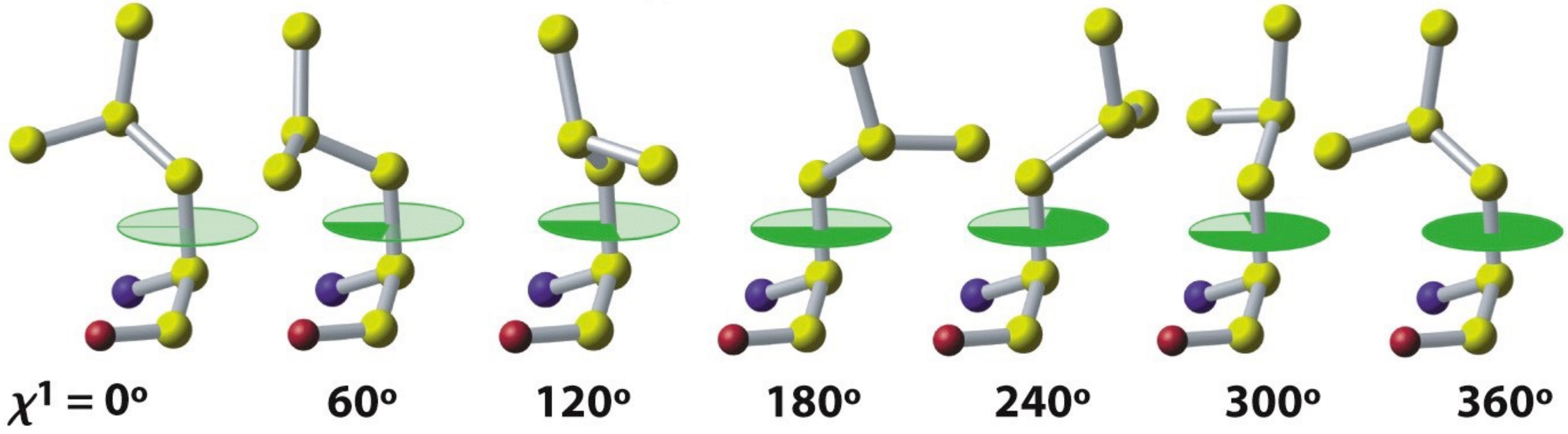
leucine (Leu, L)



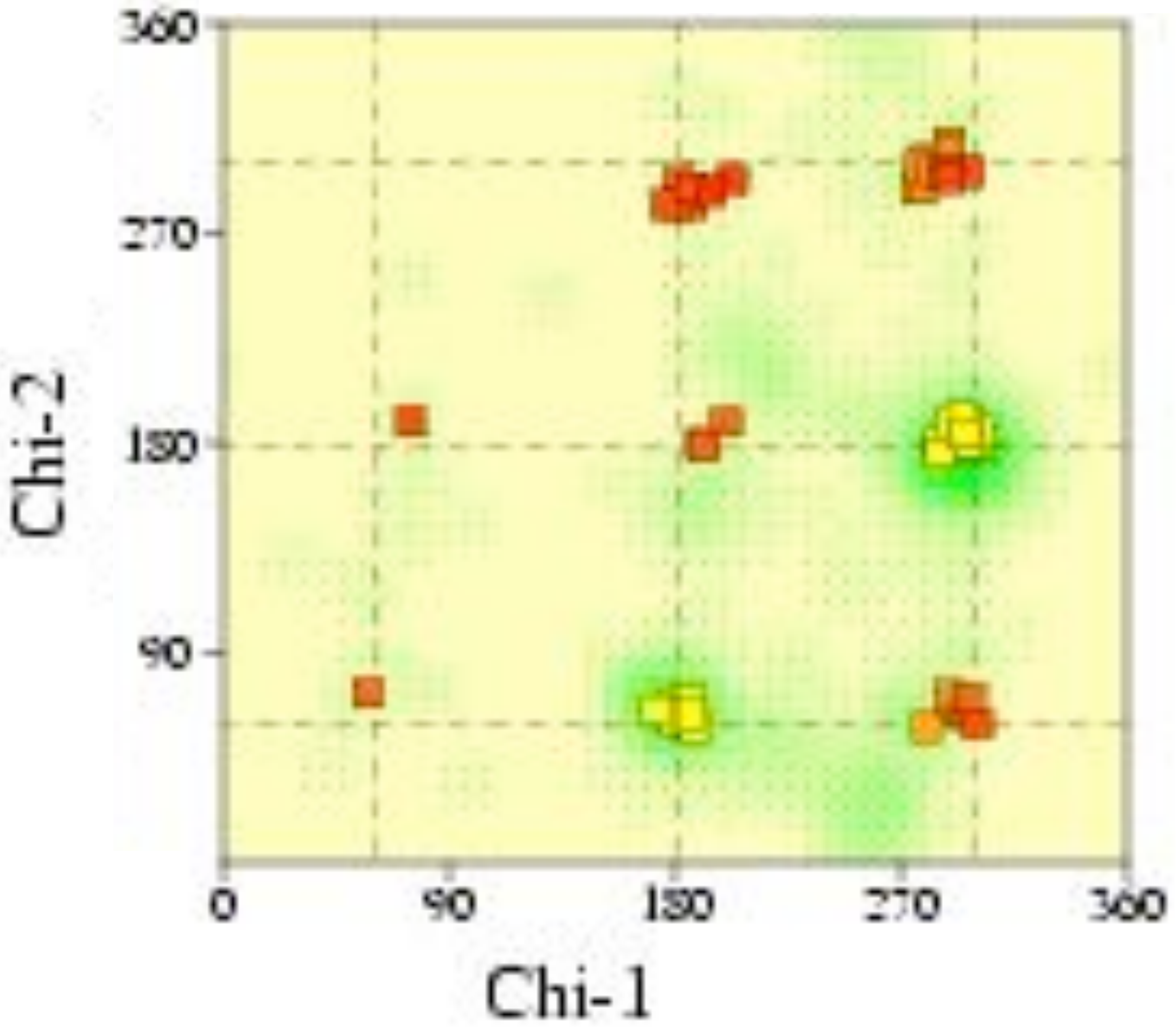
$$\chi^1 = \phi$$

$$U(\phi) = \frac{k_\phi}{2}(1 + cos3\phi)$$



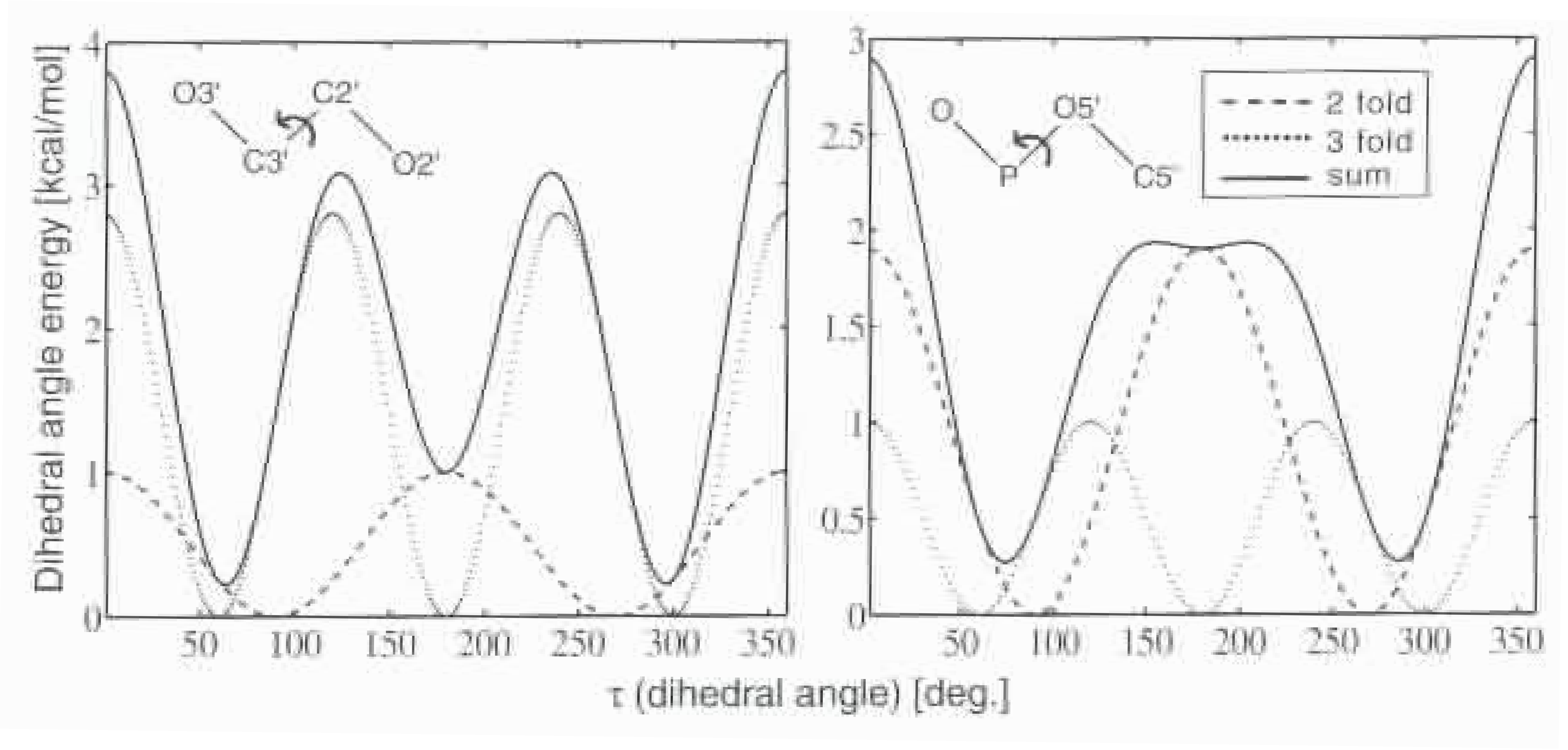


leucine (Leu, L)



Procheck: http://www.ebi.ac.uk/pdbsum/

$$U_{torsion}(\phi) = \sum_n \frac{k_{\phi,n}}{2}[1 + cos(n\phi - \delta)]$$

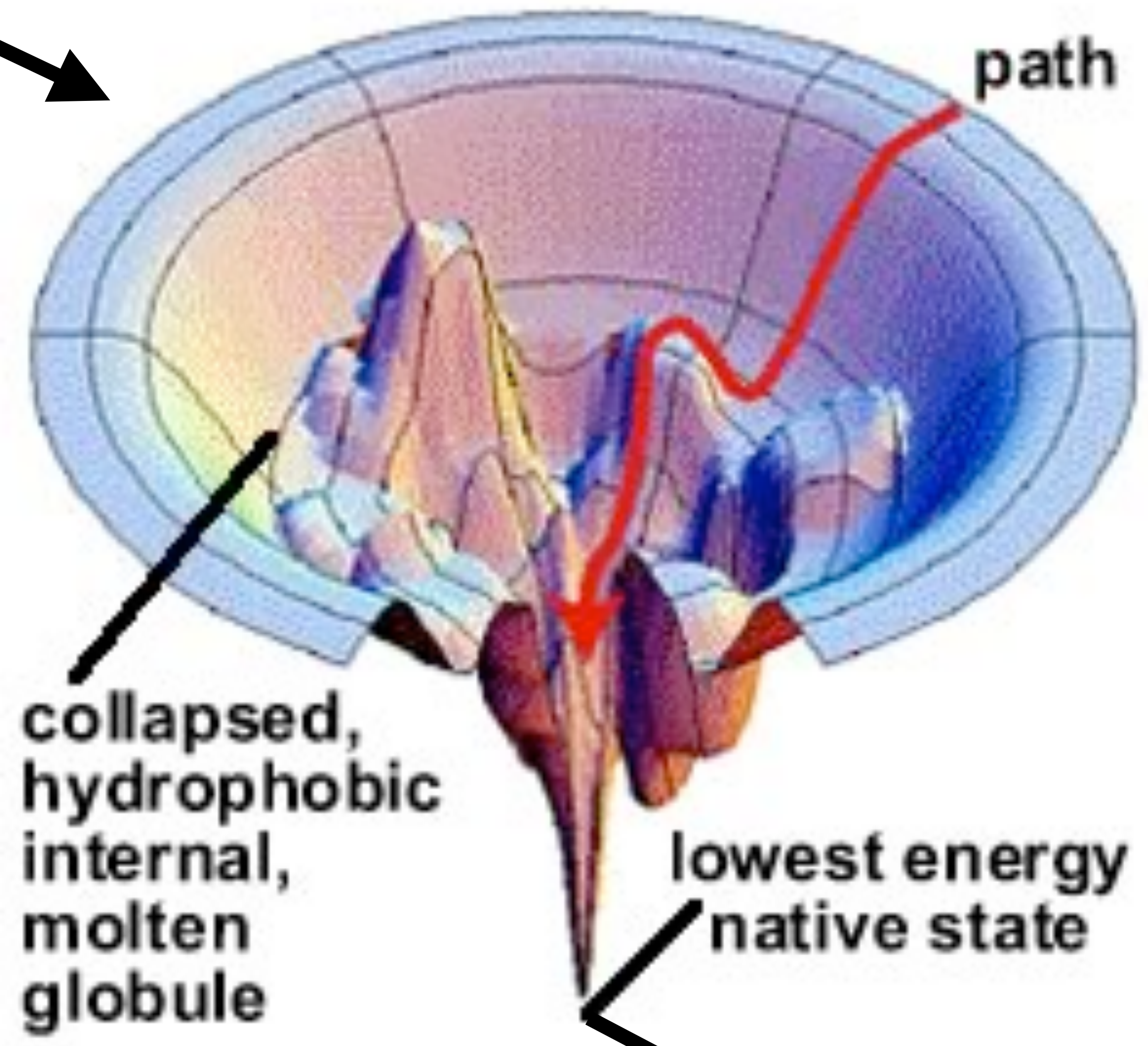


Fast energy funnel from sequence to structure

MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
TPDDHLVARWCATPVAEVAEKSARF
PPRPEELLLPREETLPEMYSLSFTA
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS

primary sequence

Anfinsen dogma



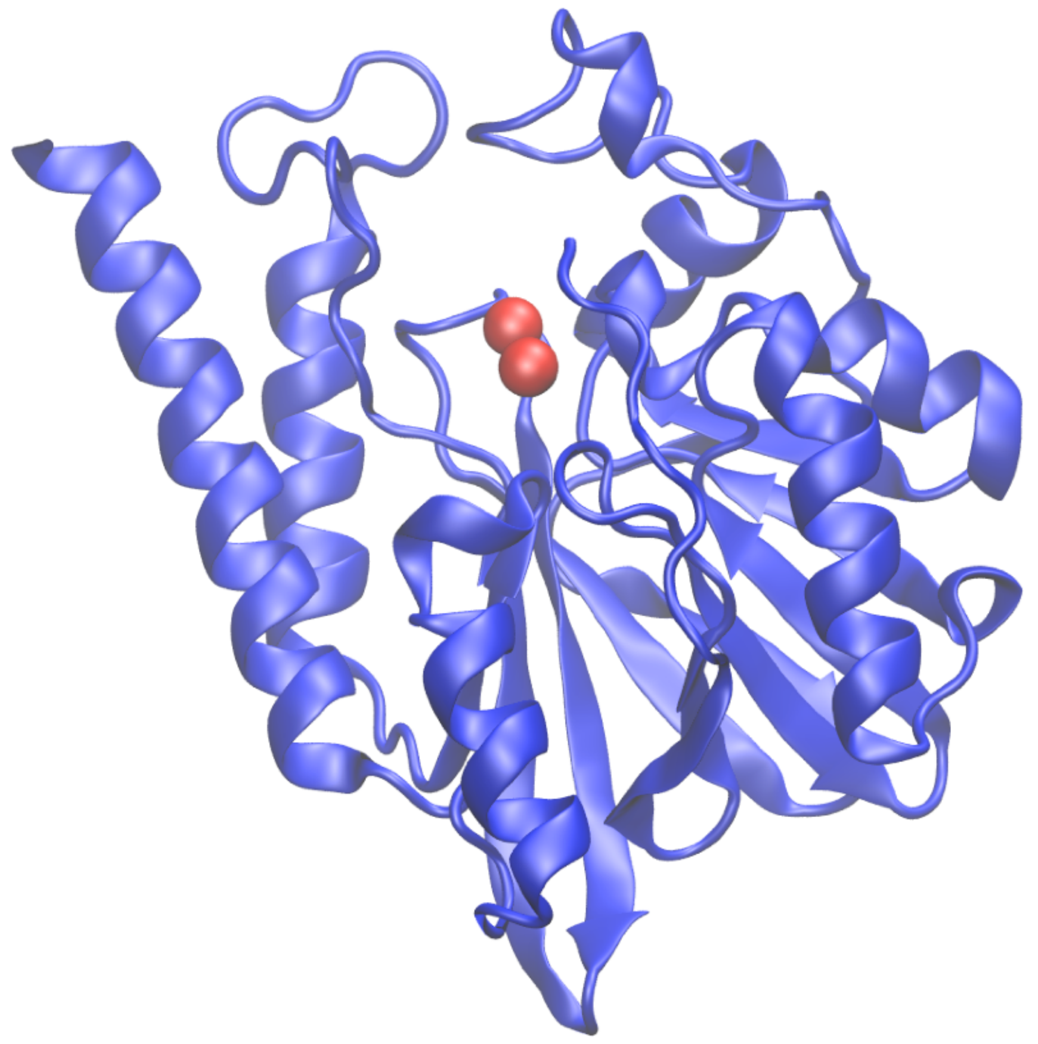
Dill@UCSF

folding pathway

Levinthal paradox ('69)

100 residue-long peptide

$3^{198} \sim 10^{94}$ torsional degrees of freedom



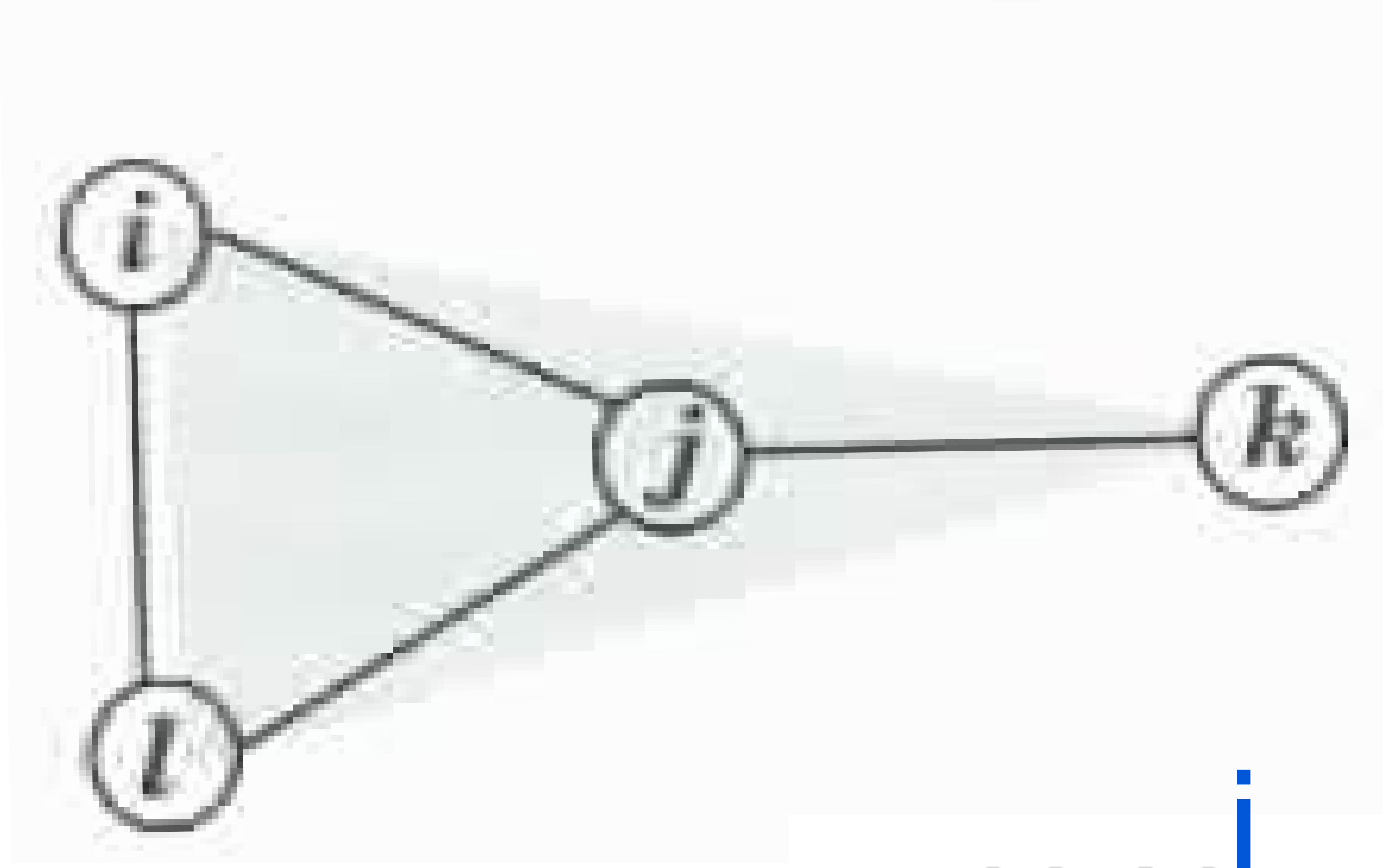
folded native structure

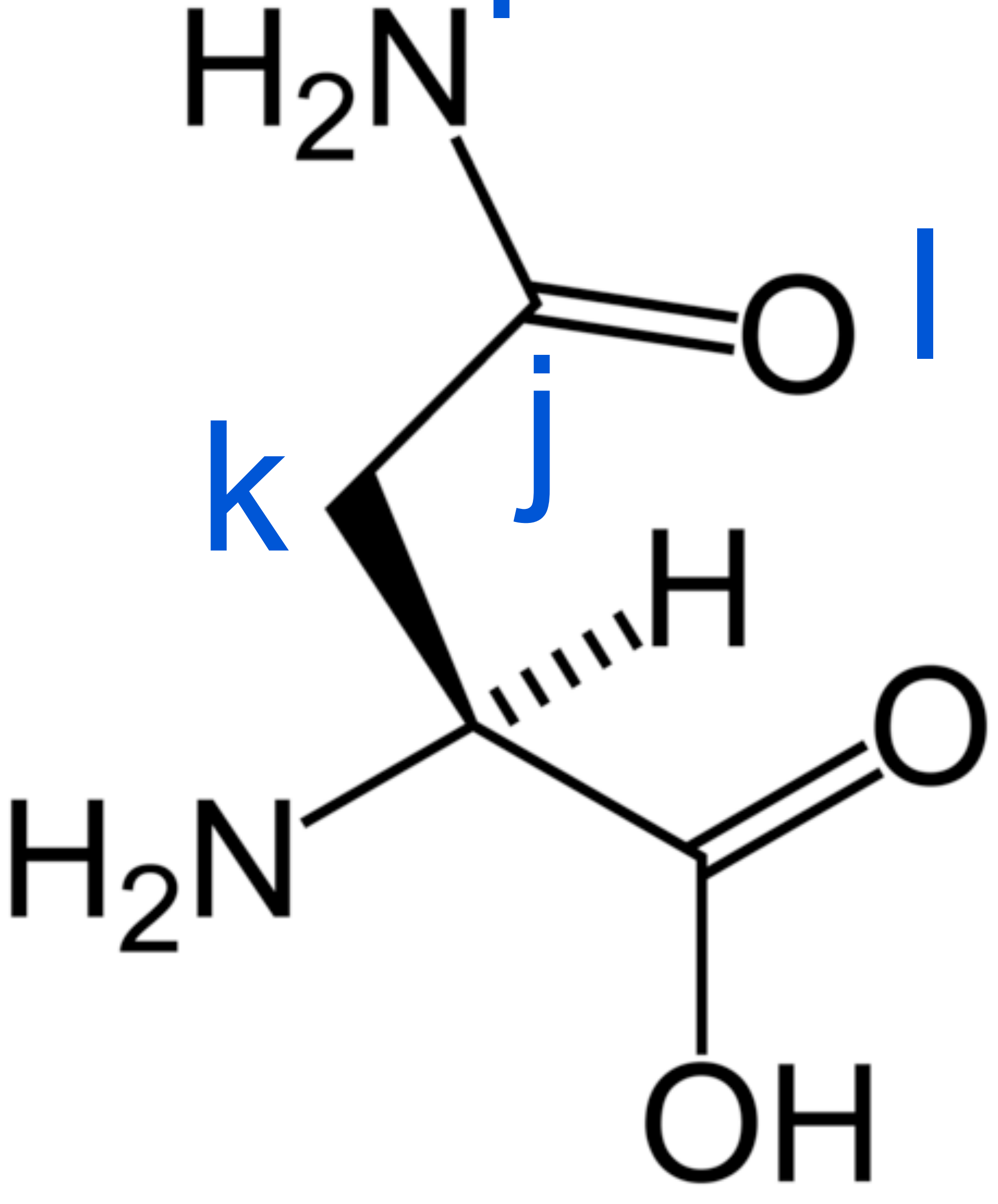
Improper torsions

- used to enforce **planarity** and **chirality** of specific groups
- χ is the improper Wilson angle (i,j,k,l) where j is the central atom
- planarity of peptide bond (-N-C$_\alpha$-C-O-), side chains of Asn, Gln, Asp, Glu, Arg.
- as before possible cross dihedral/bond, dihedral/bend terms can be added

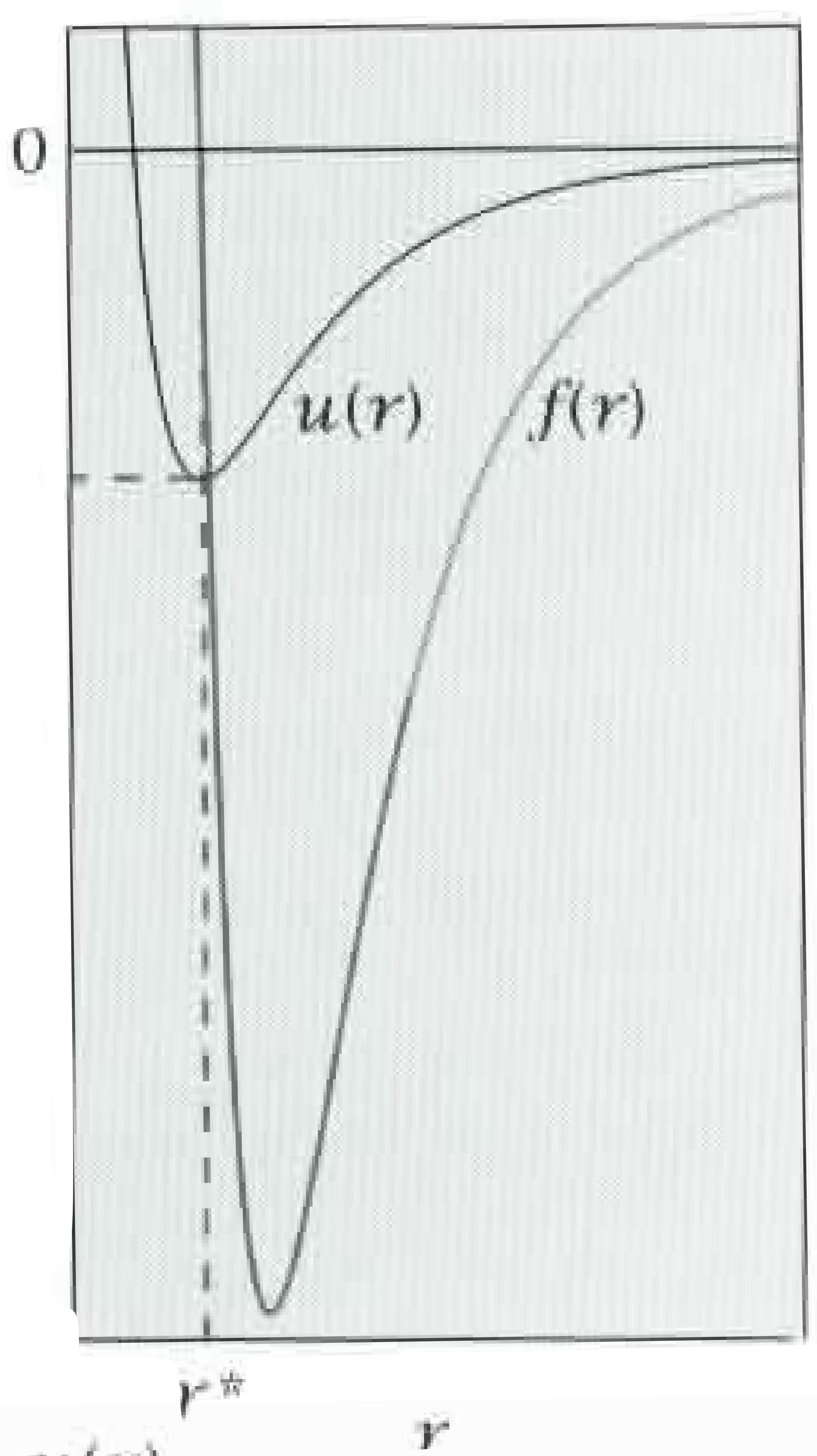
$$U_{improper}(\chi) = \frac{k_\chi}{2}[1 - cos(2\chi)]$$

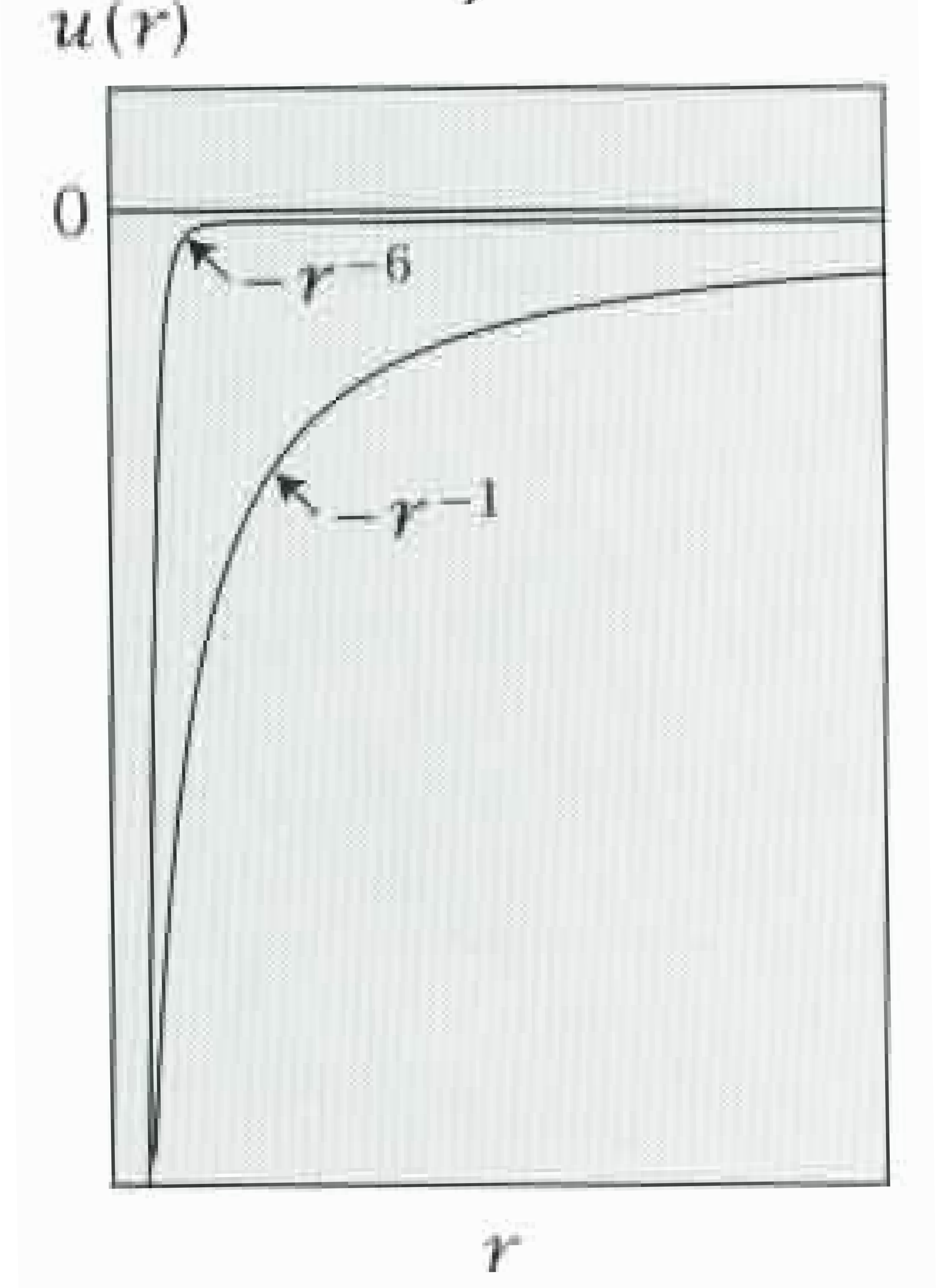
$$U_{improper}(\chi) = \frac{k_\chi}{2}\chi^2$$

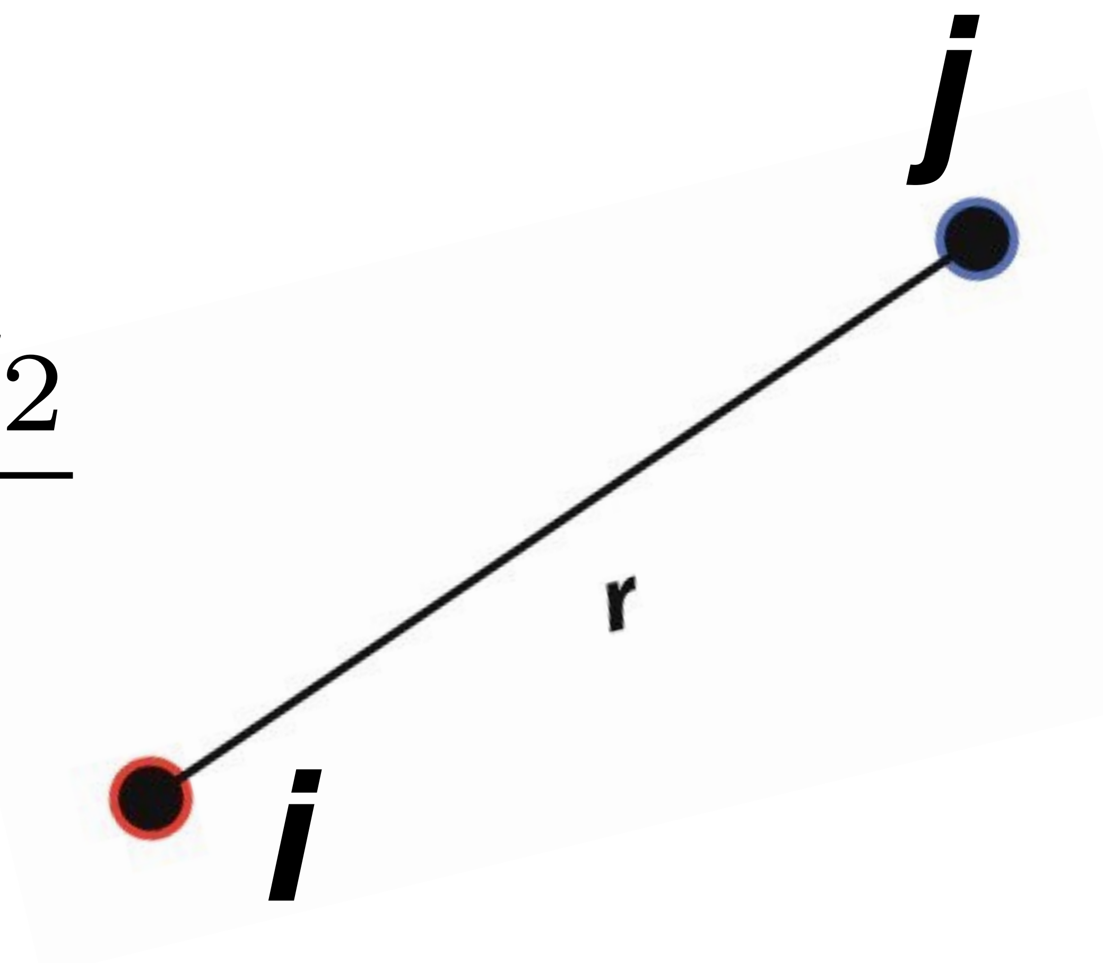




Non-bonded interaction are modeled with pair potential *U(r)*







$$u(r) \propto \frac{1}{r^p}$$ ***e.g.*** $$U(r) = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r}$$

$$f(r) = -\frac{du(r)}{dr}$$

non-bonded interactions usually modeled as a power law

$p \leqslant 3$ **long-ranged:** $u(r) \propto \pm\frac{1}{r}$

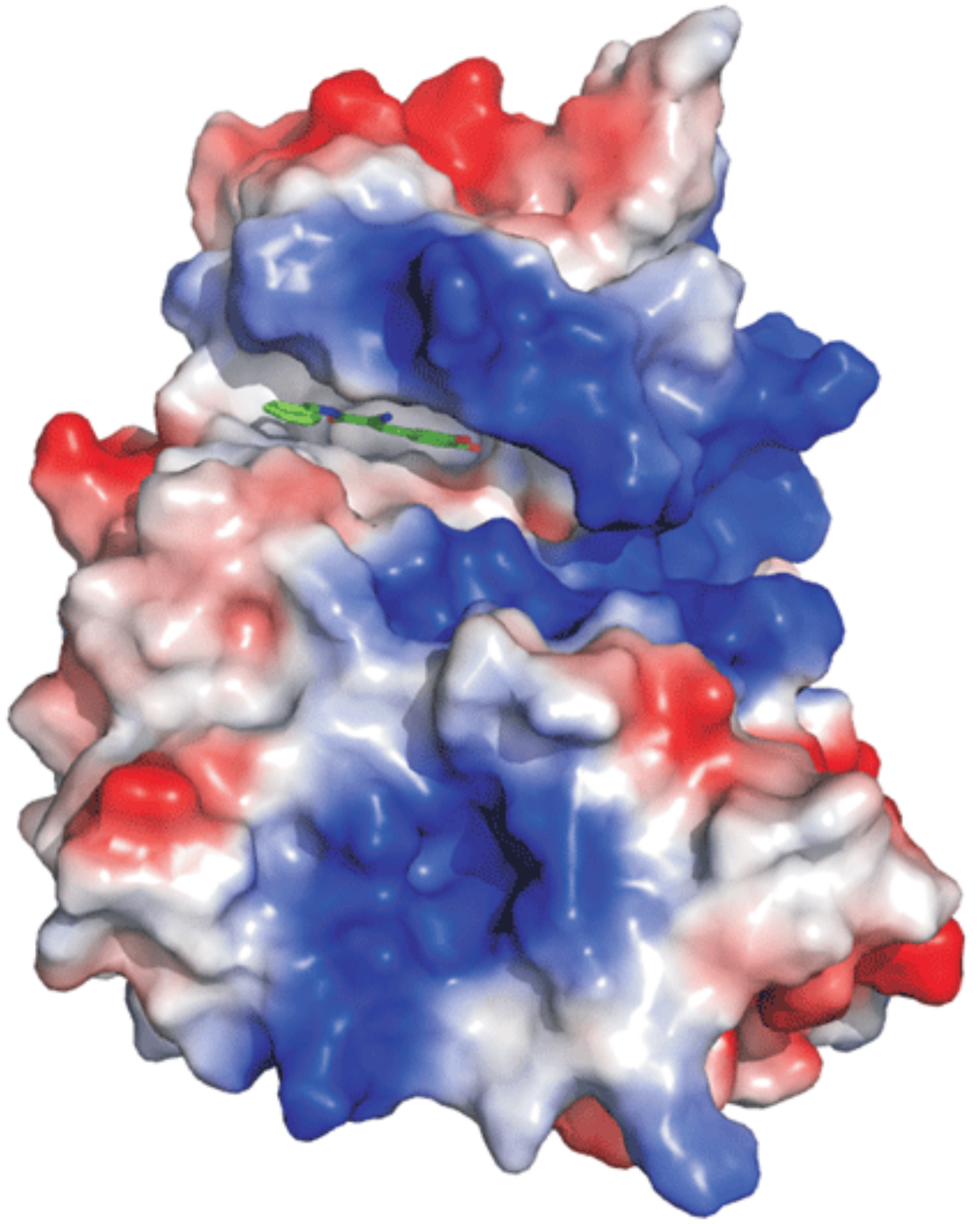
$p > 3$ **short-ranged:** $u(r) \propto \pm\frac{1}{r^6}$

- weaker, but still can be described as electrostatic interactions
- very short-range *p=9,12,14* (Pauli principle)

Electrostatics of proteins

- acidic and basic amino-acids mainly define the global electrostatic signature of each protein

- hydrophobic amino-acids usually have a neutral distribution of charge



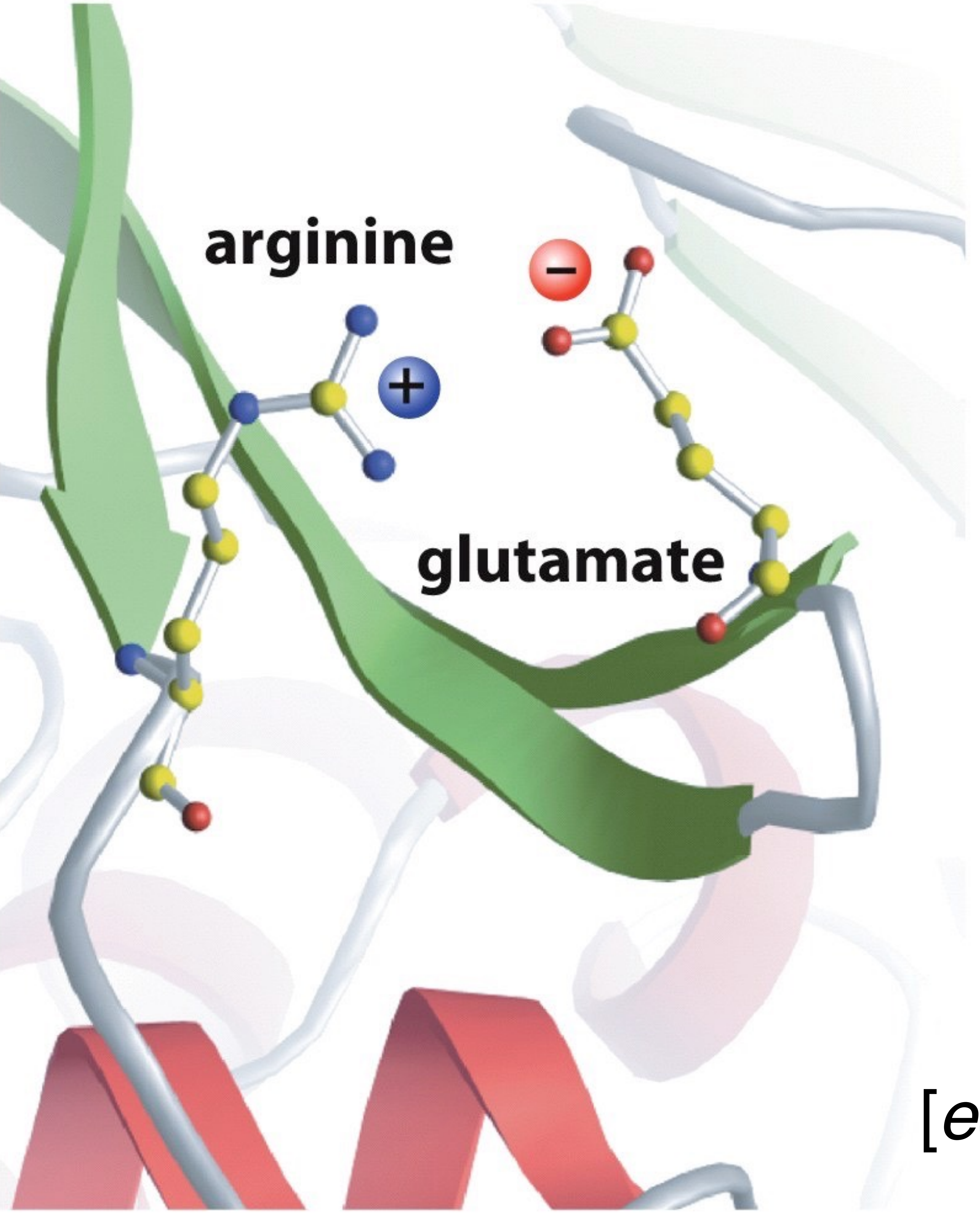
Protein kinase CK2 in complex with the inhibitor tyrphostin

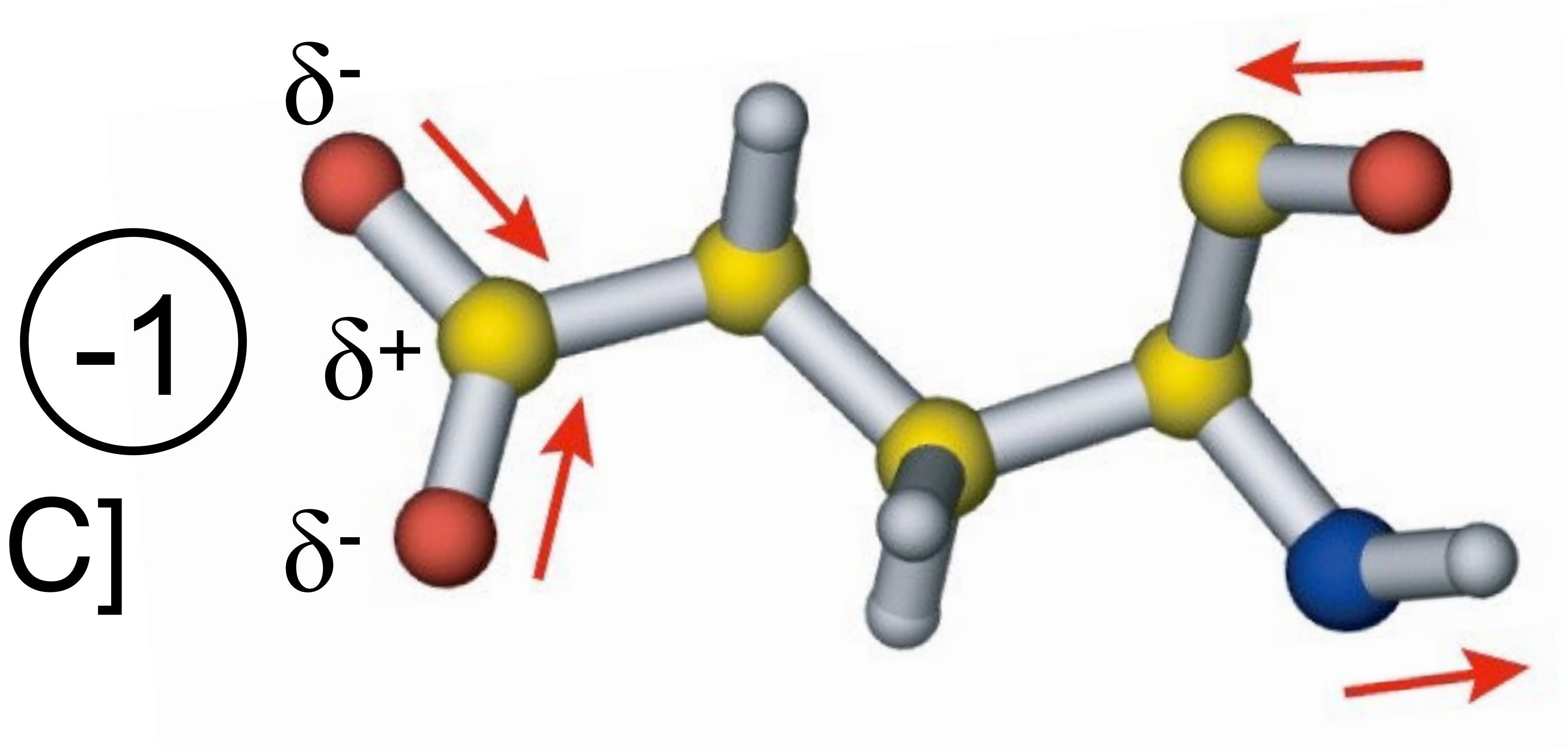
Electrostatics of proteins

$$U_{electrostatic}(r) = \frac{1}{4\pi\epsilon_0}\frac{q_iq_j}{r_{ij}}$$

$$= 1391\ kJ \cdot mol^{-1}\frac{q_iq_j}{r_{ij}[\mathring{A}]}$$

- Coulomb's law determines the interaction between charged molecules
- in proteins formation of **salt-bridges** with estimated energy of ~100 kcal/mol
- effect of water solvation has a large screening effect (dielectric constant of water is ~80)



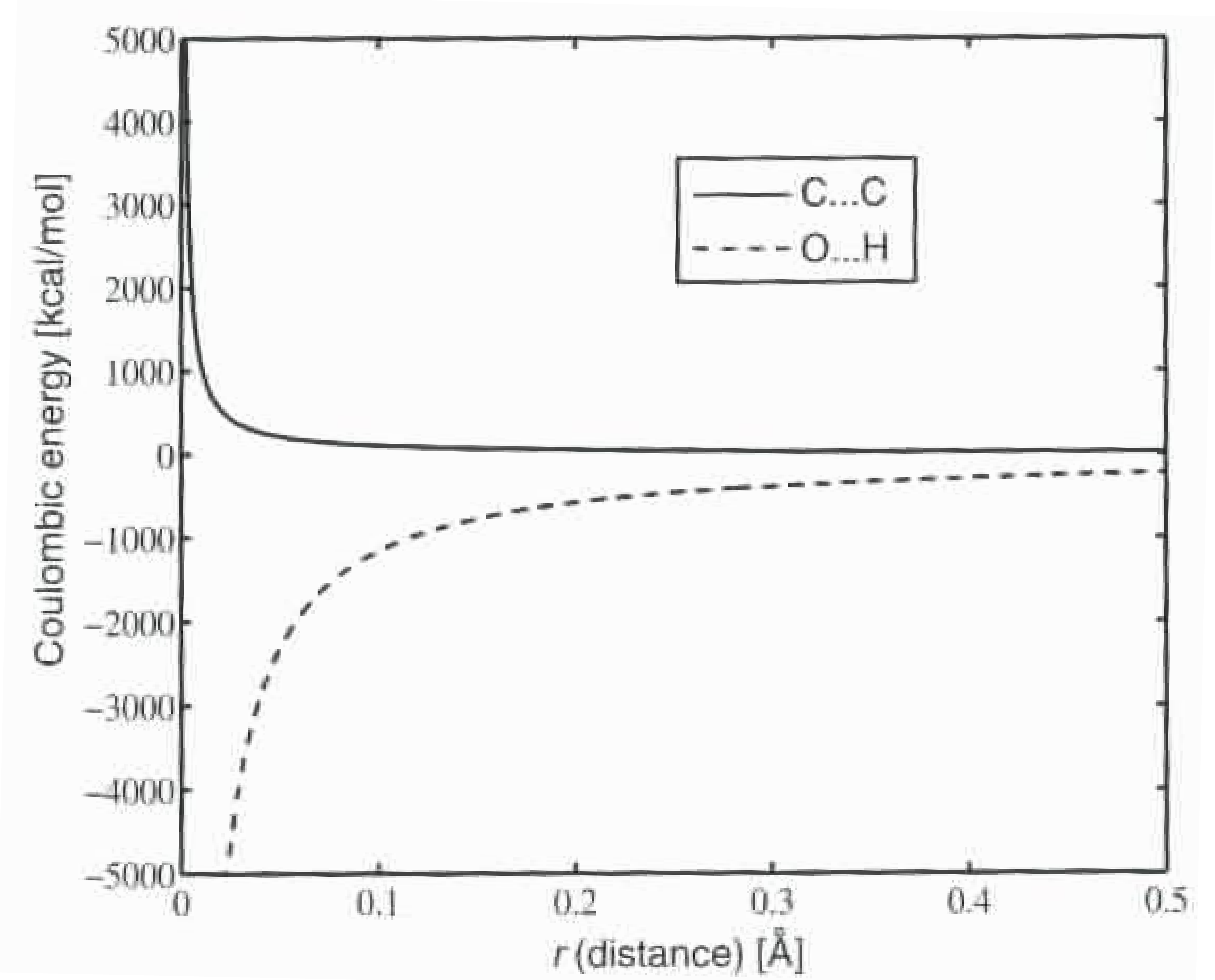


[e= 1.602*10^{-19} C]

Electrostatic potential

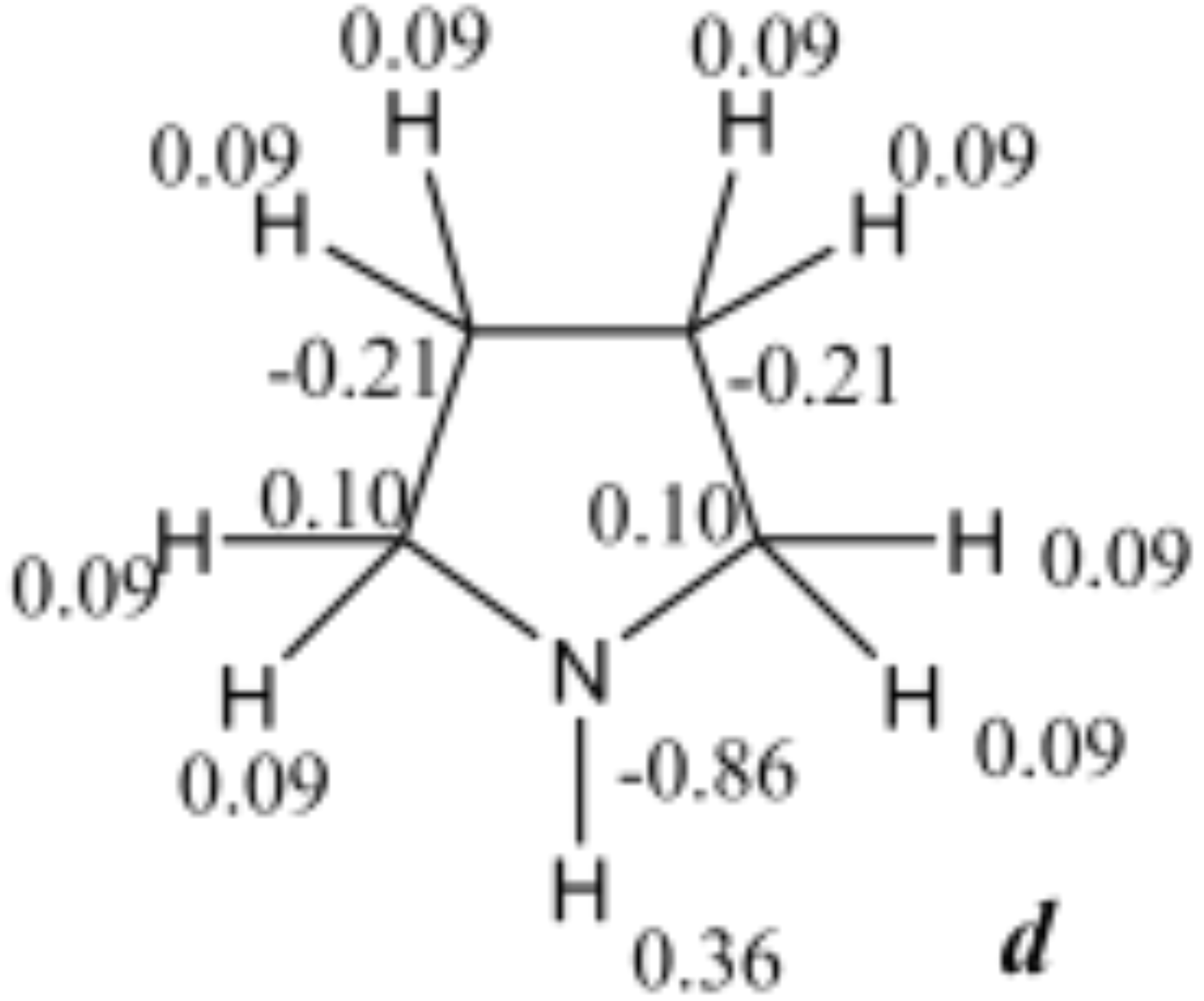
- long-range interactions, slow decay with distance
- very important for protein stabilization and folding
- N*(N-1)/2 interactions, scale as *O(N²)*, bottleneck of the calculation
- requires use of cutoff methods or efficient algorithms for electrostatics(e.g. Ewald summation or Particle Mesh Ewald, PME)

$$U_{electrostatic}(r) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}}$$



Partial atomic charges

- easy for simple molecules (e.g. HF with dipole 1.82 D)
- set to reproduce thermodynamic properties
- derive charges from QM calculations, but there is not a univocal way to do it



$$\phi(\mathbf{r}) = \phi_{nuclei}(\mathbf{r}) + \phi_{electrons}(\mathbf{r}) = \sum_{i=1}^{N} \frac{Z_i}{|\mathbf{r} - \mathbf{R_i}|} - \int \frac{d\mathbf{r}'\,\rho(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|}$$

- least-squares fitting procedure is used to map of the QM electrostatic potential on atomic positions (e.g. RESP in AMBER)

$$R = \sum_{i=1}^{N_{points}} w_i(\phi_i^0 - \phi_i^{calc})^2$$

Polarizability of molecules

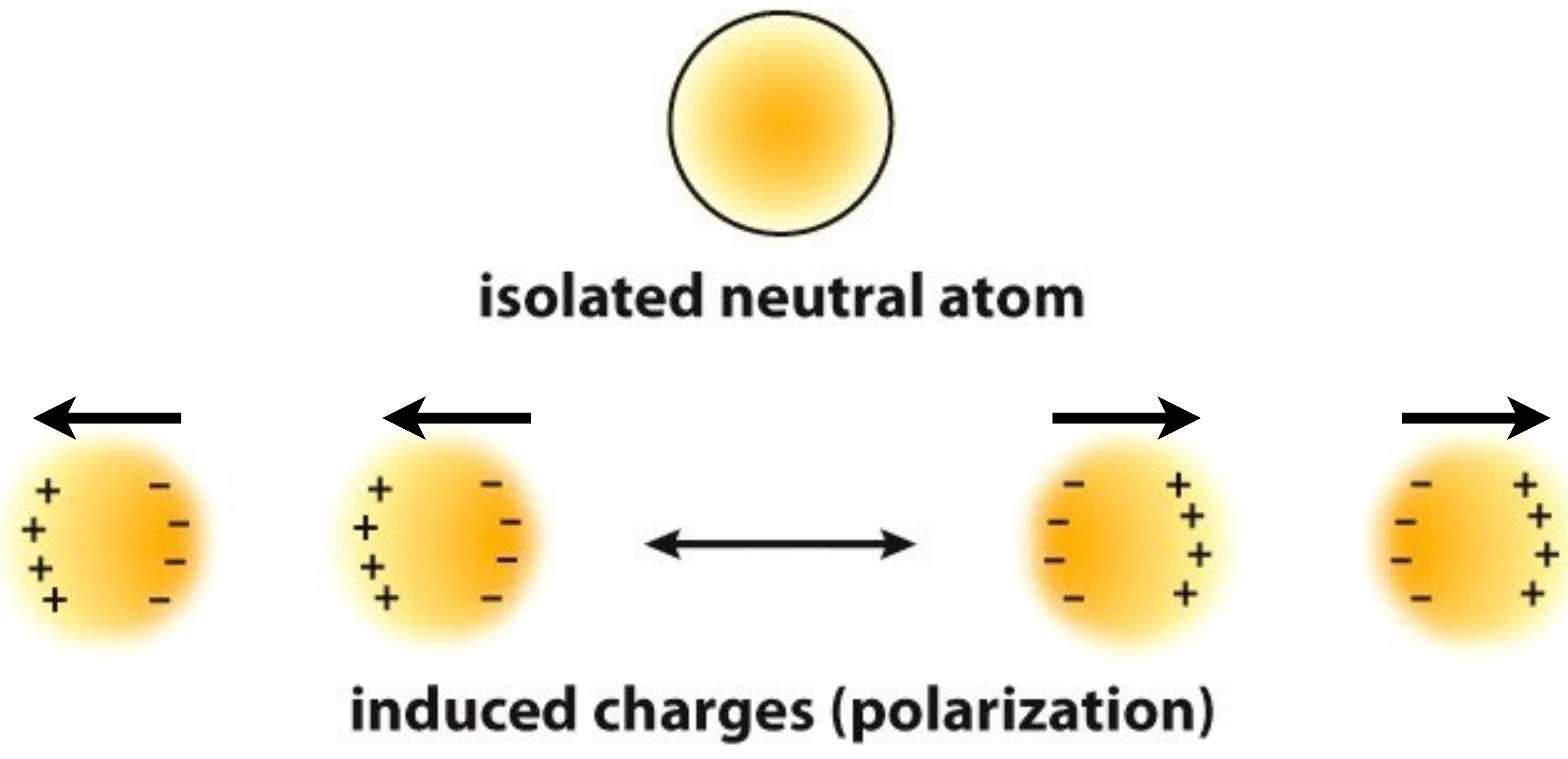
Attractive interactions are universal:
(London or dispersion forces, 1937, inert gases)

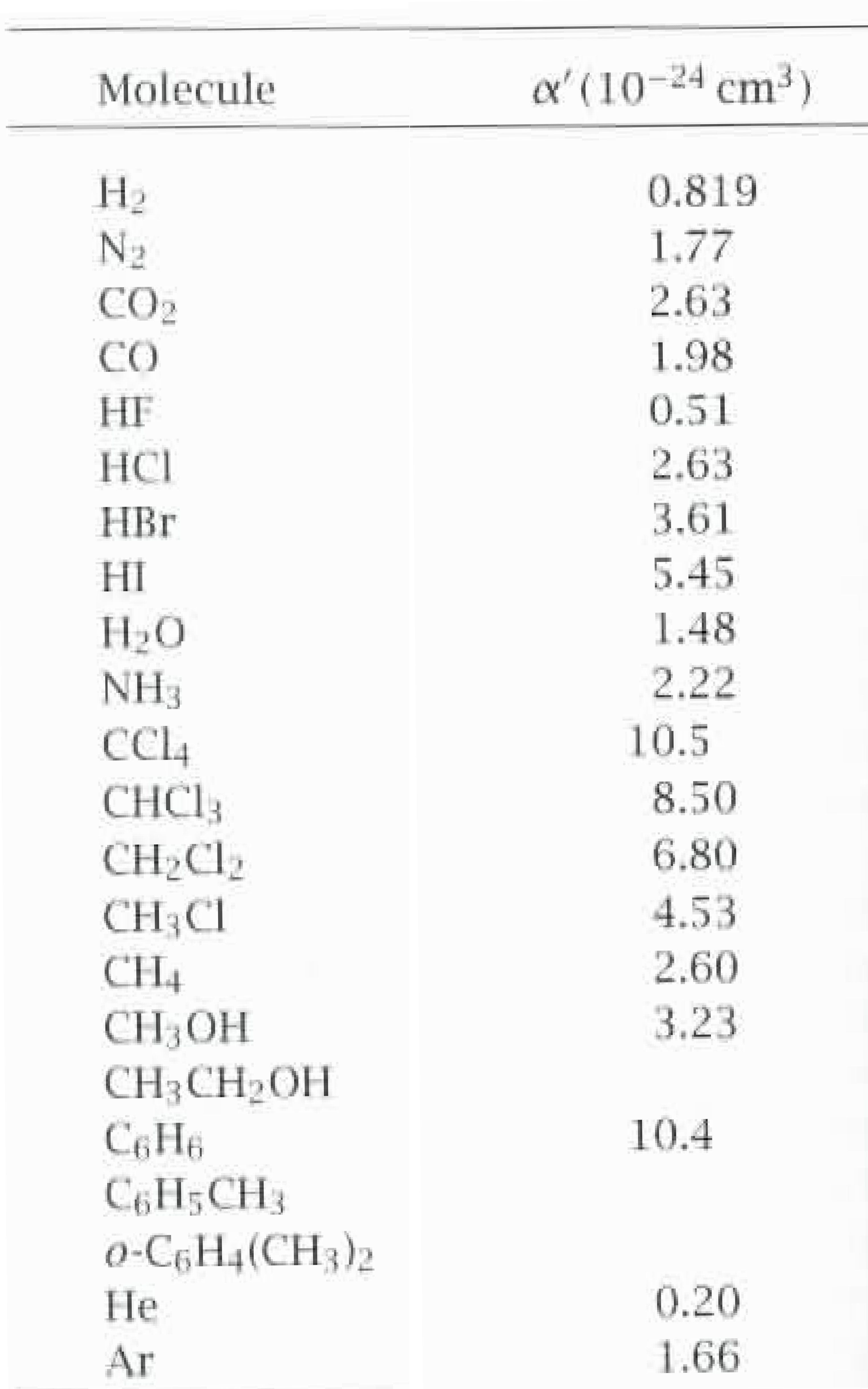
Polarizable atoms/molecules respond to external field, if **E** is small the induced dipole is proportional:

$$\mu_{ind} = \alpha \mathbf{E}$$

$$\alpha : polarizability$$

dipole-induced dipole: $u(r) \ \propto \dfrac{1}{r^6}$
always attractive

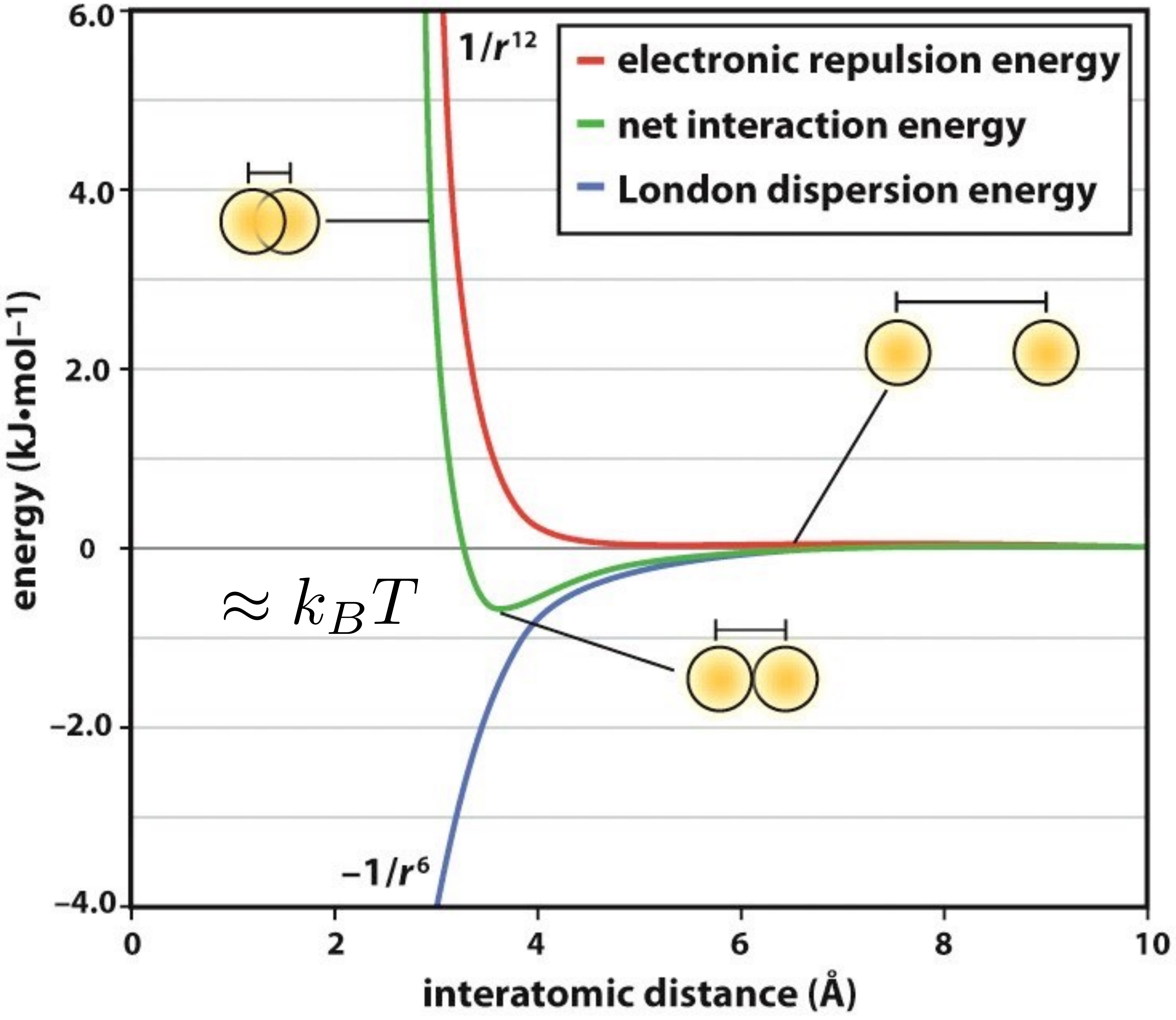




Molecule	$\alpha'(10^{-24}\ \mathrm{cm}^3)$
H_2	0.819
N_2	1.77
CO_2	2.63
CO	1.98
HF	0.51
HCl	2.63
HBr	3.61
HI	5.45
H_2O	1.48
NH_3	2.22
CCl_4	10.5
$CHCl_3$	8.50
CH_2Cl_2	6.80
CH_3Cl	4.53
CH_4	2.60
CH_3OH	3.23
CH_3CH_2OH	
C_6H_6	10.4
$C_6H_5CH_3$	
$o\text{-}C_6H_4(CH_3)_2$	
He	0.20
Ar	1.66

van der Waals interactions

London forces are at the basis of the attractive interactions and are commonly modeled by **Lennard-Jones** (12-6) energy potentials:



$$U_{VdW} = \sum_{i>j}^{N} \epsilon \left[\left(\frac{r_0}{r_{ij}}\right)^{12} - 2\left(\frac{r_0}{r_{ij}}\right)^{6} \right]$$

$$U_{VdW} = \sum_{i>j}^{N} \frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{6}}$$

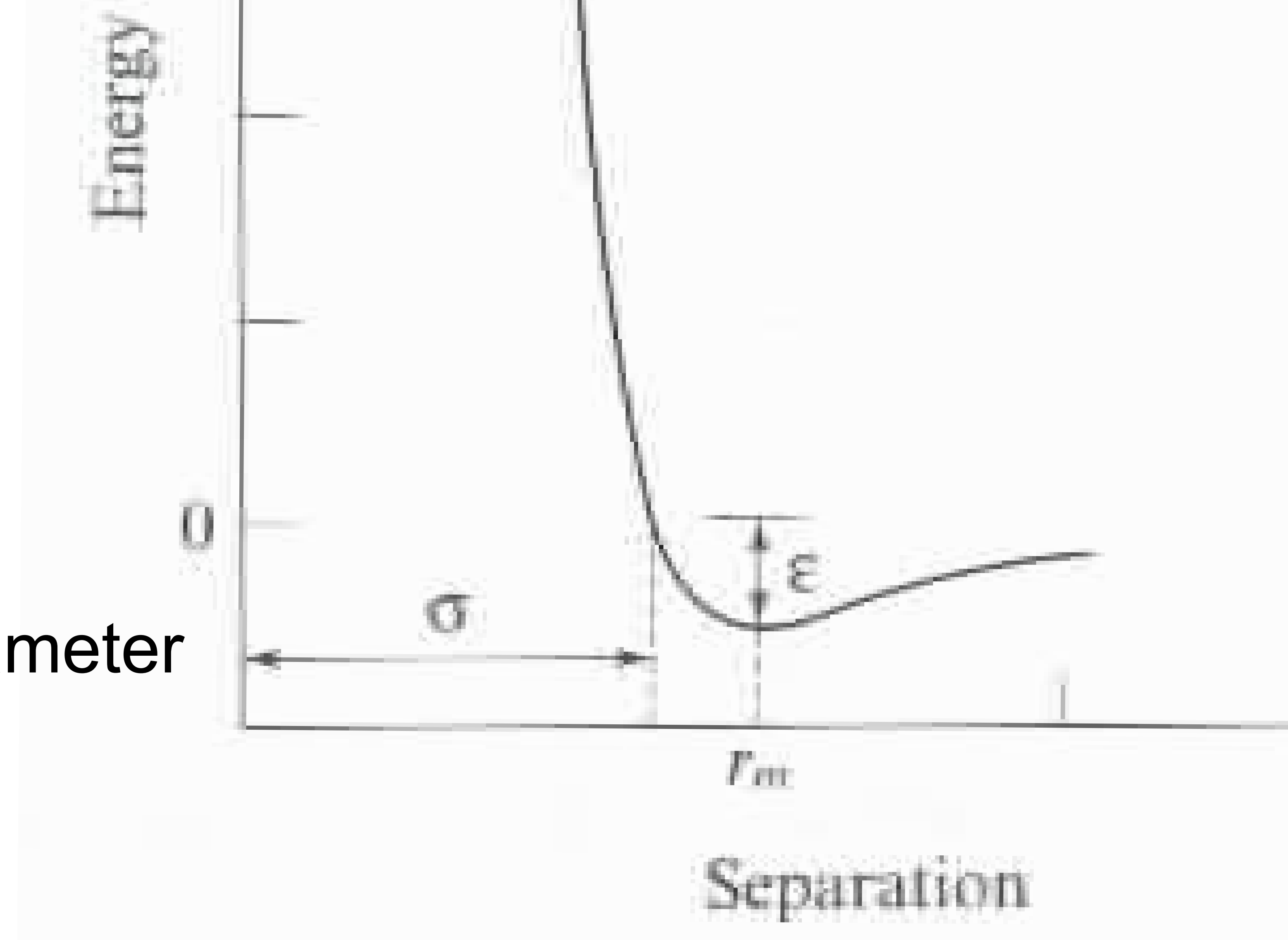
r_0 = eq. interatomic separation,
r_0/2 = vdW radius if *i* and *j* are the same atom type

Van der Waals potentials

- non-bonded **short-range** Lennard-Jones potential (allows for a **cutoff** radius)
- 6/12 form adopted for computational convenience
- Lorentz-Berthelot mixing rules to define parameters for polyatomic systems
- general form:

$$U_{VdW} = \sum_{i>j}^{N} k\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{n} - \left(\frac{\sigma}{r_{ij}} \right)^{m} \right]$$

$$k = \frac{n}{n-m} \left(\frac{n}{m} \right)^{m/(n-m)}$$



collision diameter

$$U_{VdW} = \sum_{i>j}^{N} \epsilon \left[\left(\frac{r_0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_0}{r_{ij}} \right)^{6} \right]$$

$$U_{VdW} = \sum_{i>j}^{N} \frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{6}}$$

for each atom type *i with* (ε,r_0)

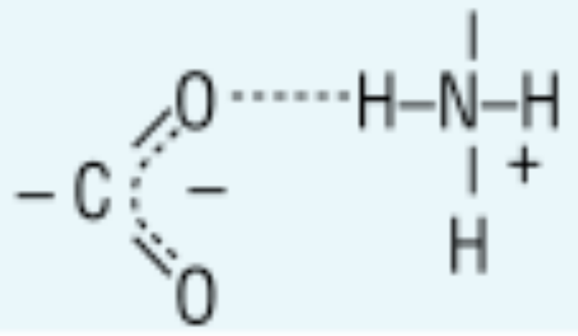
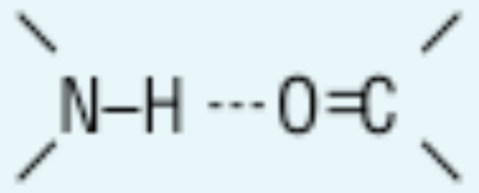
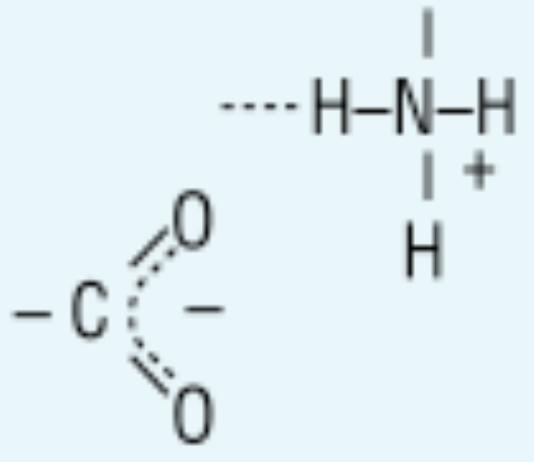
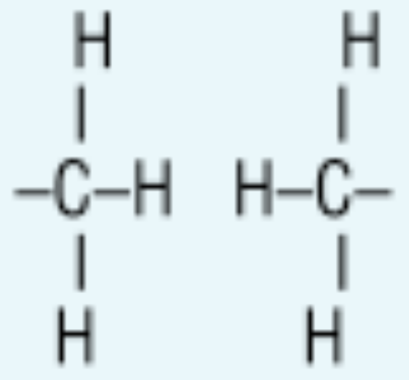
$$A = \epsilon r_0^{12} = 4\epsilon\sigma^{12}$$

$$C = 2\epsilon r_0^{6} = 4\epsilon\sigma^{6} \quad r_0 = 2^{1/6}\sigma$$

for each atom pair {i,j}

$$\begin{cases} r_{0,AB} = \frac{1}{2} r_{0,AA} + \frac{1}{2} r_{0,BB} \\ \sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \\ \epsilon_{AB} = \sqrt{\epsilon_{AA}\epsilon_{BB}} \end{cases}$$

Molecular interactions in biomolecules (bonded and non-bonded)

Interaction	Example	Distance dependence	Typical distance	Free energy (bond dissociation enthalpies for the covalent bonds)
Covalent bond	$-C_\alpha-C-$	-	1.5 Å	356 kJ/mole (610 kJ/mole for a C=C bond)
Disulfide bond	–Cys–S–S–Cys–	-	2.2 Å	167 kJ/mole
Salt bridge		Donor (here N), and acceptor (here O) atoms <3.5 Å	2.8 Å	12.5–17 kJ/mole; may be as high as 30 kJ/mole for fully or partially buried salt bridges (see text), less if the salt bridge is external
Hydrogen bond		Donor (here N), and acceptor (here O) atoms <3.5 Å	3.0 Å	2–6 kJ/mole in water; 12.5–21 kJ/mole if either donor or acceptor is charged
Long-range electrostatic interaction		Depends on dielectric constant of medium. Screened by water. 1/r dependence	Variable	Depends on distance and environment. Can be very strong in nonpolar region but very weak in water
Van der Waals interaction		Short range. Falls off rapidly beyond 4 Å separation. $1/r^6$ dependence	3.5 Å	4 kJ/mole (4–17 in protein interior) depending on the size of the group (for comparison, the average thermal energy of molecules at room temperature is 2.5 kJ/mole)

k_BT(300 K) = 0.6 kcal/mole / 2.5kJ/mole / 4.1 pN nm / 4.1 * 10^{-21} J
(1 kcal/mol = 4.184 kJ/mole)

Empirical potential energy function

$$U_{MM}(r) = \boxed{\sum_{bonds} \frac{k_b}{2}(r-r_0)^2 + \sum_{angles}\frac{k_\theta}{2}(\theta-\theta_0)^2 + \sum_{torsions,n}\frac{k_{\phi,n}}{2}[1+cos(n\phi-\delta)]} +$$

$$+ \boxed{\sum_{i>j}^{N}\left(\frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{6}}\right) + \sum_{i>j}^{N}\frac{1}{4\pi\epsilon_0}\frac{q_iq_j}{r_{ij}}}$$

- large number of **parameters** fitted to represent experimental data or QM calculated quantities (usually structure and thermodynamic of small molecules)

- “trial and error” or least-squares fitting methods to converge to a consistent set of parameters

- coupling/correlation between parameters, thus parameterization of a **force field (FF)** is a global task

- assumption that parameters can be **transferable** to different contexts (specialized *vs.* generalized FF)

2013 NOBEL PRIZE IN CHEMISTRY

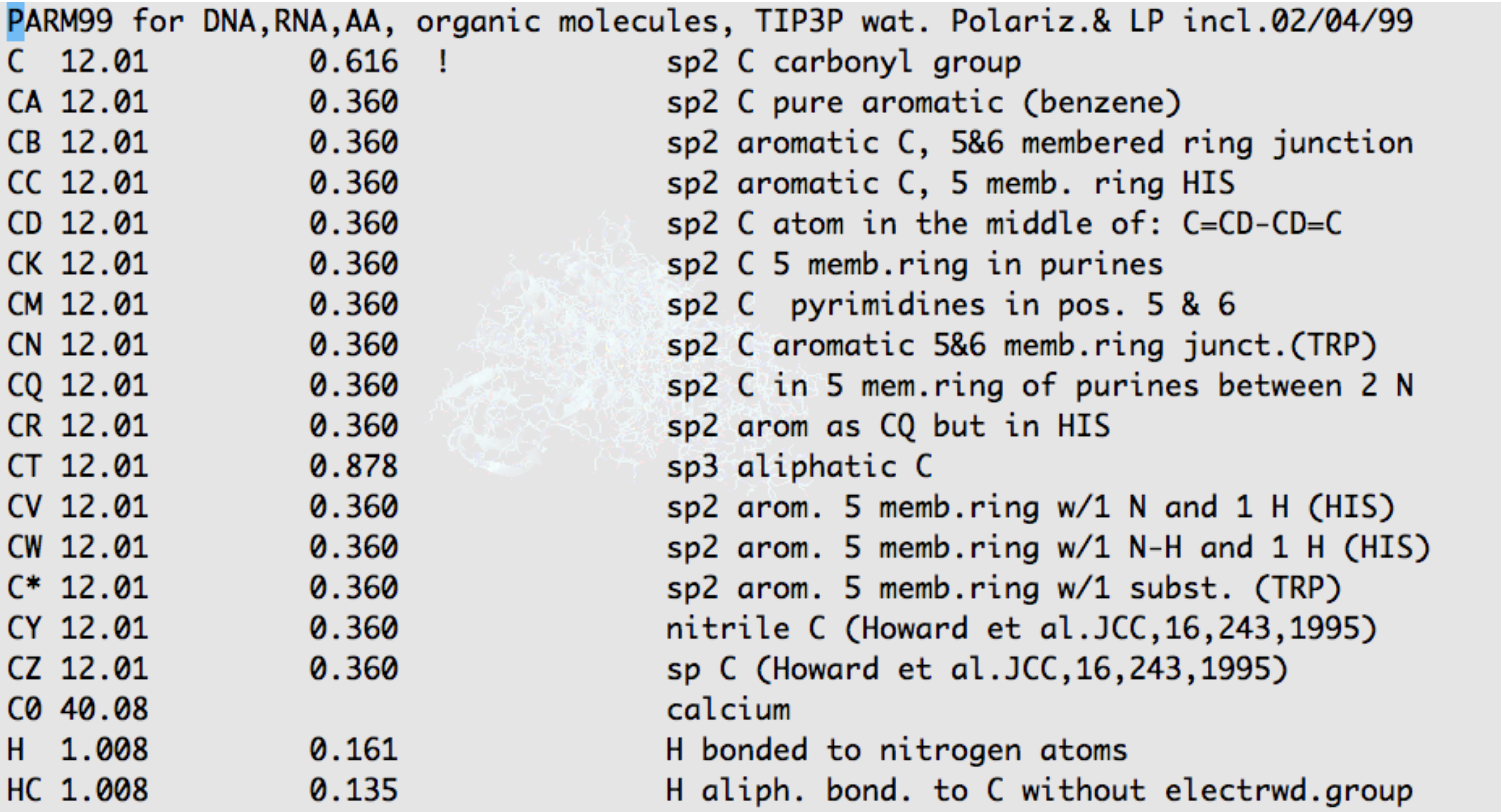
Martin Karplus
Michael Levitt
Arieh Warshel

© ® The Nobel Foundation. Photo: Lovisa Engblom.

"for the development of multiscale models for complex chemical systems"

A parameter file

- atom types

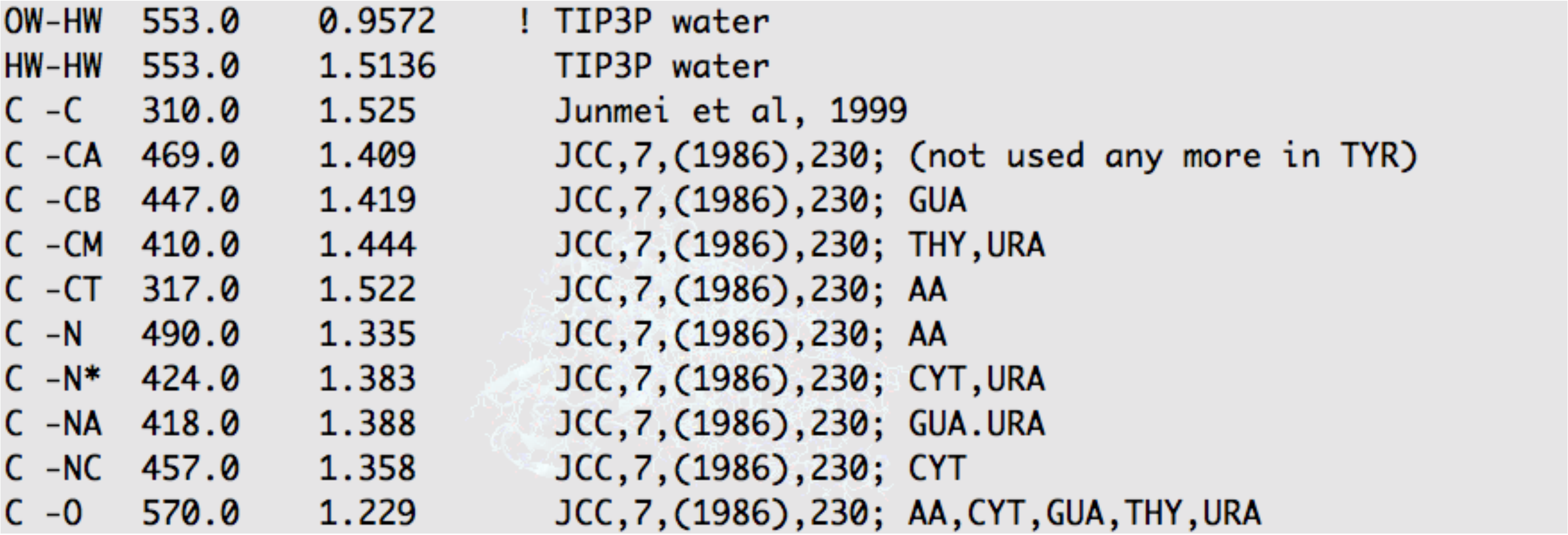


```
PARM99 for DNA,RNA,AA, organic molecules, TIP3P wat. Polariz.& LP incl.02/04/99
C  12.01         0.616  !            sp2 C carbonyl group
CA 12.01         0.360               sp2 C pure aromatic (benzene)
CB 12.01         0.360               sp2 aromatic C, 5&6 membered ring junction
CC 12.01         0.360               sp2 aromatic C, 5 memb. ring HIS
CD 12.01         0.360               sp2 C atom in the middle of: C=CD-CD=C
CK 12.01         0.360               sp2 C 5 memb.ring in purines
CM 12.01         0.360               sp2 C  pyrimidines in pos. 5 & 6
CN 12.01         0.360               sp2 C aromatic 5&6 memb.ring junct.(TRP)
CQ 12.01         0.360               sp2 C in 5 mem.ring of purines between 2 N
CR 12.01         0.360               sp2 arom as CQ but in HIS
CT 12.01         0.878               sp3 aliphatic C
CV 12.01         0.360               sp2 arom. 5 memb.ring w/1 N and 1 H (HIS)
CW 12.01         0.360               sp2 arom. 5 memb.ring w/1 N-H and 1 H (HIS)
C* 12.01         0.360               sp2 arom. 5 memb.ring w/1 subst. (TRP)
CY 12.01         0.360               nitrile C (Howard et al.JCC,16,243,1995)
CZ 12.01         0.360               sp C (Howard et al.JCC,16,243,1995)
C0 40.08                             calcium
H  1.008         0.161               H bonded to nitrogen atoms
HC 1.008         0.135               H aliph. bond. to C without electrwd.group
```

AMBER FF (parm99)

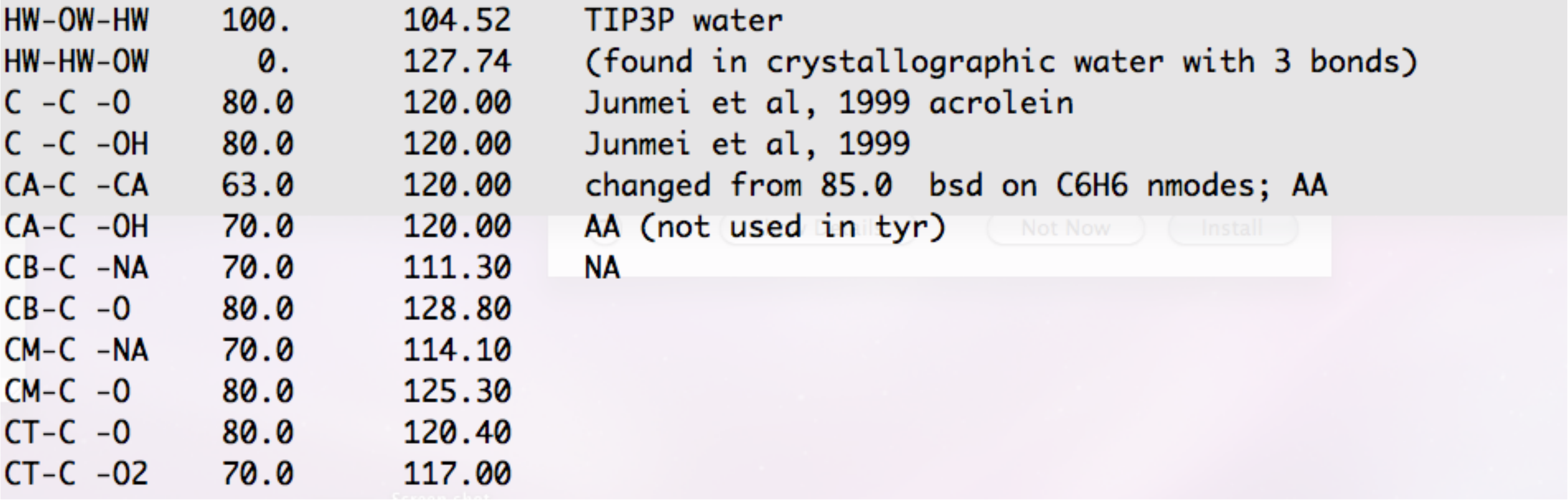
A parameter file

- bond types



```
OW-HW  553.0    0.9572    ! TIP3P water
HW-HW  553.0    1.5136      TIP3P water
C -C   310.0    1.525       Junmei et al, 1999
C -CA  469.0    1.409       JCC,7,(1986),230; (not used any more in TYR)
C -CB  447.0    1.419       JCC,7,(1986),230; GUA
C -CM  410.0    1.444       JCC,7,(1986),230; THY,URA
C -CT  317.0    1.522       JCC,7,(1986),230; AA
C -N   490.0    1.335       JCC,7,(1986),230; AA
C -N*  424.0    1.383       JCC,7,(1986),230; CYT,URA
C -NA  418.0    1.388       JCC,7,(1986),230; GUA.URA
C -NC  457.0    1.358       JCC,7,(1986),230; CYT
C -O   570.0    1.229       JCC,7,(1986),230; AA,CYT,GUA,THY,URA
```

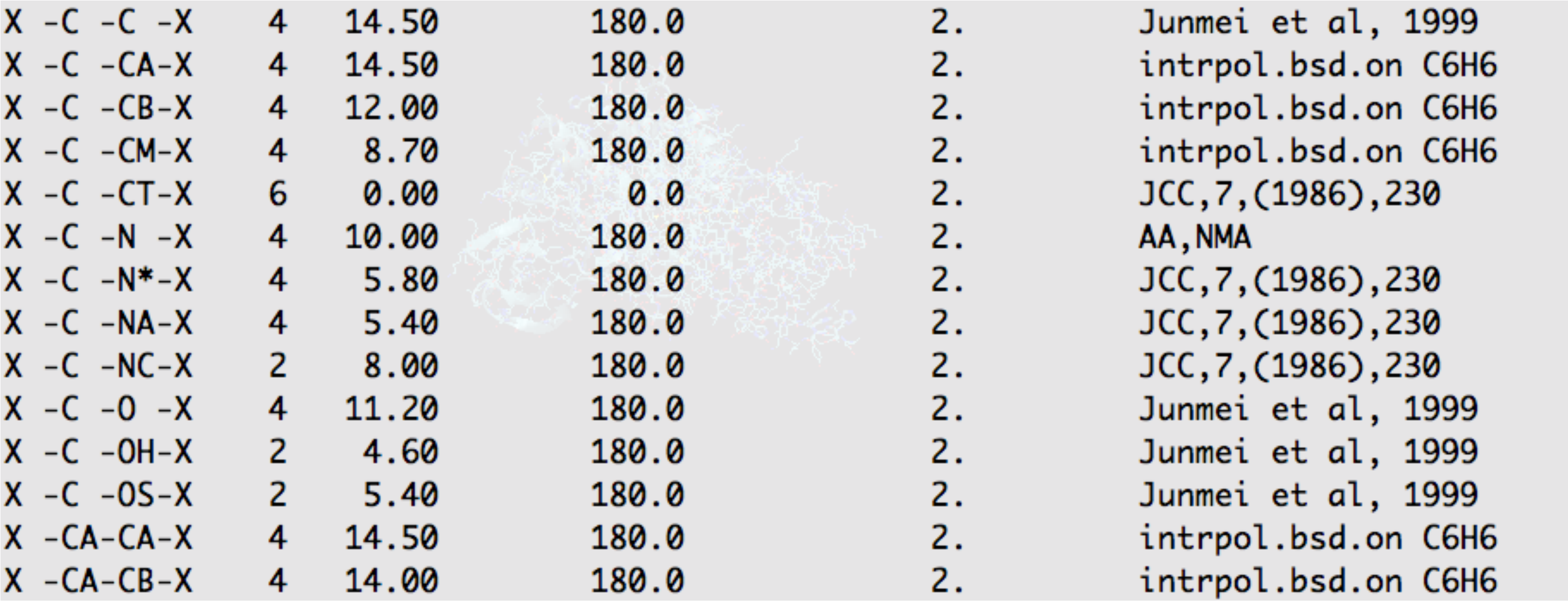
- angle types



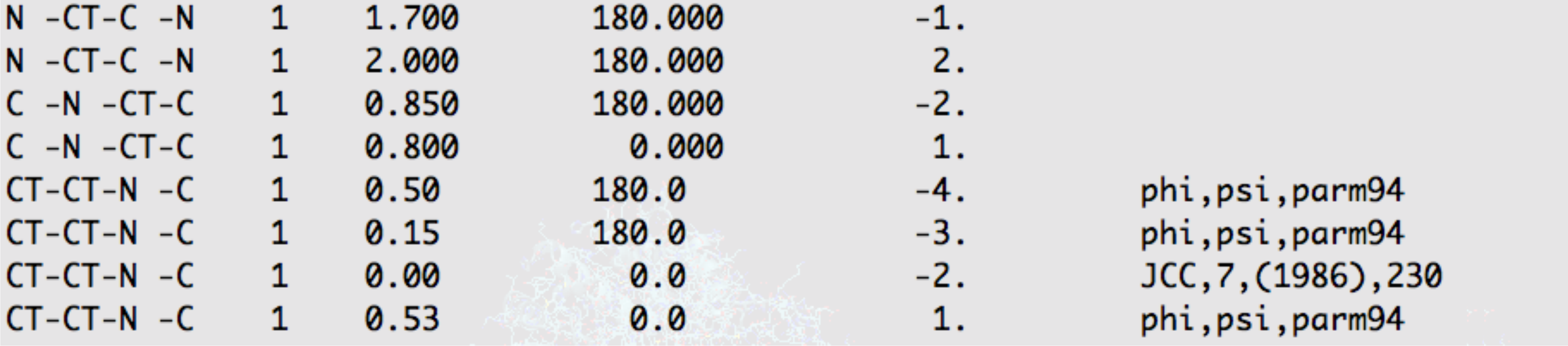
```
HW-OW-HW    100.       104.52    TIP3P water
HW-HW-OW      0.       127.74    (found in crystallographic water with 3 bonds)
C -C -O     80.0       120.00    Junmei et al, 1999 acrolein
C -C -OH    80.0       120.00    Junmei et al, 1999
CA-C -CA    63.0       120.00    changed from 85.0  bsd on C6H6 nmodes; AA
CA-C -OH    70.0       120.00    AA (not used in tyr)
CB-C -NA    70.0       111.30    NA
CB-C -O     80.0       128.80
CM-C -NA    70.0       114.10
CM-C -O     80.0       125.30
CT-C -O     80.0       120.40
CT-C -O2    70.0       117.00
```

A parameter file

- torsion types



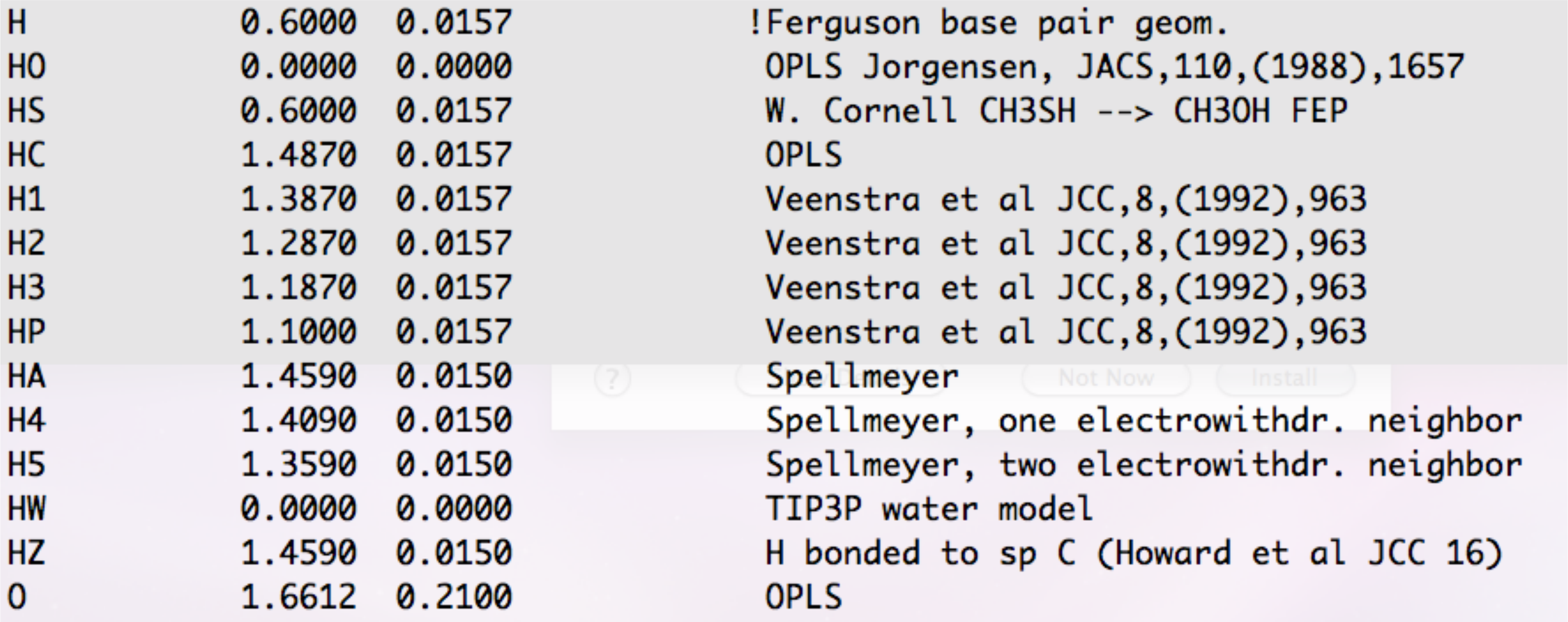
```
X -C -C -X    4   14.50          180.0             2.          Junmei et al, 1999
X -C -CA-X    4   14.50          180.0             2.          intrpol.bsd.on C6H6
X -C -CB-X    4   12.00          180.0             2.          intrpol.bsd.on C6H6
X -C -CM-X    4    8.70          180.0             2.          intrpol.bsd.on C6H6
X -C -CT-X    6    0.00            0.0             2.          JCC,7,(1986),230
X -C -N -X    4   10.00          180.0             2.          AA,NMA
X -C -N*-X    4    5.80          180.0             2.          JCC,7,(1986),230
X -C -NA-X    4    5.40          180.0             2.          JCC,7,(1986),230
X -C -NC-X    2    8.00          180.0             2.          JCC,7,(1986),230
X -C -O -X    4   11.20          180.0             2.          Junmei et al, 1999
X -C -OH-X    2    4.60          180.0             2.          Junmei et al, 1999
X -C -OS-X    2    5.40          180.0             2.          Junmei et al, 1999
X -CA-CA-X    4   14.50          180.0             2.          intrpol.bsd.on C6H6
X -CA-CB-X    4   14.00          180.0             2.          intrpol.bsd.on C6H6
```



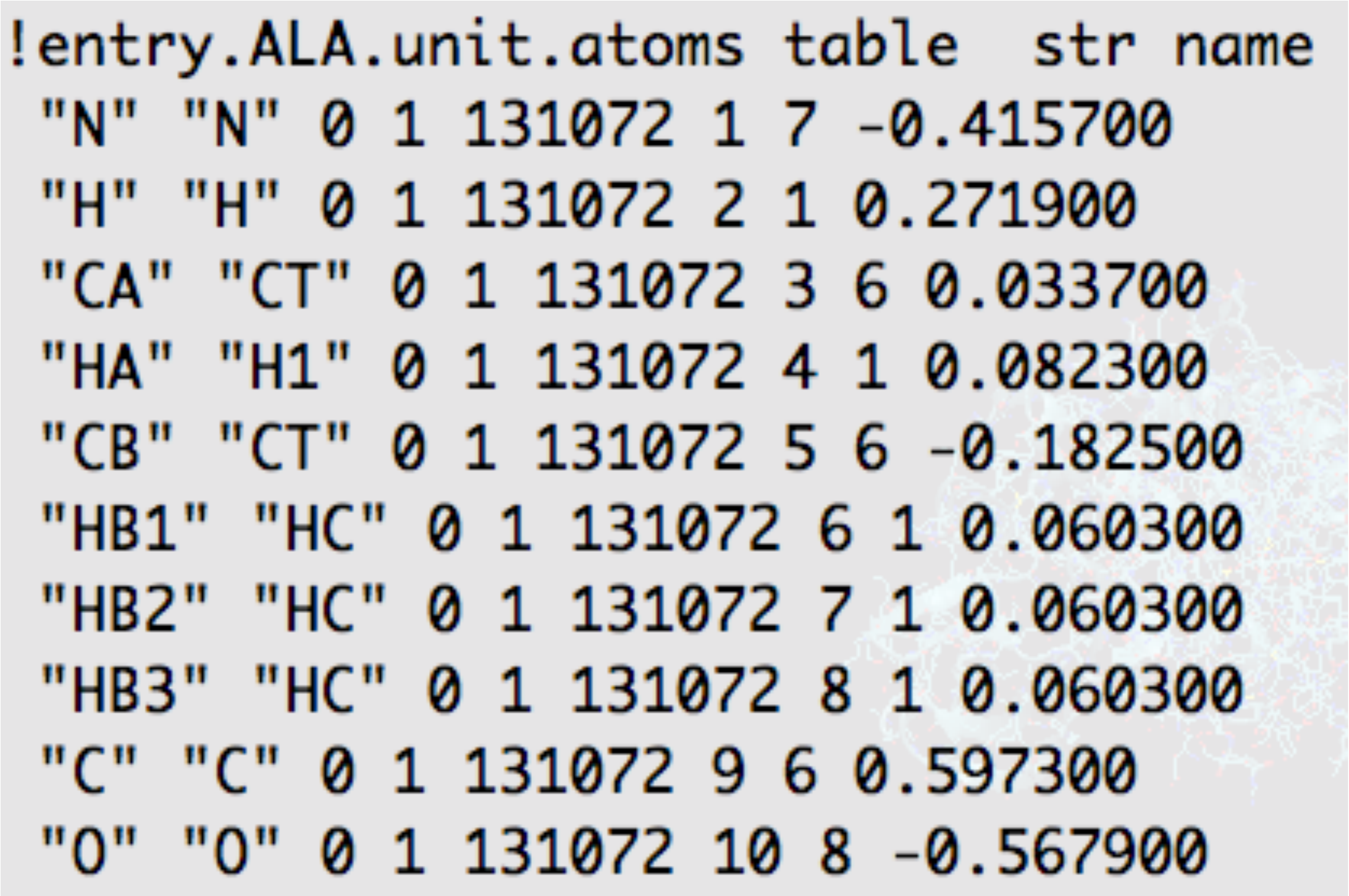
```
N -CT-C -N    1    1.700         180.000          -1.
N -CT-C -N    1    2.000         180.000           2.
C -N -CT-C    1    0.850         180.000          -2.
C -N -CT-C    1    0.800           0.000           1.
CT-CT-N -C    1    0.50          180.0            -4.          phi,psi,parm94
CT-CT-N -C    1    0.15          180.0            -3.          phi,psi,parm94
CT-CT-N -C    1    0.00            0.0            -2.          JCC,7,(1986),230
CT-CT-N -C    1    0.53            0.0             1.          phi,psi,parm94
```

A parameter file

- LJ parameters



```
H            0.6000  0.0157           !Ferguson base pair geom.
HO           0.0000  0.0000            OPLS Jorgensen, JACS,110,(1988),1657
HS           0.6000  0.0157            W. Cornell CH3SH --> CH3OH FEP
HC           1.4870  0.0157            OPLS
H1           1.3870  0.0157            Veenstra et al JCC,8,(1992),963
H2           1.2870  0.0157            Veenstra et al JCC,8,(1992),963
H3           1.1870  0.0157            Veenstra et al JCC,8,(1992),963
HP           1.1000  0.0157            Veenstra et al JCC,8,(1992),963
HA           1.4590  0.0150            Spellmeyer
H4           1.4090  0.0150            Spellmeyer, one electrowithdr. neighbor
H5           1.3590  0.0150            Spellmeyer, two electrowithdr. neighbor
HW           0.0000  0.0000            TIP3P water model
HZ           1.4590  0.0150            H bonded to sp C (Howard et al JCC 16)
O            1.6612  0.2100            OPLS
```

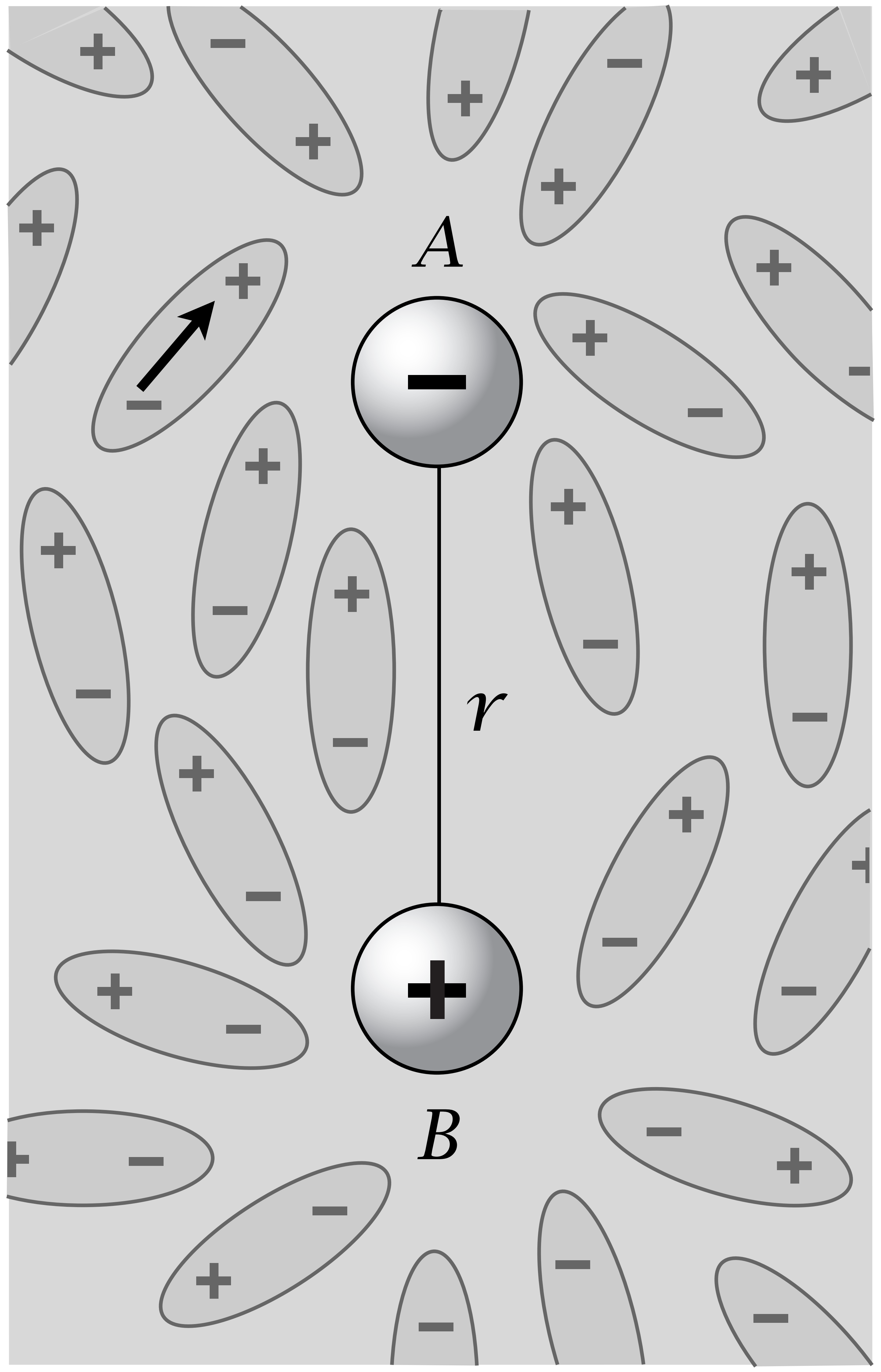


```
!entry.ALA.unit.atoms table  str name
 "N" "N" 0 1 131072 1 7 -0.415700
 "H" "H" 0 1 131072 2 1 0.271900
 "CA" "CT" 0 1 131072 3 6 0.033700
 "HA" "H1" 0 1 131072 4 1 0.082300
 "CB" "CT" 0 1 131072 5 6 -0.182500
 "HB1" "HC" 0 1 131072 6 1 0.060300
 "HB2" "HC" 0 1 131072 7 1 0.060300
 "HB3" "HC" 0 1 131072 8 1 0.060300
 "C" "C" 0 1 131072 9 6 0.597300
 "O" "O" 0 1 131072 10 8 -0.567900
```

- partial atomic charges

Electrostatics of proteins in solution

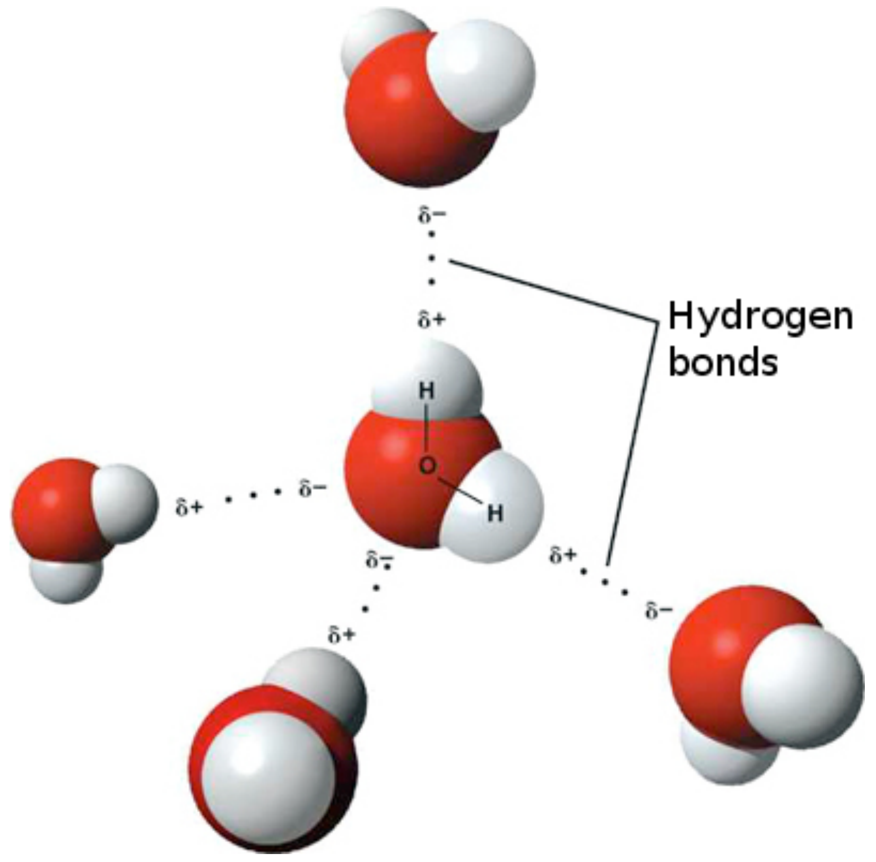
$$U(r) = \frac{1}{4\pi\epsilon_0}\frac{1}{D}\frac{q_1 q_2}{r}$$

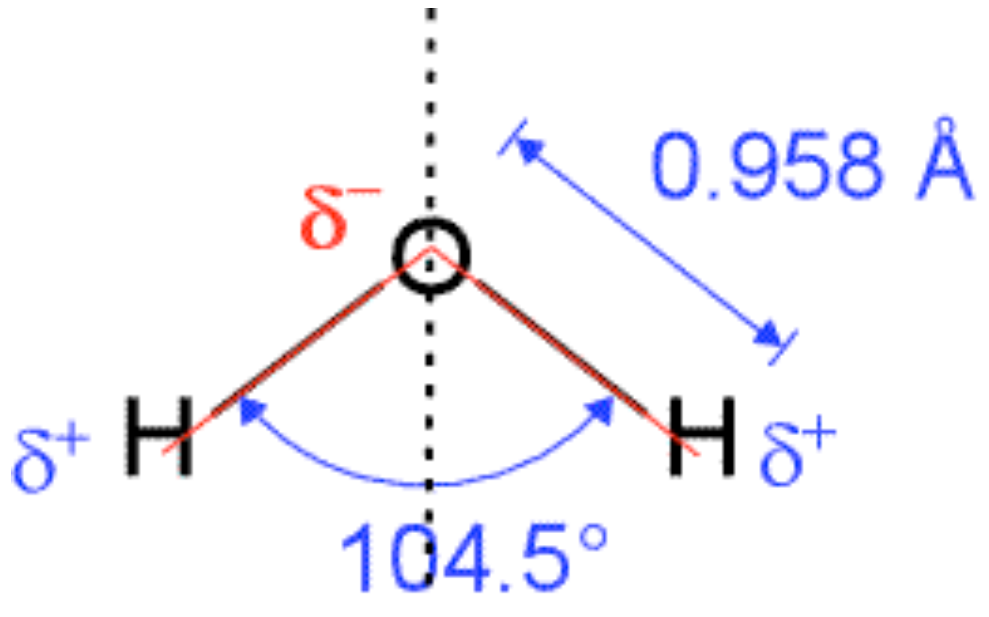


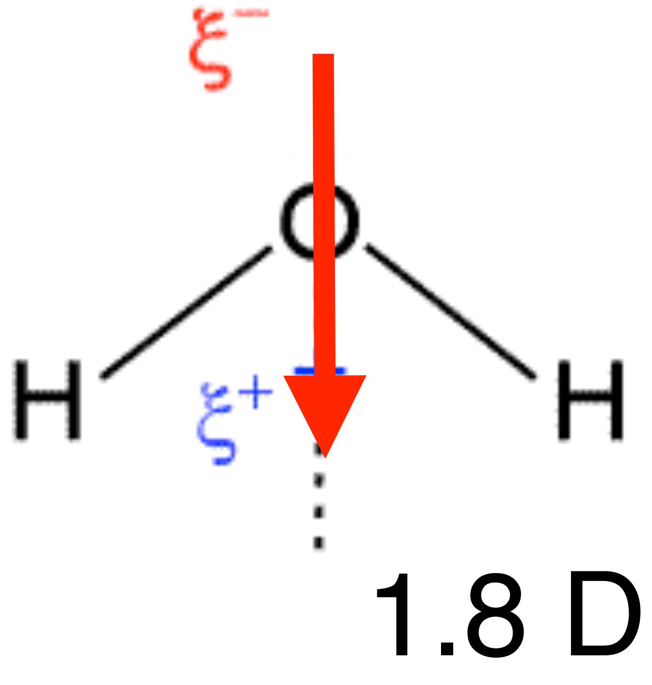
medium is polarizable or
has permanent dipole

D = dielectric constant
[air(T=0°C) D=1.00059]

Liquid	*T* (°C)	*D*
Heptane	0	1.958
Heptane	30	1.916
Methanol	25	33
Formamide	20	109
Formic acid	16	58
Nitrobenzene	25	35
HCN	0	158
HCN	20	114
Glycol	25	37
Water	0	88.00
Water	25	78.54

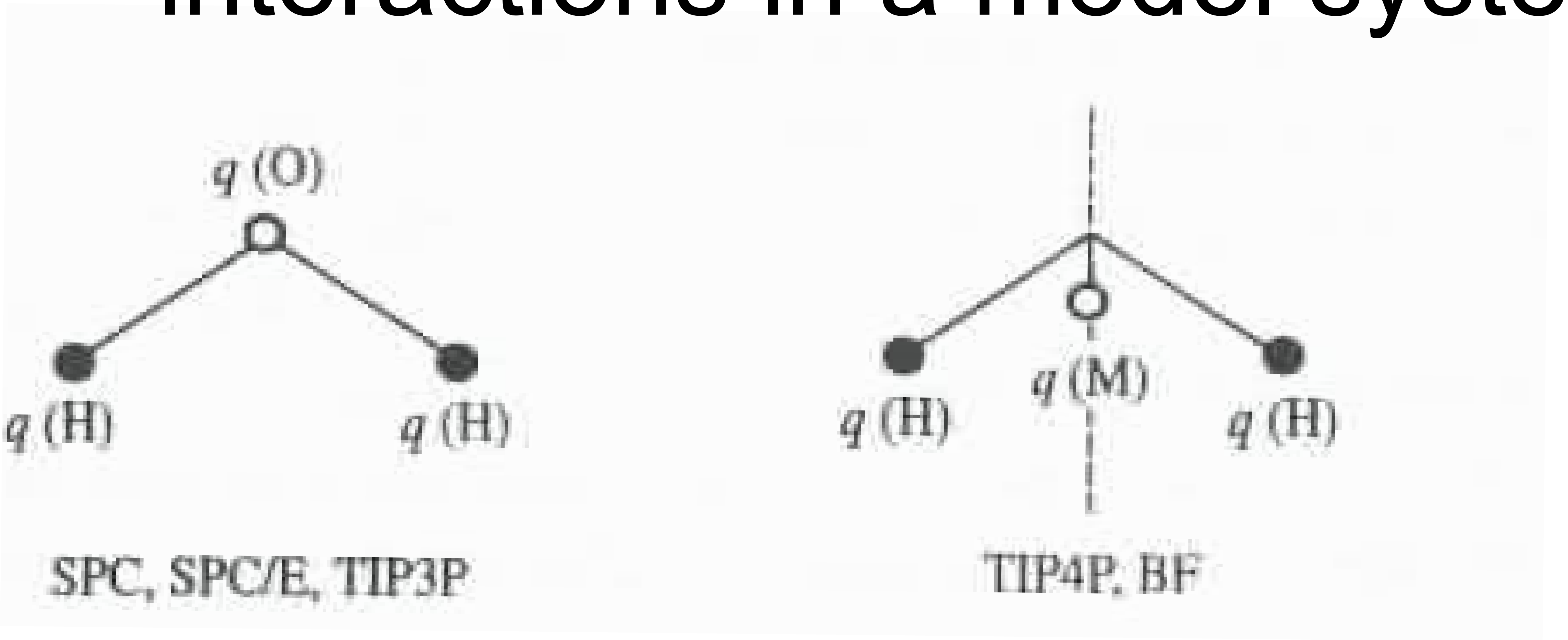


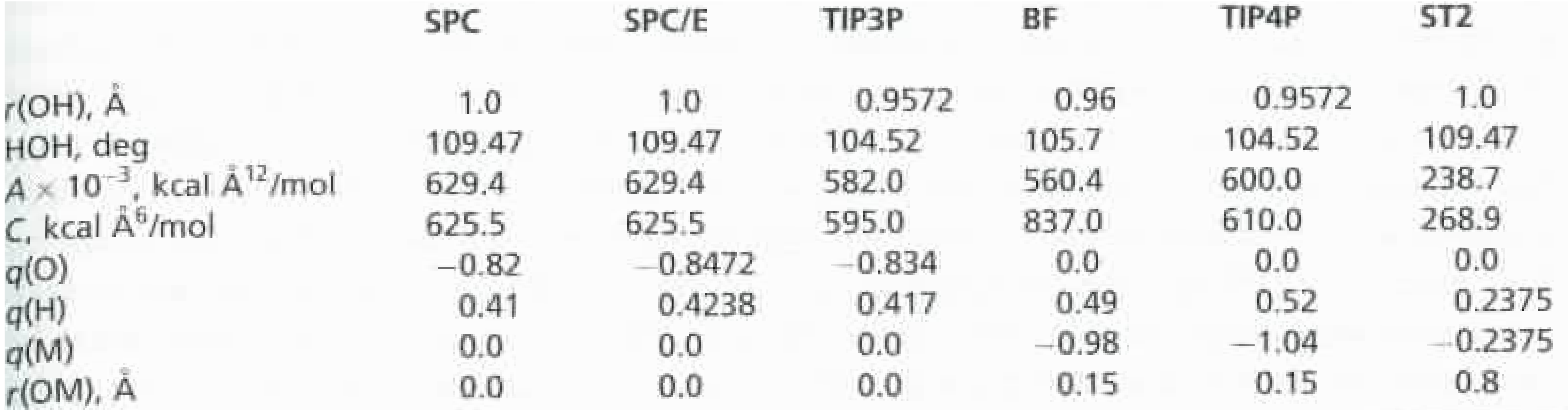




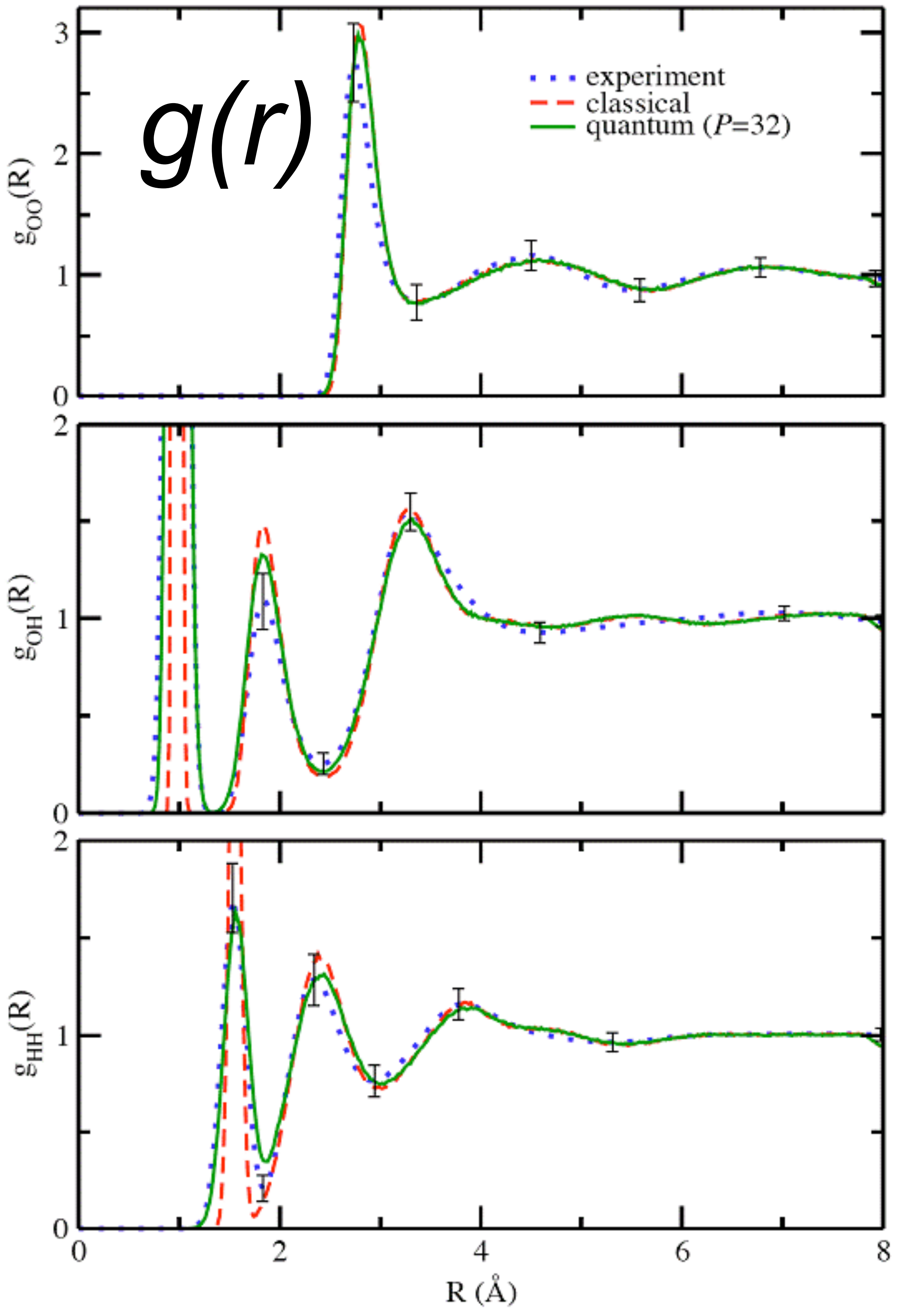
Solvation and water models

- several models (e.g. rigid, flexible, polarizable)
- fitted using simulations on density, g(r), heat capacity, diffusion coefficient, dielectric constant, etc.
- explicit water treatment is costly (~2/3 of non-bonded interactions in a model system)



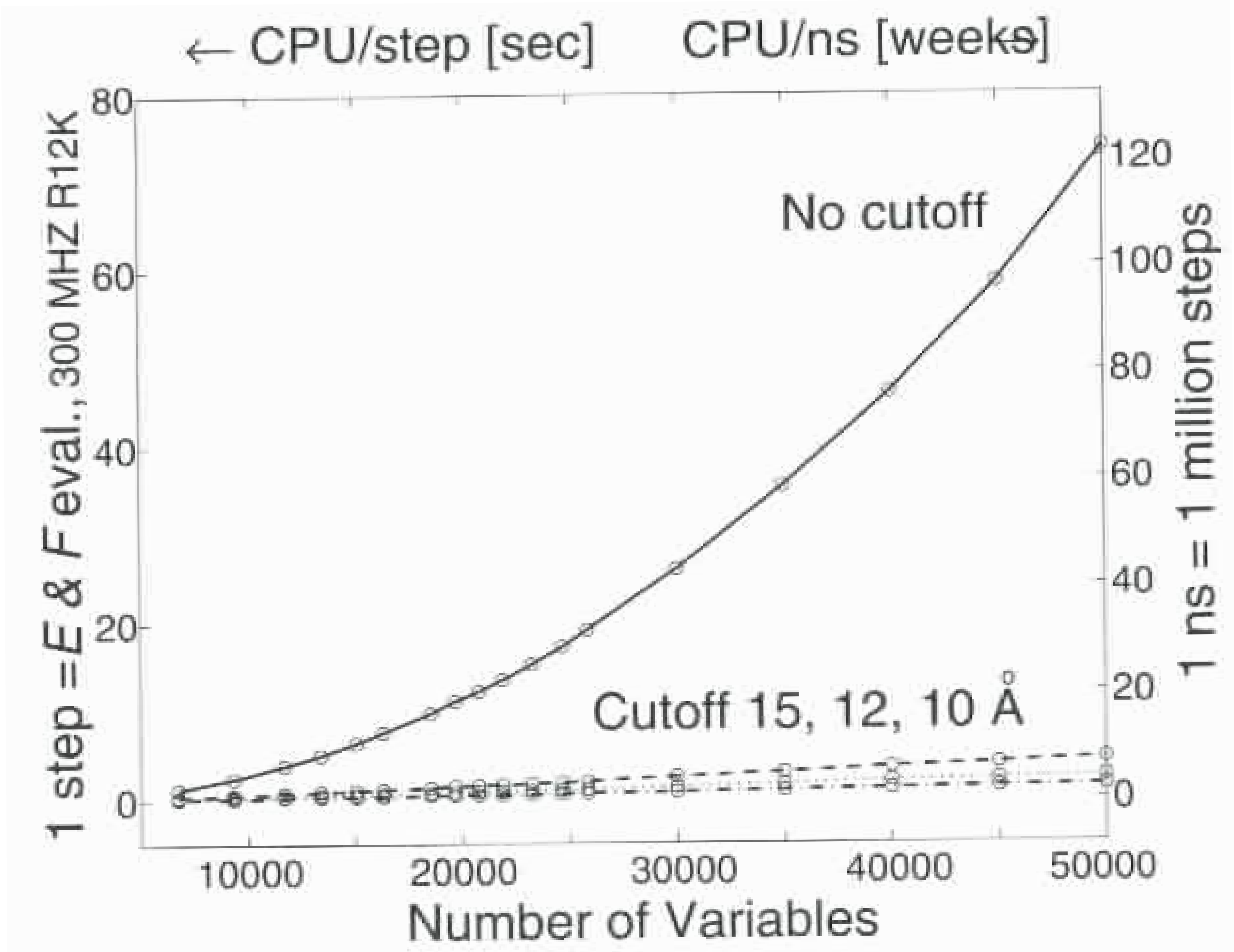


	SPC	SPC/E	TIP3P	BF	TIP4P	ST2
r(OH), Å	1.0	1.0	0.9572	0.96	0.9572	1.0
HOH, deg	109.47	109.47	104.52	105.7	104.52	109.47
$A \times 10^{-3}$, kcal Å^{12}/mol	629.4	629.4	582.0	560.4	600.0	238.7
C, kcal Å^{6}/mol	625.5	625.5	595.0	837.0	610.0	268.9
q(O)	−0.82	−0.8472	−0.834	0.0	0.0	0.0
q(H)	0.41	0.4238	0.417	0.49	0.52	0.2375
q(M)	0.0	0.0	0.0	−0.98	−1.04	−0.2375
r(OM), Å	0.0	0.0	0.0	0.15	0.15	0.8



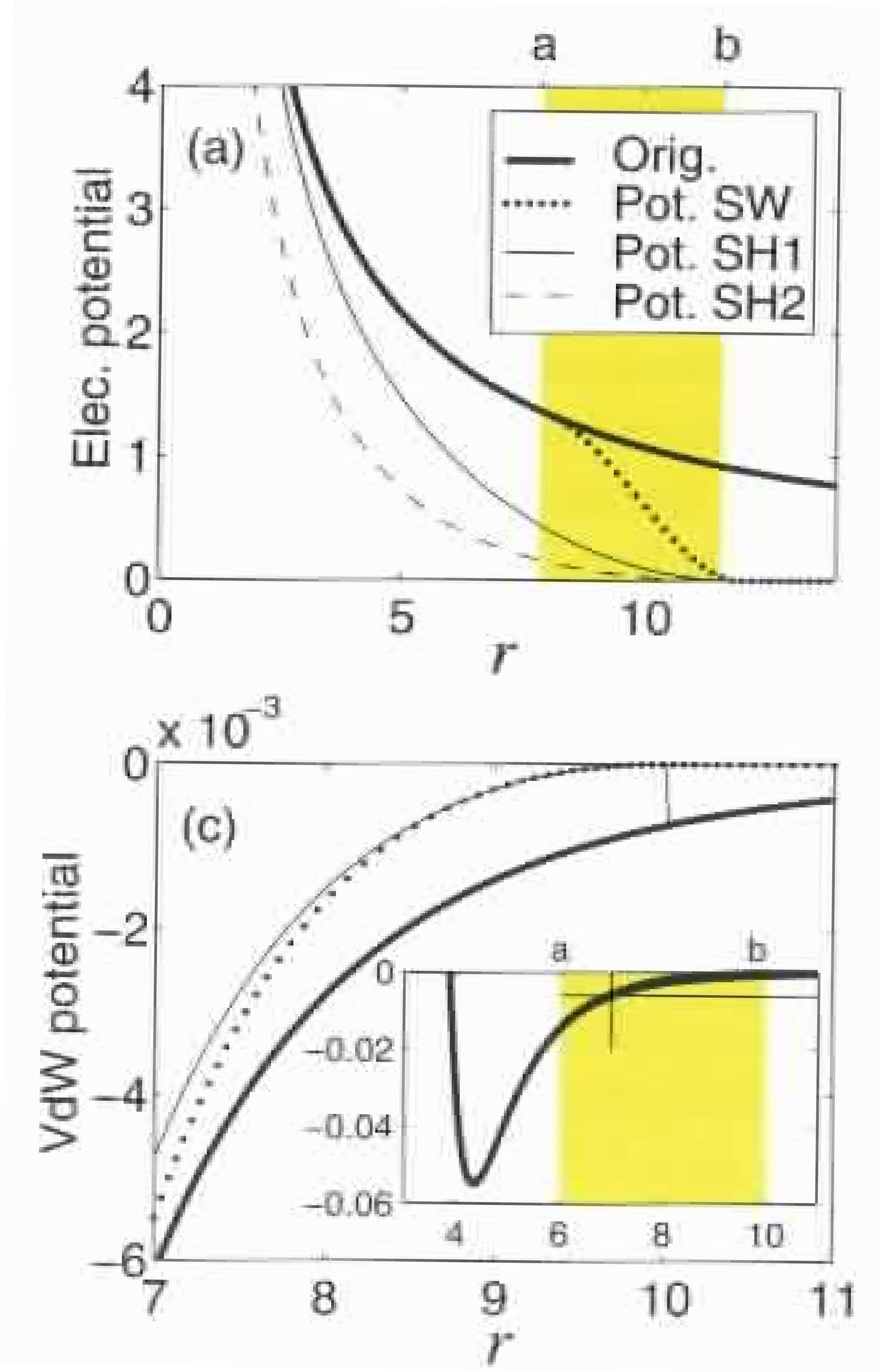
Non-bonded computation

- configuration space of a system on N atoms is 3N-6 (rigid-body rotation and translation of the system are removed for energy conservation)
- for protein and nucleic acid N~10^4 or more, if you add water N~10^5
- non-bonded part is real computational bottleneck, *O*(N^2) as compared to linear-scaling of bonded interactions
- different methods to decrease the computational cost:
- 1. cutoff schemes O(N)
- 2. Ewald methods O(N logN)
- 3. fast multipole schemes



Spherical cutoff schemes

- **truncation**: cutoff radius *r* set equal to *b*, interactions are abruptly set to zero for *r>b*, and are not altered for *r<b*
- **switching***:* define a switching region *[a,b]* where interactions are smoothly brought to zero
- **shift**: interactions are gradually smoothed for all *r<b*

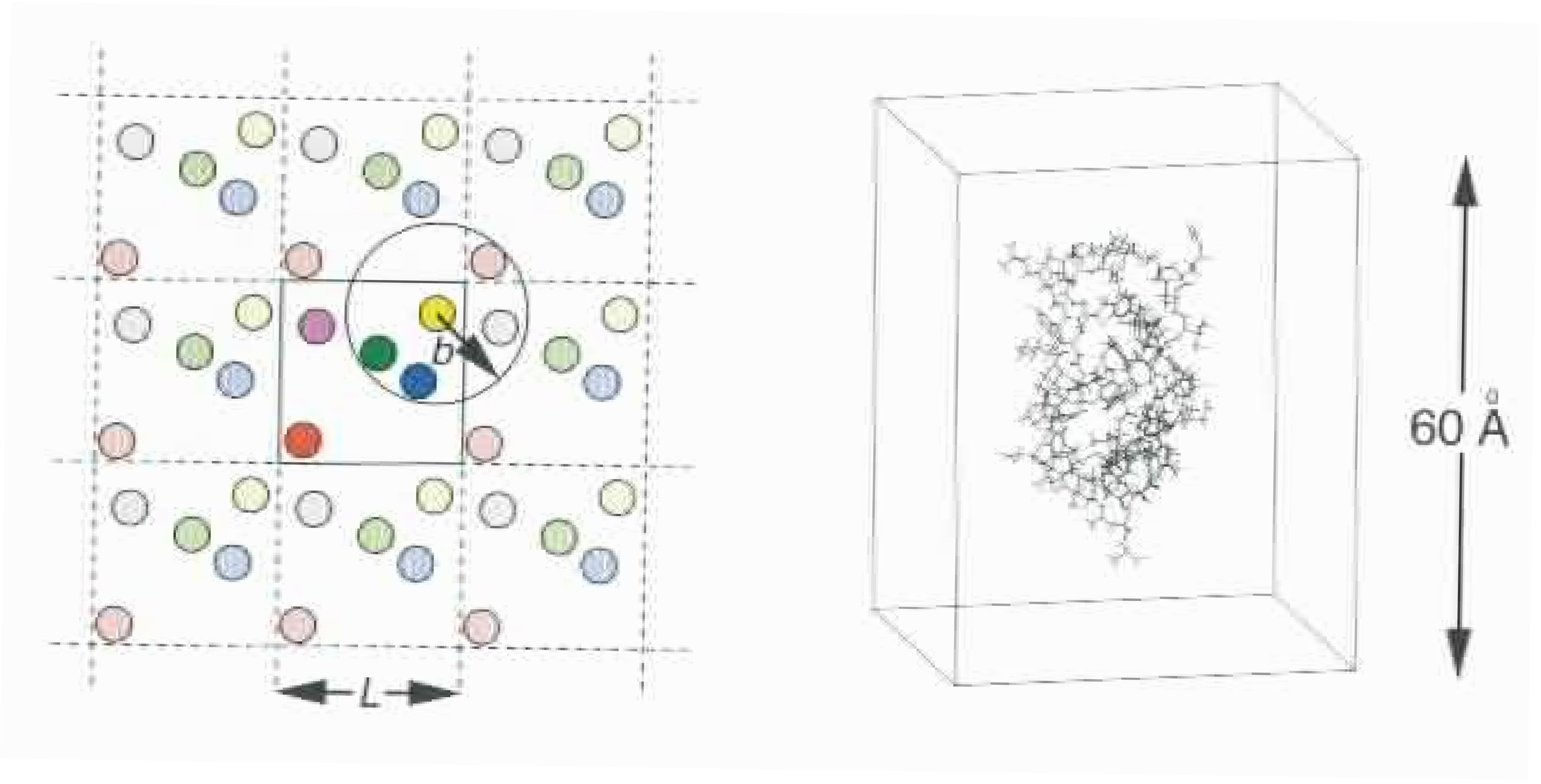


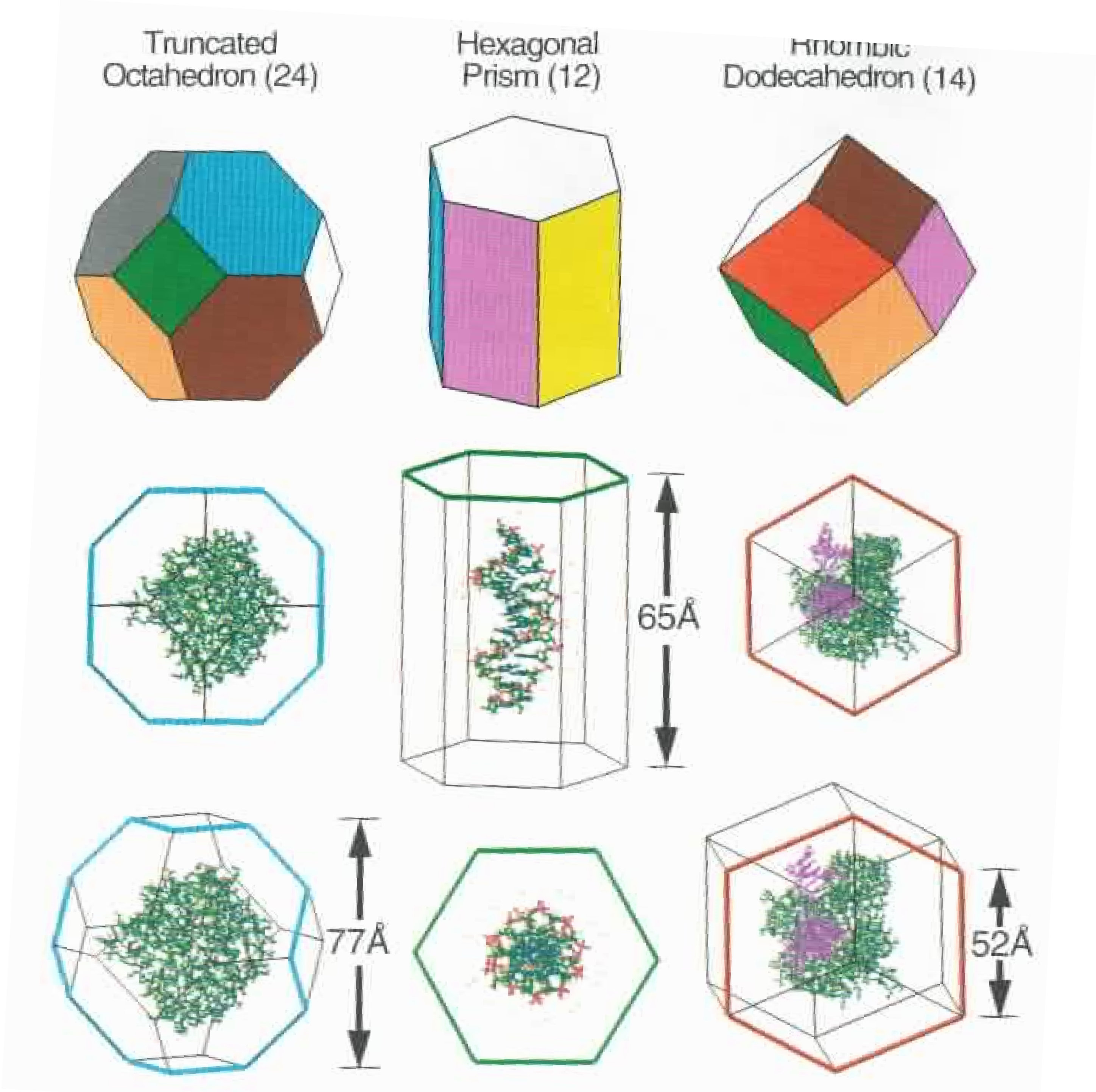
$$U_{NB}(r) = \sum_{i>j}^{N} w_{ij} S(r_{ij}) \left[\left(\frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{6}} \right) + \sum_{i>j}^{N} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \right]$$

- *S(r_{ij})* is the distance-dependent switch or shift function

Ewald methods

- used for calculating electrostatic energy of systems in **periodic boundary conditions** (unit cell charge = 0)
- **minimum-image convention:** each atom interacts with the closest periodic image of the other N-1 atoms
- different unit cell lattice geometry
- use of **fast fourier transforms** to compute the electrostatic energy in the real and reciprocal lattice





MM FF limitations

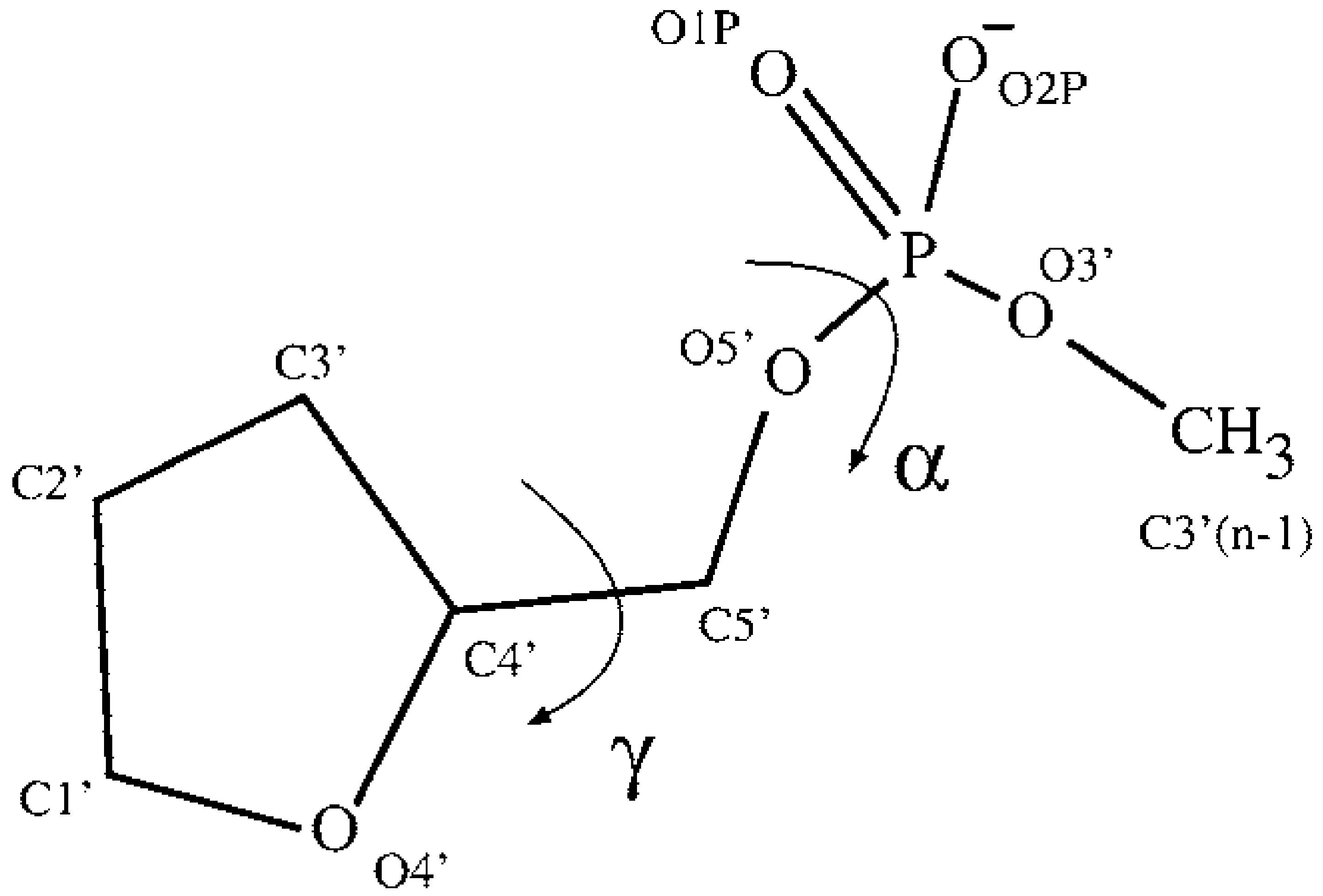
- transferability
- accuracy of parametrization
- functional form (e.g. can add polarizability)

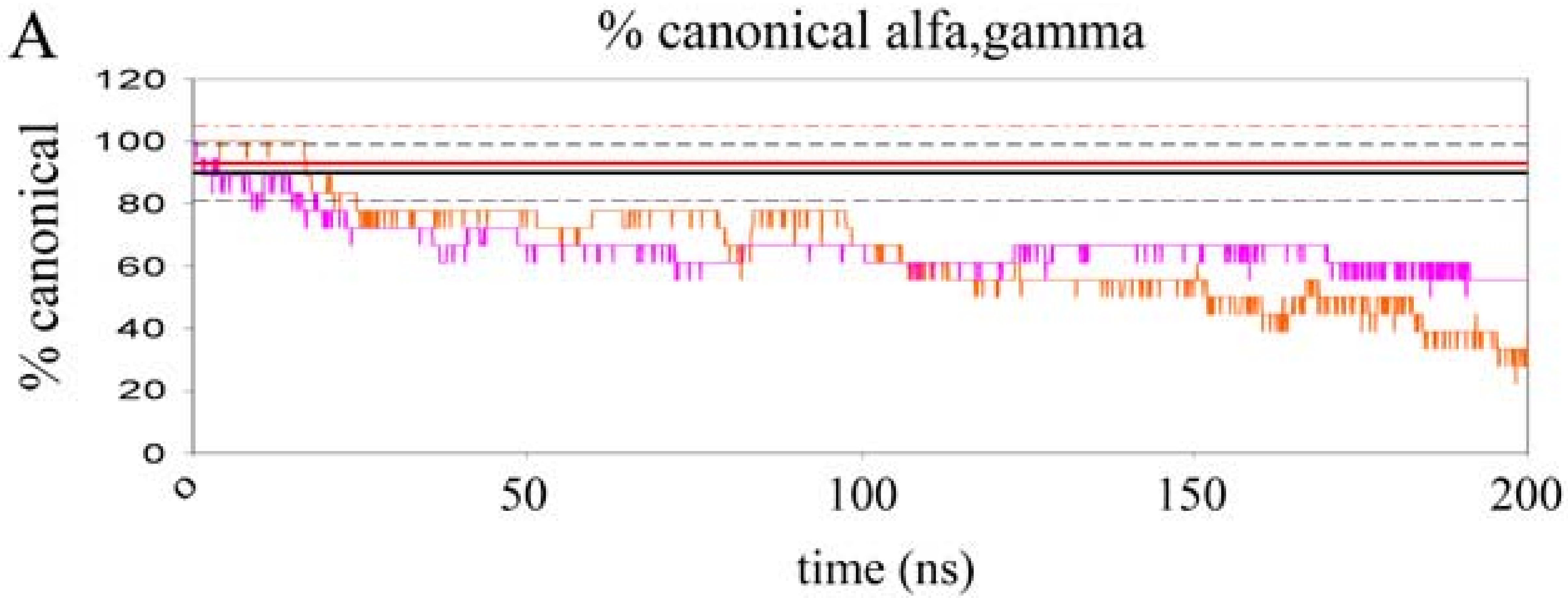
  $\mu_{ind} = \alpha \mathbf{E} \qquad \alpha : polarizability$

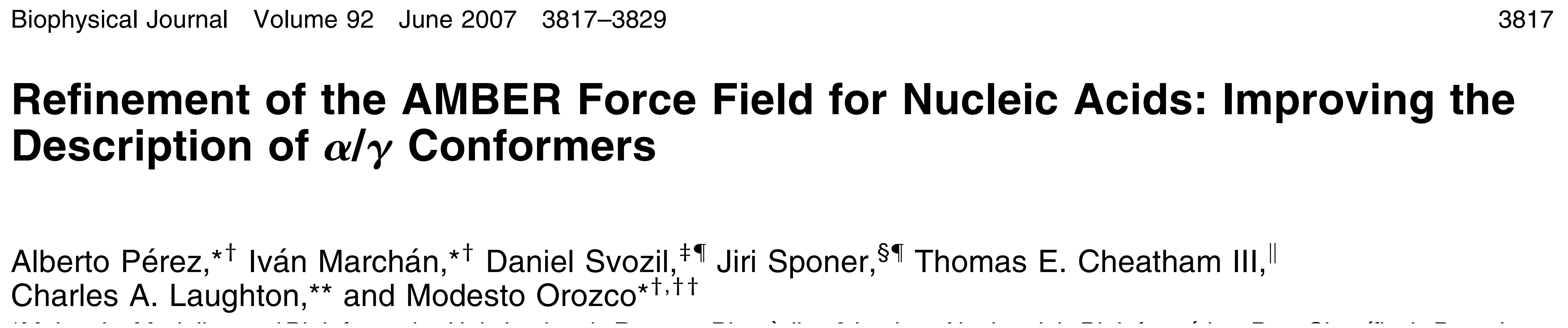
  or many-body terms
- many different force fields (specific vs. generalized)
- approximation in treating long-range interactions
- can be expensive for very large systems (e.g. ~10^6 atoms)

Failure of a force field

- enhanced computer power allows to run longer MD simulations, and to discover failures in the models



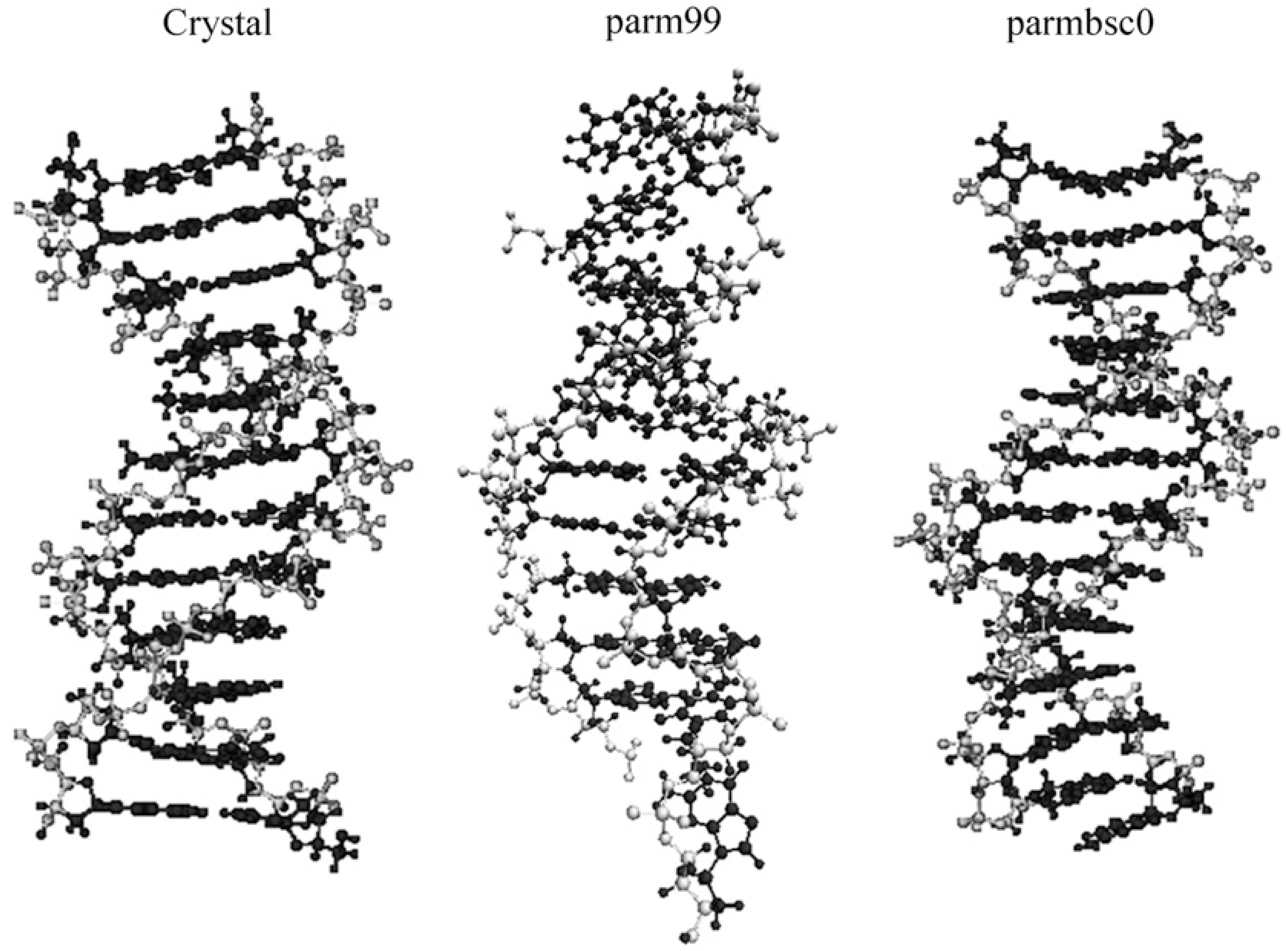




Biophysical Journal Volume 92 June 2007 3817–3829 3817

Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers

Alberto Pérez,*† Iván Marchán,*† Daniel Svozil,‡¶ Jiri Sponer,§¶ Thomas E. Cheatham III,‖ Charles A. Laughton,** and Modesto Orozco*†,††



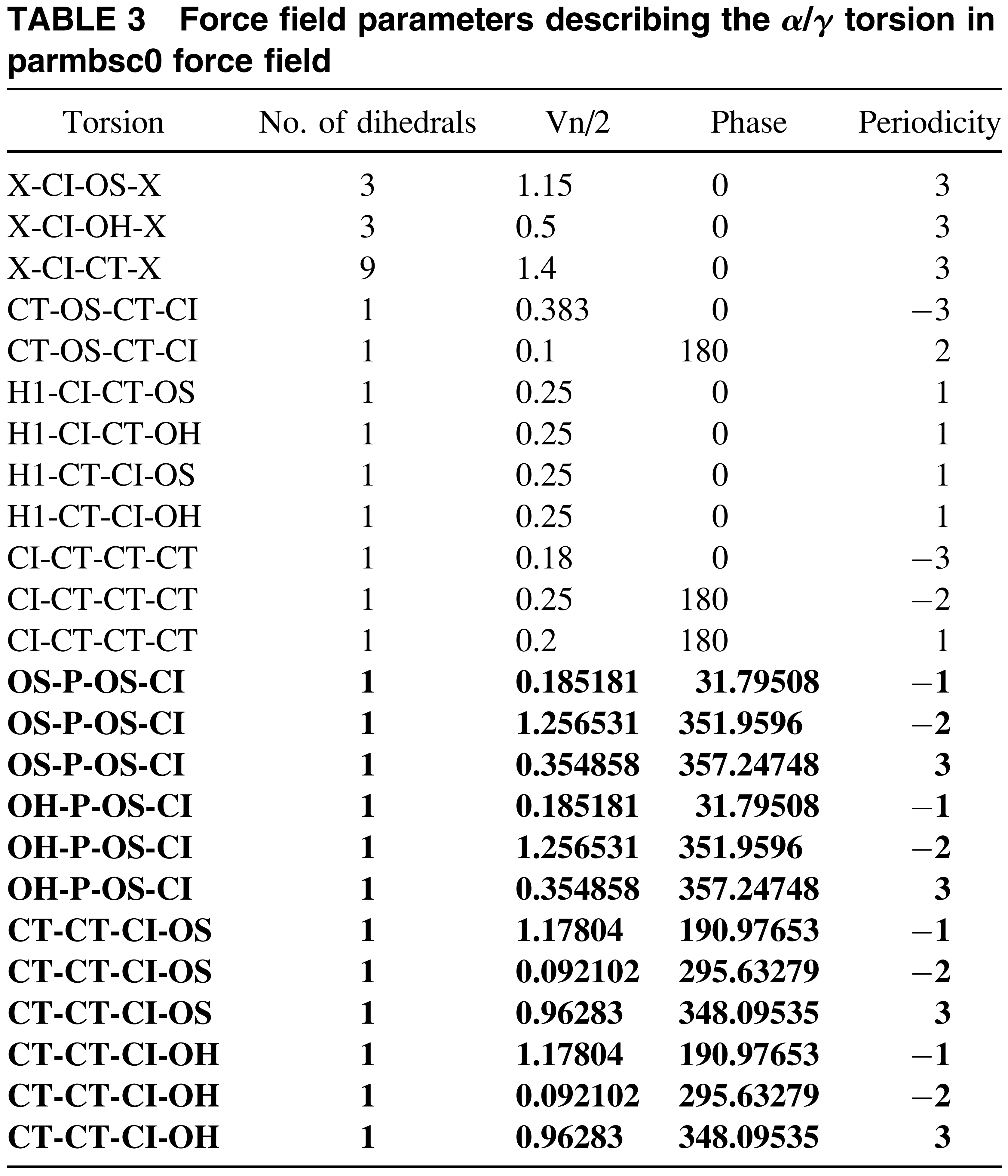


TABLE 3 Force field parameters describing the α/γ torsion in parmbsc0 force field

Torsion	No. of dihedrals	Vn/2	Phase	Periodicity
X-CI-OS-X	3	1.15	0	3
X-CI-OH-X	3	0.5	0	3
X-CI-CT-X	9	1.4	0	3
CT-OS-CT-CI	1	0.383	0	−3
CT-OS-CT-CI	1	0.1	180	2
H1-CI-CT-OS	1	0.25	0	1
H1-CI-CT-OH	1	0.25	0	1
H1-CT-CI-OS	1	0.25	0	1
H1-CT-CI-OH	1	0.25	0	1
CI-CT-CT-CT	1	0.18	0	−3
CI-CT-CT-CT	1	0.25	180	−2
CI-CT-CT-CT	1	0.2	180	1
OS-P-OS-CI	**1**	**0.185181**	**31.79508**	**−1**
OS-P-OS-CI	**1**	**1.256531**	**351.9596**	**−2**
OS-P-OS-CI	**1**	**0.354858**	**357.24748**	**3**
OH-P-OS-CI	**1**	**0.185181**	**31.79508**	**−1**
OH-P-OS-CI	**1**	**1.256531**	**351.9596**	**−2**
OH-P-OS-CI	**1**	**0.354858**	**357.24748**	**3**
CT-CT-CI-OS	**1**	**1.17804**	**190.97653**	**−1**
CT-CT-CI-OS	**1**	**0.092102**	**295.63279**	**−2**
CT-CT-CI-OS	**1**	**0.96283**	**348.09535**	**3**
CT-CT-CI-OH	**1**	**1.17804**	**190.97653**	**−1**
CT-CT-CI-OH	**1**	**0.092102**	**295.63279**	**−2**
CT-CT-CI-OH	**1**	**0.96283**	**348.09535**	**3**

Current common force fields

- **CHARMM**: Karplus Harvard, http://www.charmm.org/
- **AMBER**: Kollman UCSF, http://ambermd.org/
- **GROMOS**: van Gunsteren, ETHZ, www.igc.ethz.ch/GROMOS/index
- **OPLS**: Jorgensen, Yale, http://zarbi.chem.yale.edu/
- **DESMOND**: Shaw, http://www.deshawresearch.com/
- **AMOEBA-2013:** polarizable protein force field
- ...

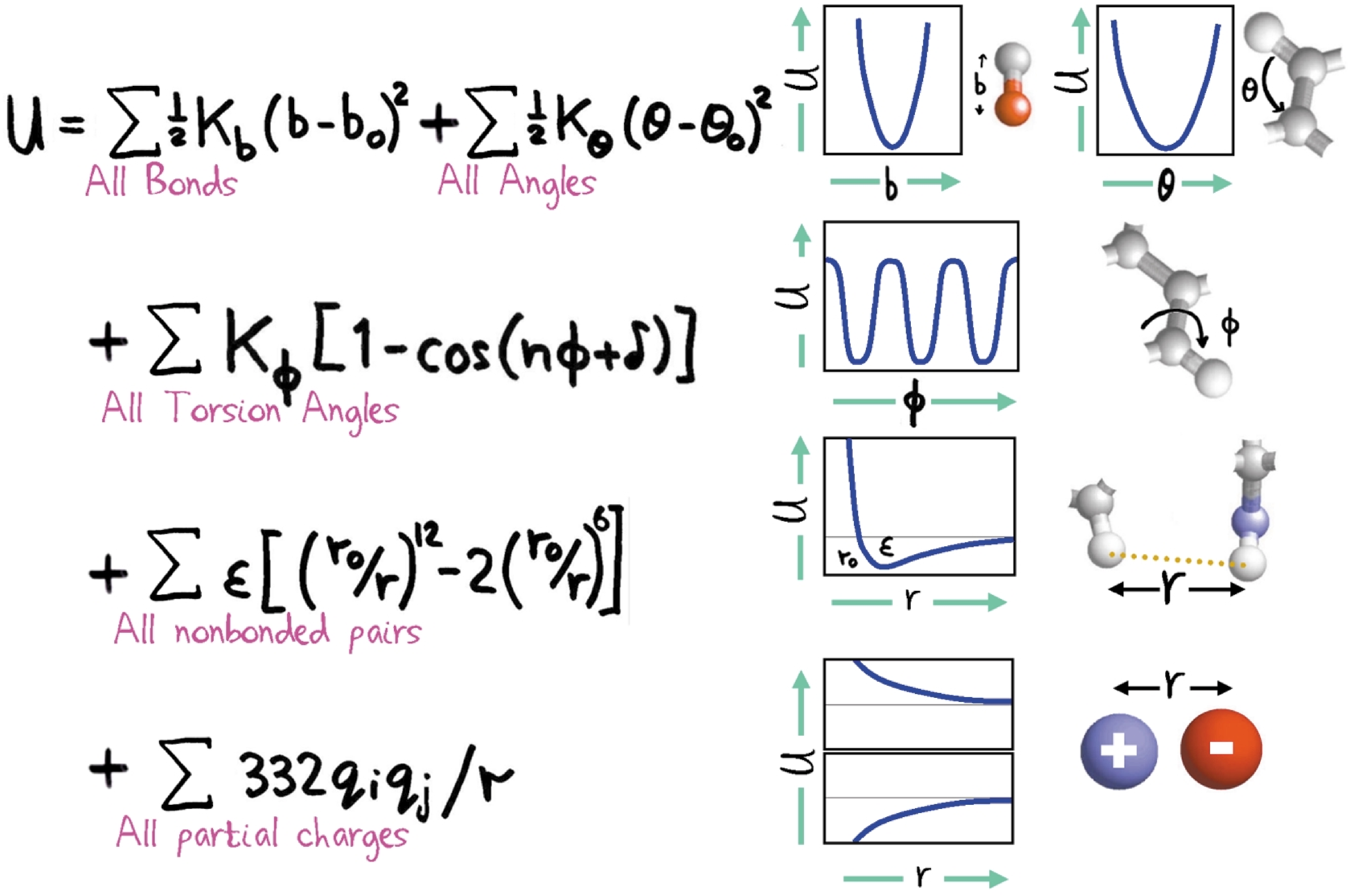
References

Molecular modeling:

- Leach, *Molecular Modeling**
- Schlick, *Molecular Modeling and Simulation **
- Dill & Bromberg, *Molecular driving forces*

Classic papers (reading assignments):

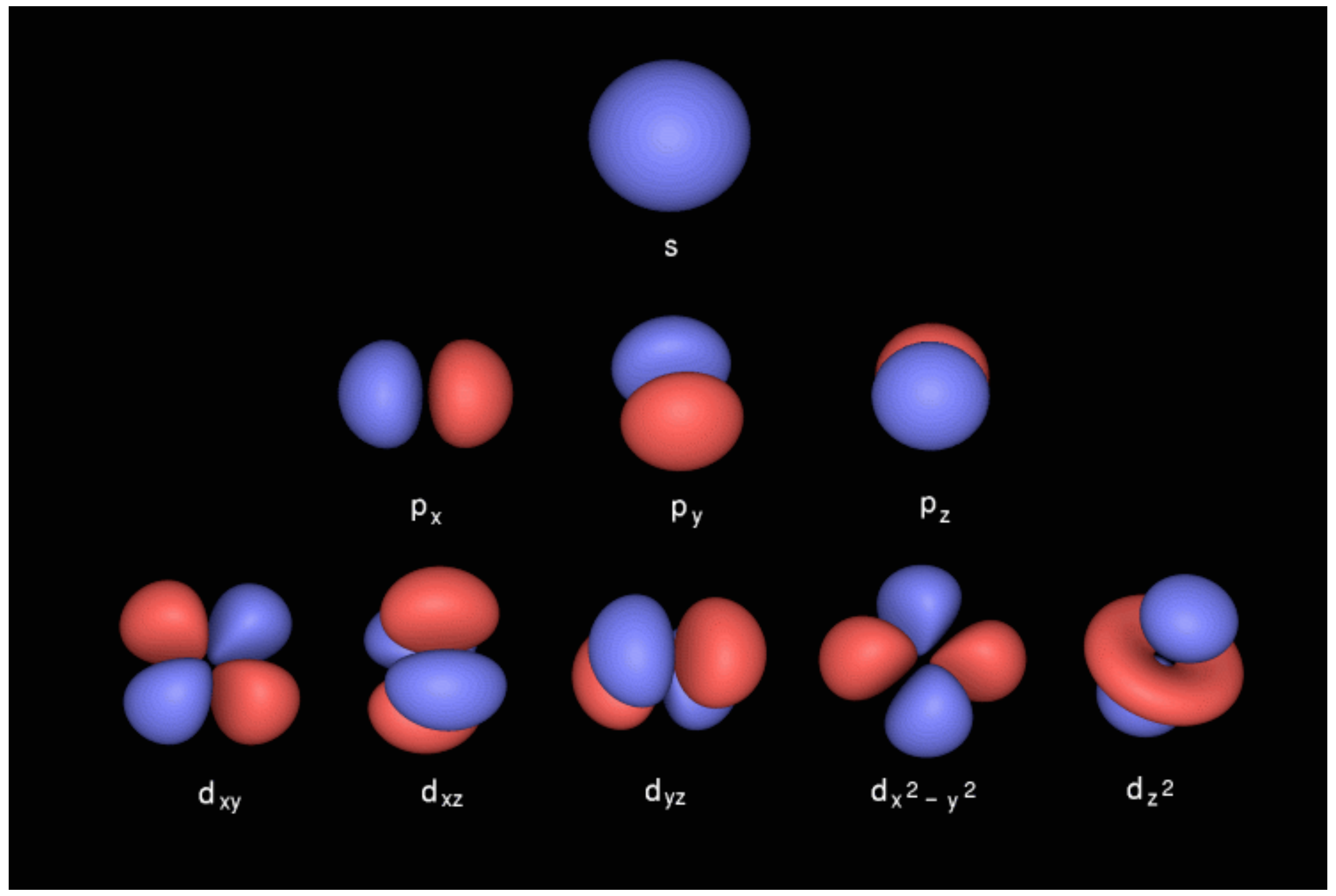
1. M. Levitt, The birth of computational structural biology, Nature Structural Biology, 8:392, 2001
2. Karplus & McCammon, Molecular dynamics simulations of biomolecules Nature Structural Biology, 9:646, 2002

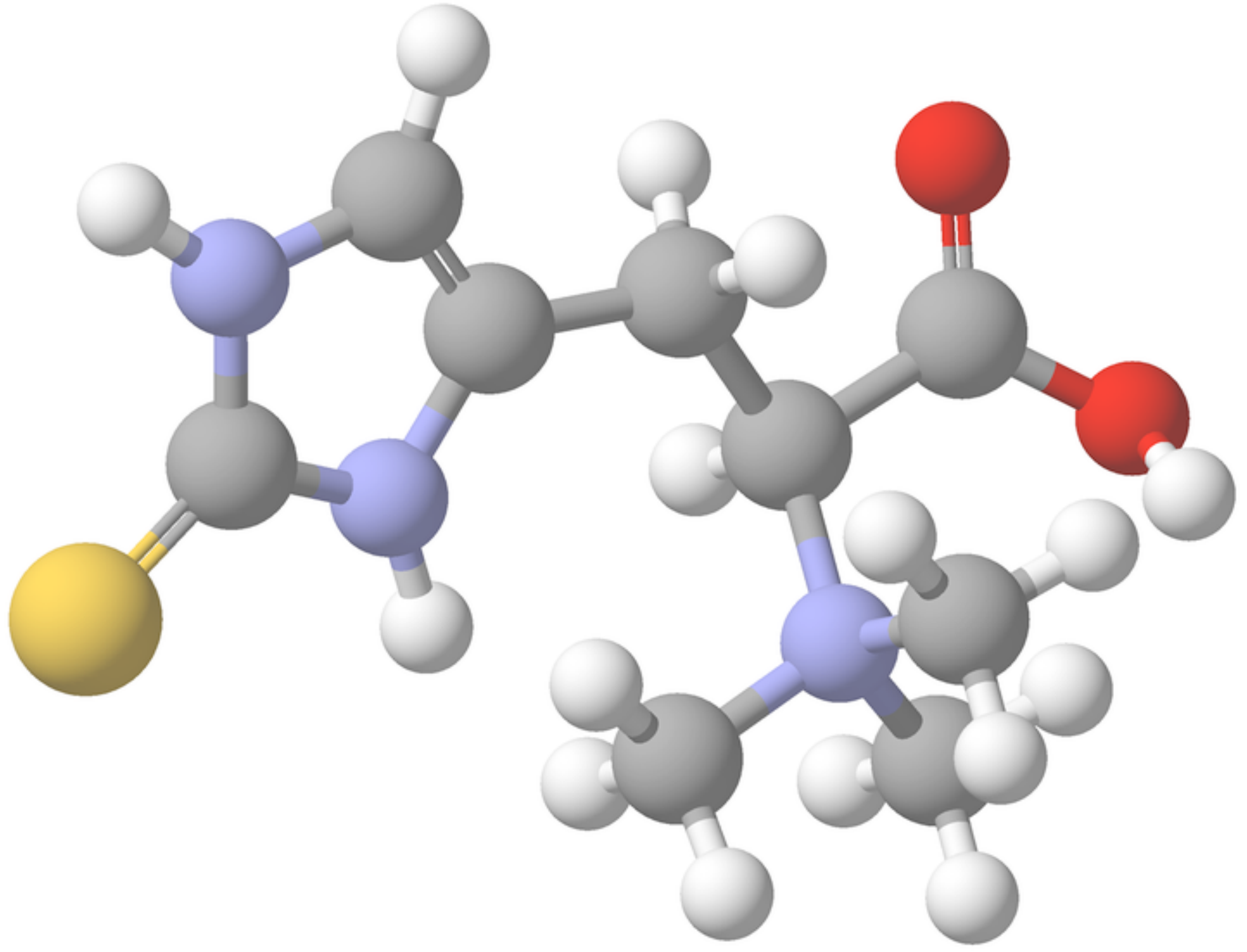


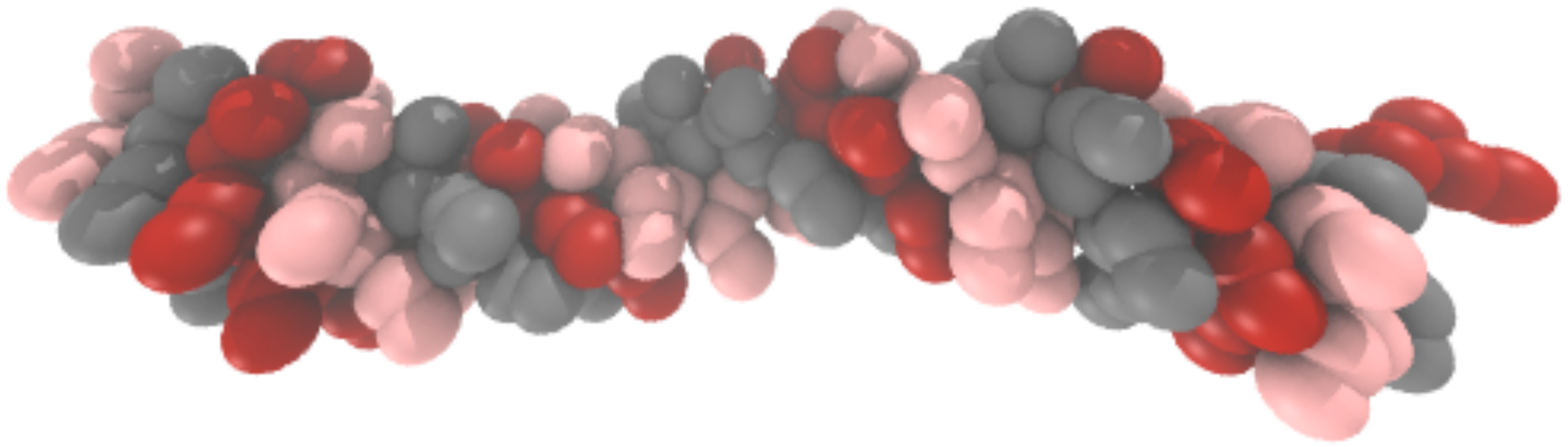
* **figures and tables taken from these books**

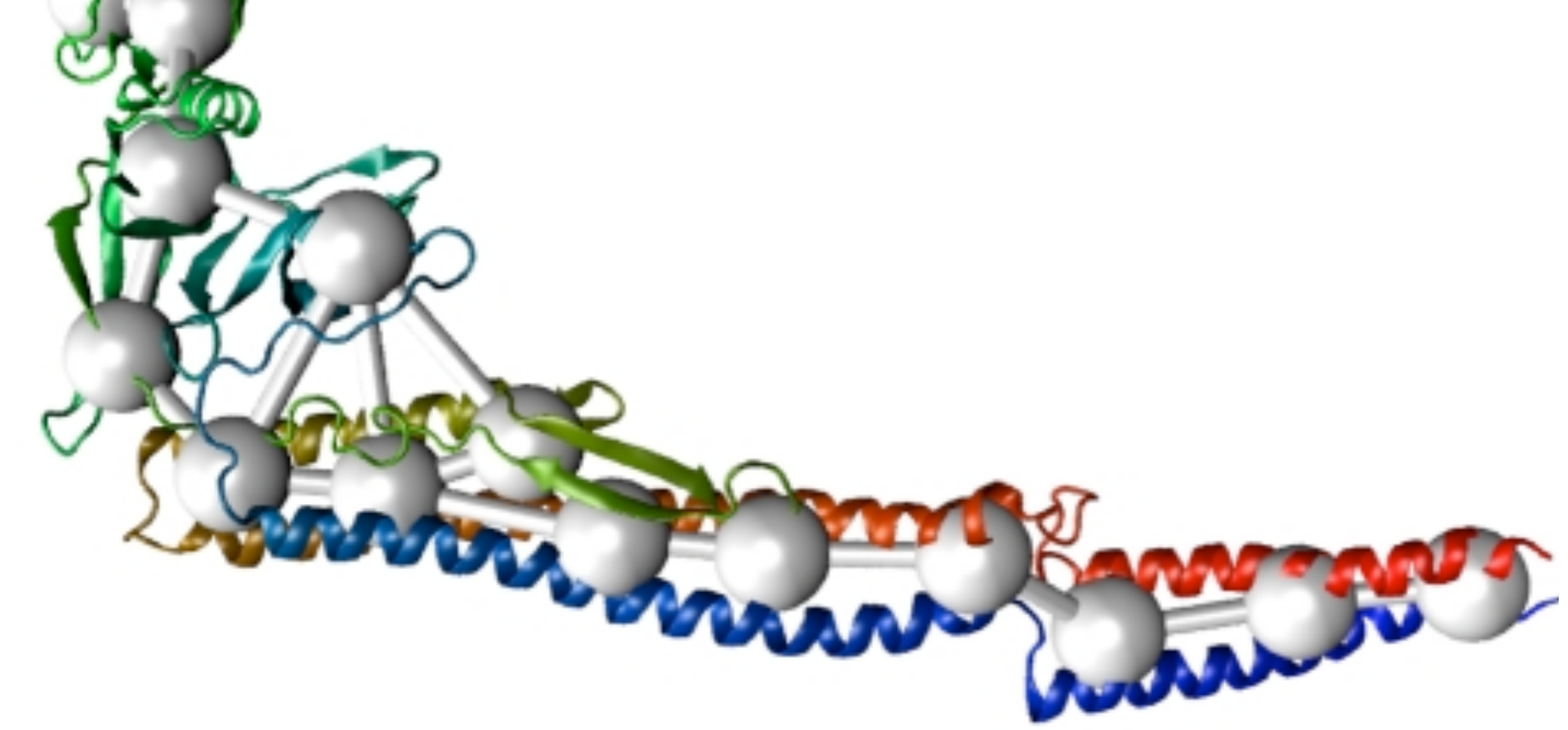
Multiscale resolution in modeling

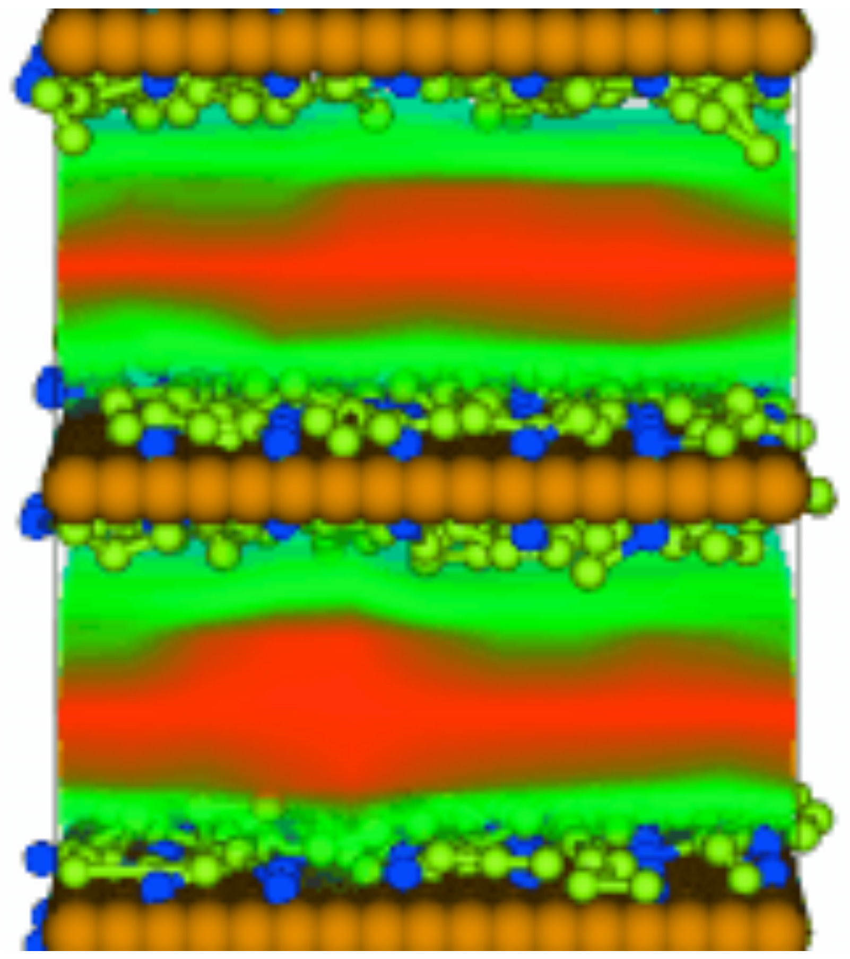
- electrons
- atoms
- amino-acids
- domains
- mesoscopic to continuum



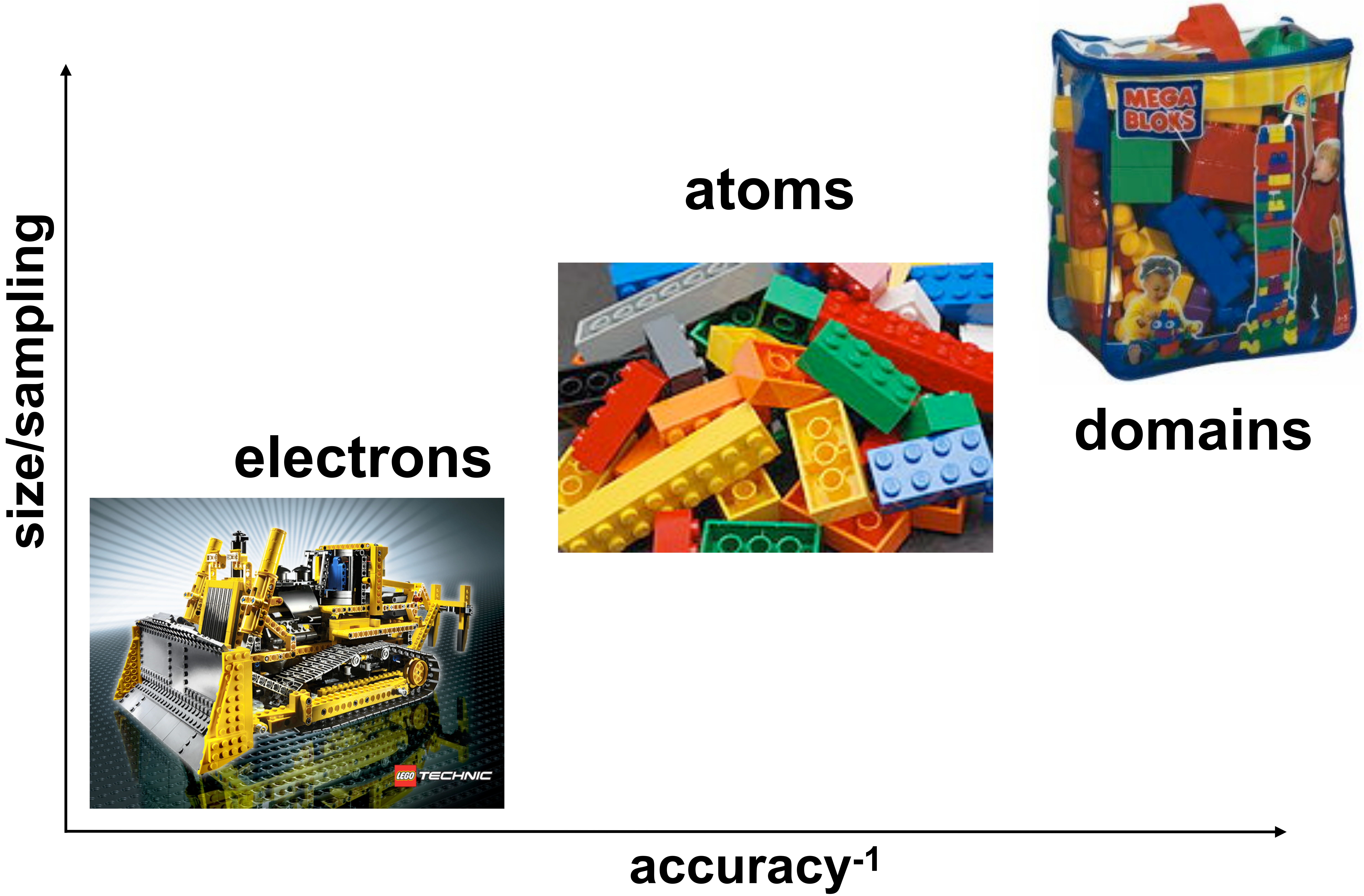








Building blocks



size/sampling

electrons

atoms

domains

accuracy^{-1}


chemical detail
sampling
binding properties
folding
mesoscale
enzymatic reactions
coarse-grained
MM-CG
10^8
10^6
10^4
10^2
size/atoms
QM-MM
MM
QM
structural fluctuations
self-assembly
fs
ps
ns
μs
ms
s
timescale

Speeding up timescales of Chemical Reactions

- **Enzymes** enhance the rate of chemical reactions by several orders of magnitude (e.g. arginine decarboxylase, alkaline phosphatase, staphylococcal nuclease **up to 10^{14} fold**)
- the transition rate depends on the activation barrier

$$\Gamma_{reactants \rightarrow products} \propto e^{-G_{barrier}/k_B T}$$

- and enzymes affect this, not the R and P states

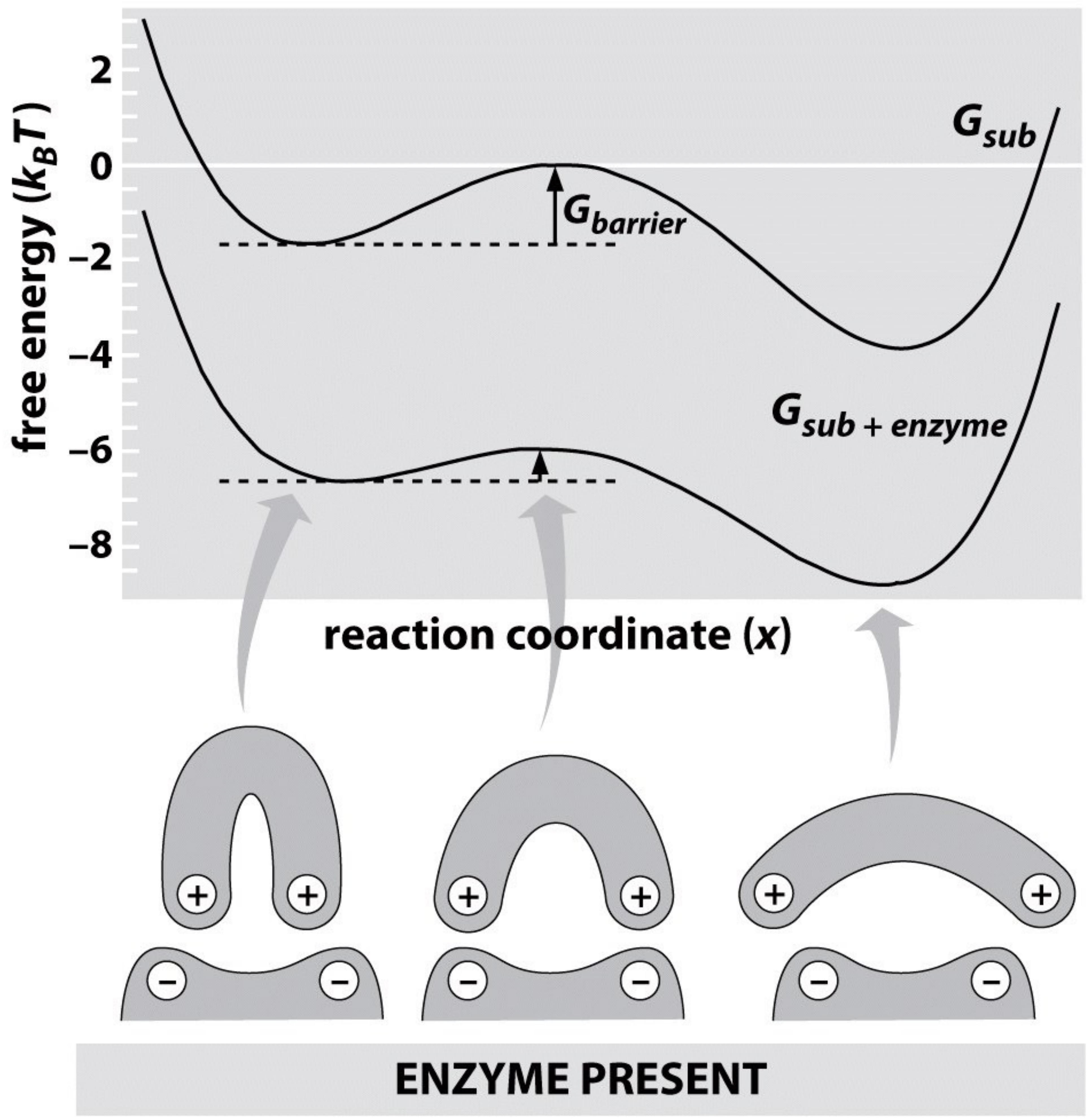
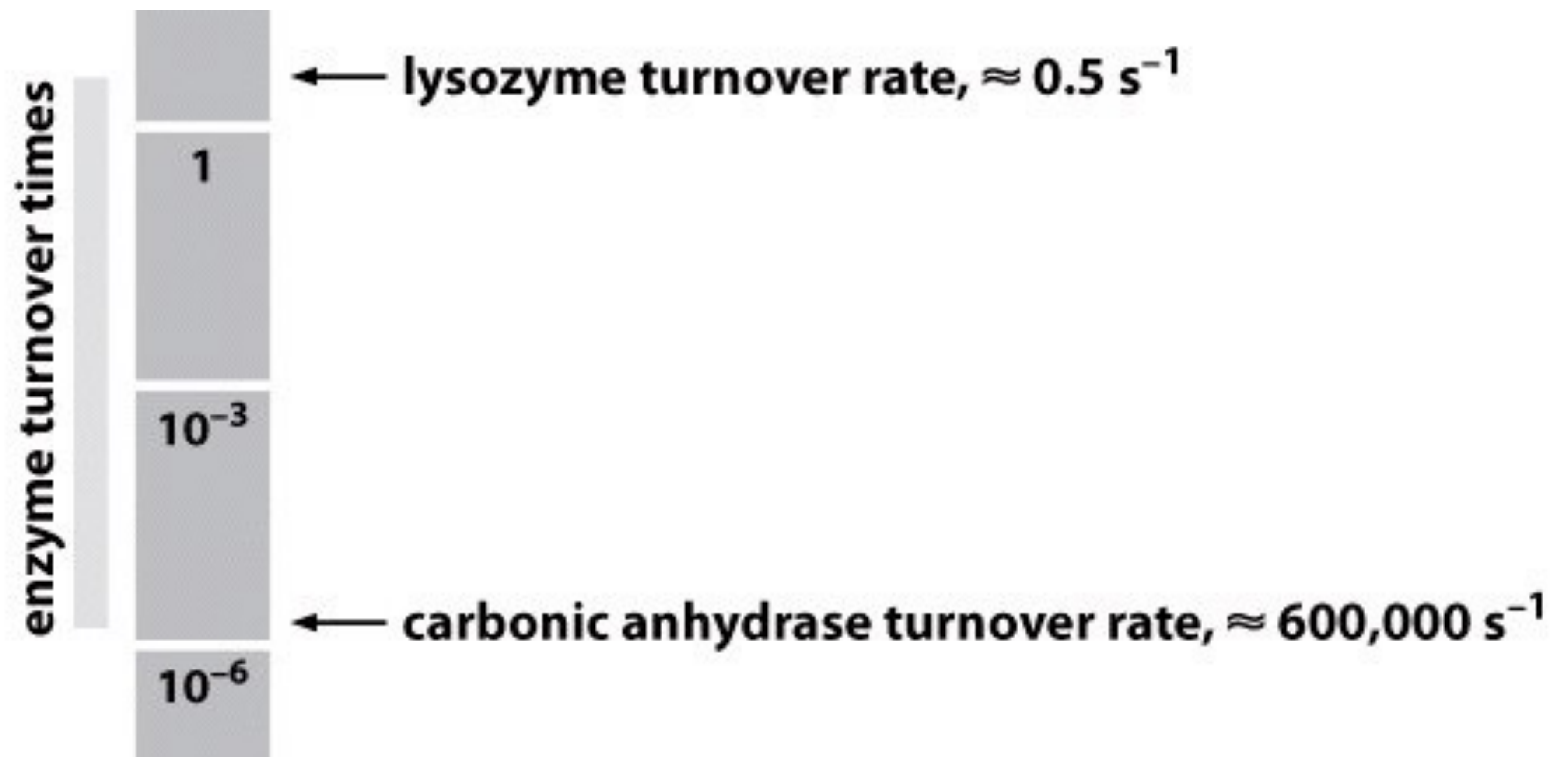
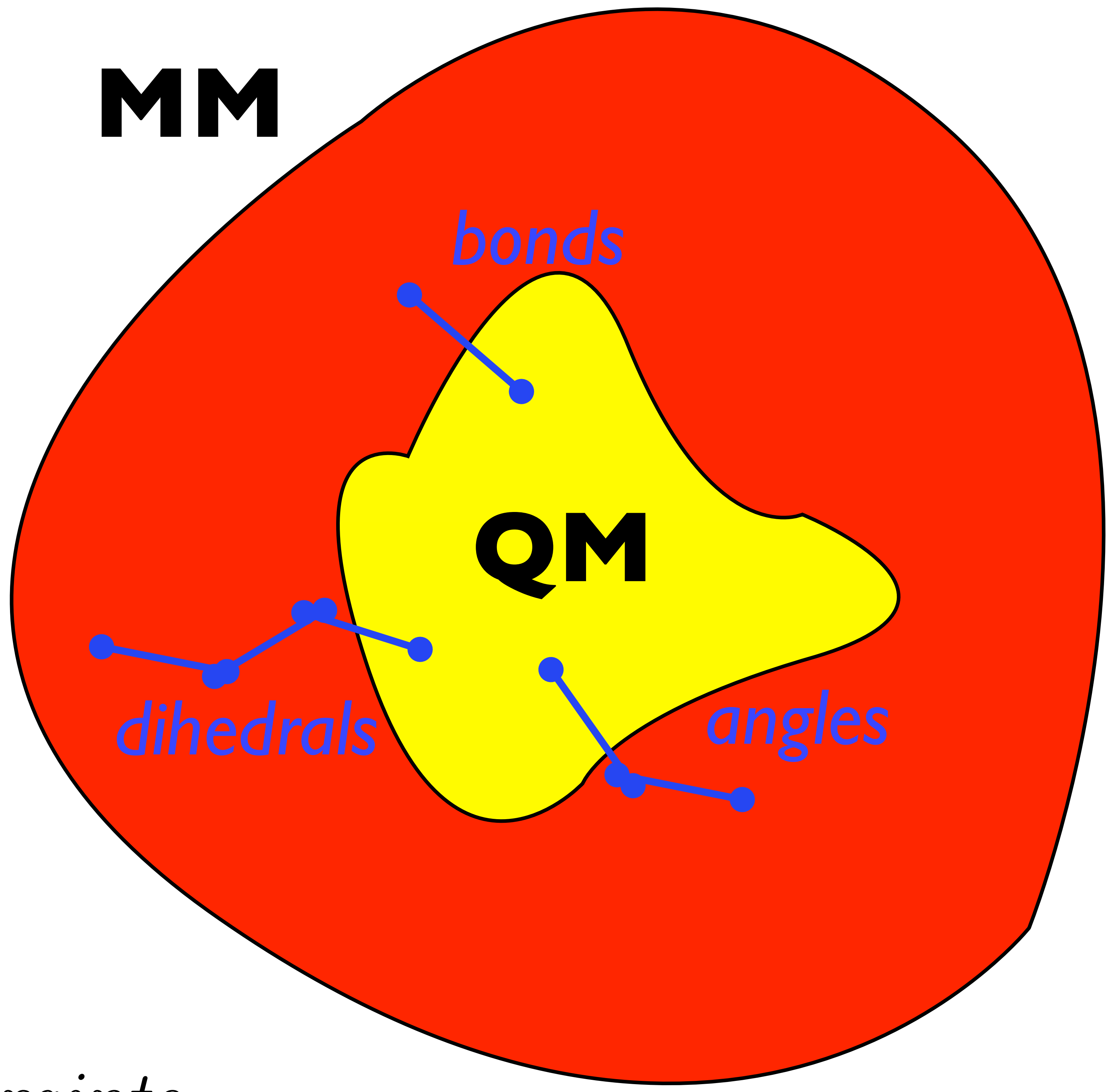


Figure 3.24b Physical Biology of the Cell (© Garland Science 2009)



Hybrid QM/MM molecular dynamics

$$H = H_{QM} + H_{MM} + \underbrace{H_{QM/MM}}_{\textit{coupling term}}$$



QM: First principles Density functional theory MD

$$\mathcal{L}_{\mathrm{CP}} = \underbrace{\sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2}\mu_i \left\langle \dot{\psi}_i \middle| \dot{\psi}_i \right\rangle}_{\text{kinetic energy}} - \underbrace{\langle \Psi_0 | \mathcal{H}_{\mathrm{e}} | \Psi_0 \rangle}_{\text{potential energy}} + \underbrace{\textit{constraints}}_{\text{orthonormality}}$$

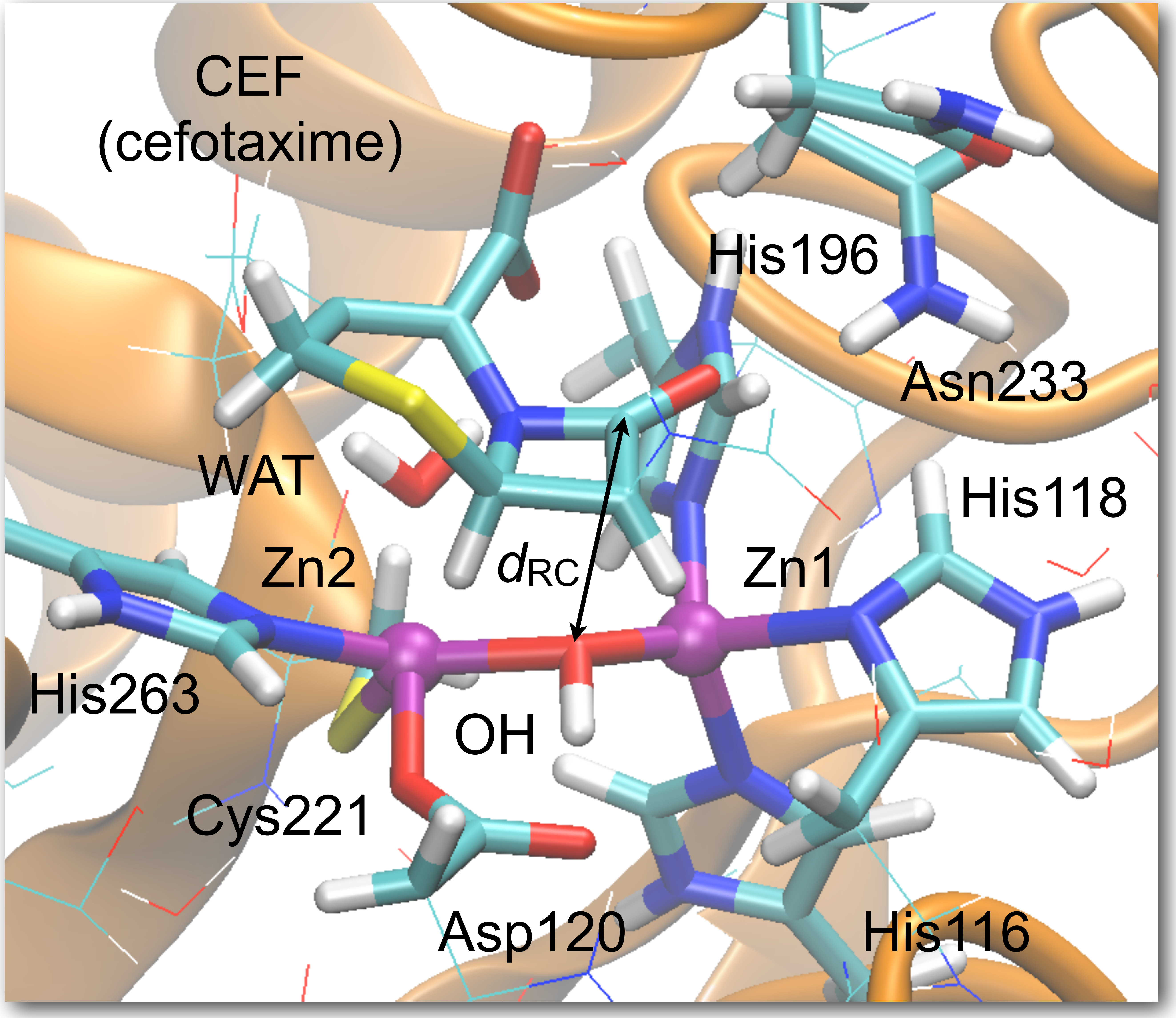
MM: Classical molecular dynamics (*e.g.* AMBER, Gromos force fields)

QM/MM: - boundary atom (*ad hoc* monovalent pseudopotential or H capping)

- hierarchical scheme to compute Coulomb interactions

Car, Parrinello, *PRL* 1985, Laio, Vandevondele, Rothlisberger *JCP* 2004, Dal Peraro *et al.*, *Curr. Opin. Struct. Biol.* 2007

CcrA MβL from *Bacteroides fragilis*



Thermodynamic integration along the reaction coordinate d_{RC}
DFT-BLYP, Martins-Troullier PPs, 70 Ry cutoff,
Nose' thermostat at 300 K,
2 reactions pathways for a total of ~150 ps trajectory

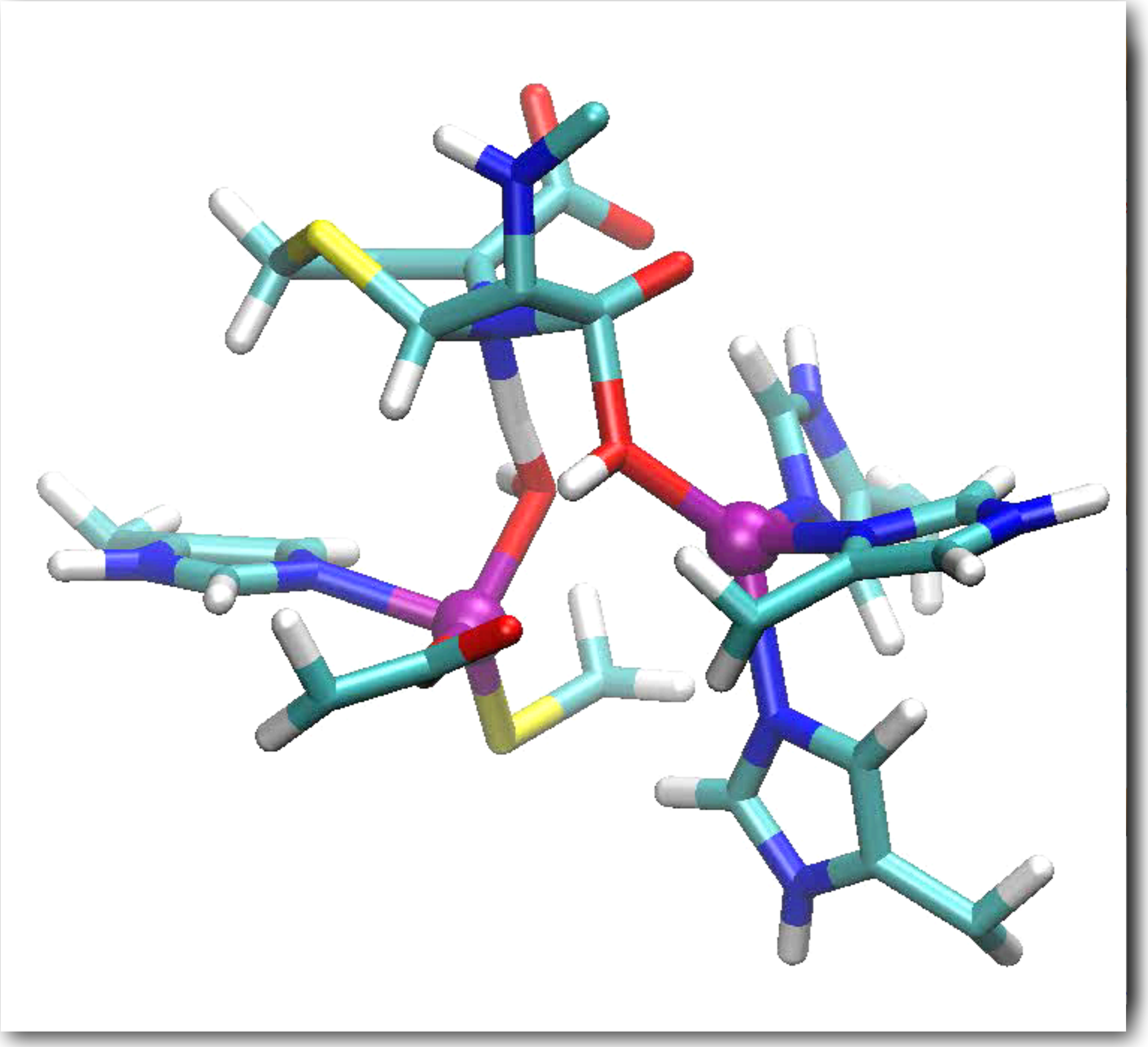
Reactant state

CcrA complexed with cefotaxime

- stable Michaelis complex OH-β-lactam distance=3.3(2)Å during 5 ns MD and 20ps QM/MM
- Zn2-bound WAT is the only water between the zinc center and CEF in 5Å

➡Classical force-field based MD is used as a tool to sample conformational space within the nanosecond timescale

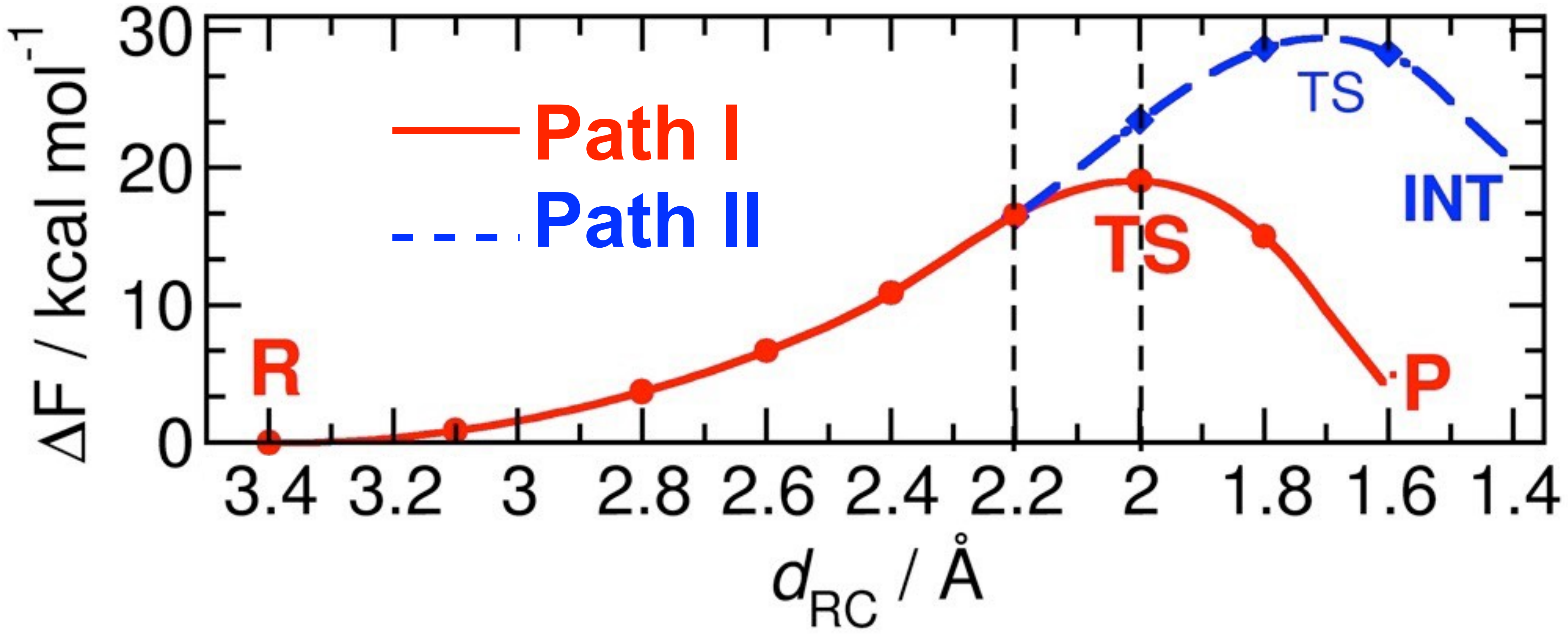
... from transition state to products



water-mediated single-step

- OH^- loses Zn2 coordination
- Zn1, Zn2 flexibility
- WAT protonates β-lactam N
- N-C β-lactam bond breaks
- WAT replaces OH^- as an hydroxide

- **ΔF = 18(2) kcal/mol** is in good agreement with experiments
- if Asn233 *does* H-bond β-lactam: formation of a high unfavorable intermediate (Path II)



Dal Peraro *et al.*, *JACS* 2007


chemical detail
sampling
binding properties
mesoscale
enzymatic reactions
coarse-grained
MM-CG
QM-MM
MD
QM
structural fluctuations
self-assembly
size/atoms
10^8
10^6
10^4
10^2
fs
ps
ns
μs
ms
s
timescale

Coarse-graining degrees of freedom

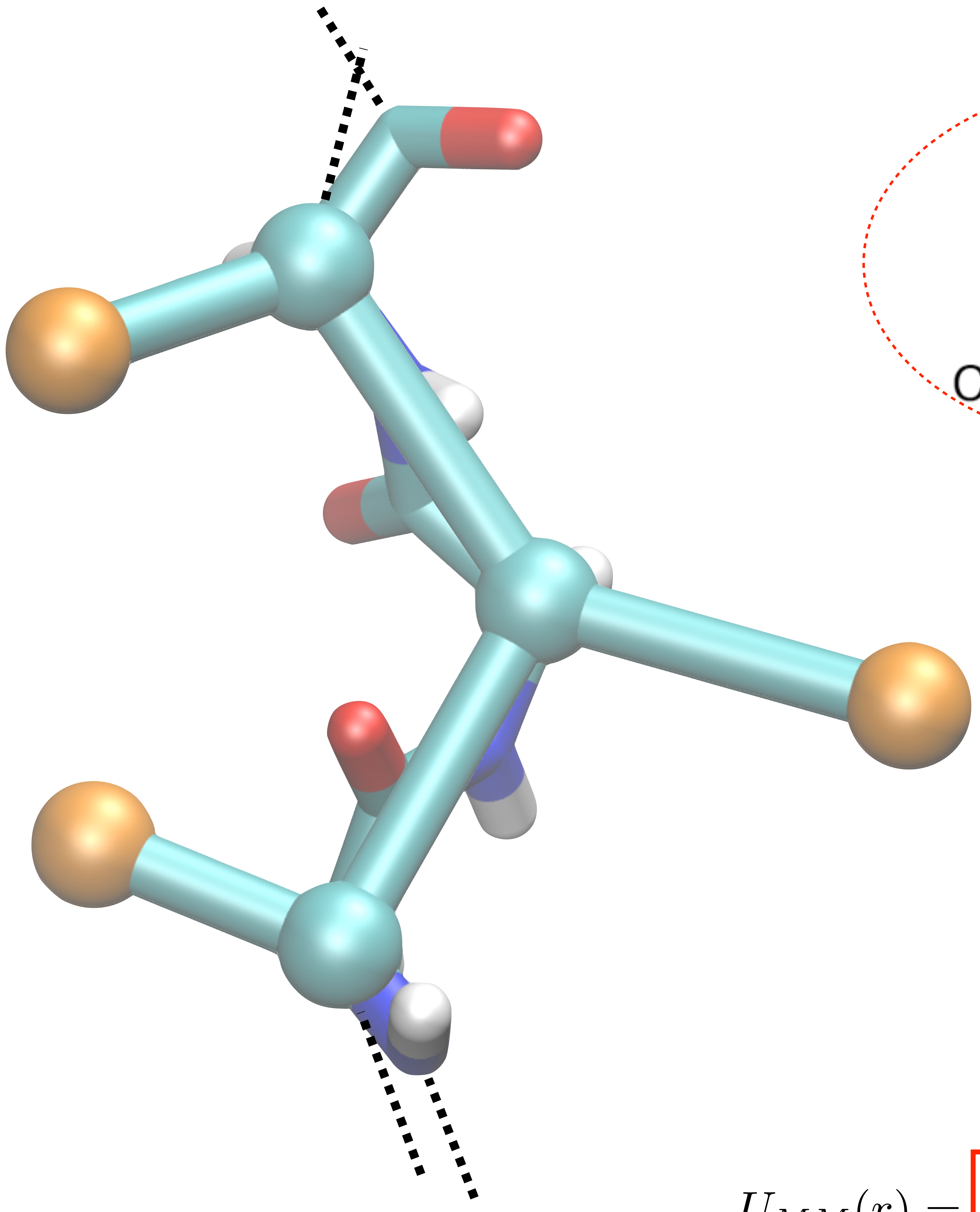
- **CG** is the process of consistently reduce the complexity of your problem integrating out degrees of freedom which can be in principle neglected for your system.

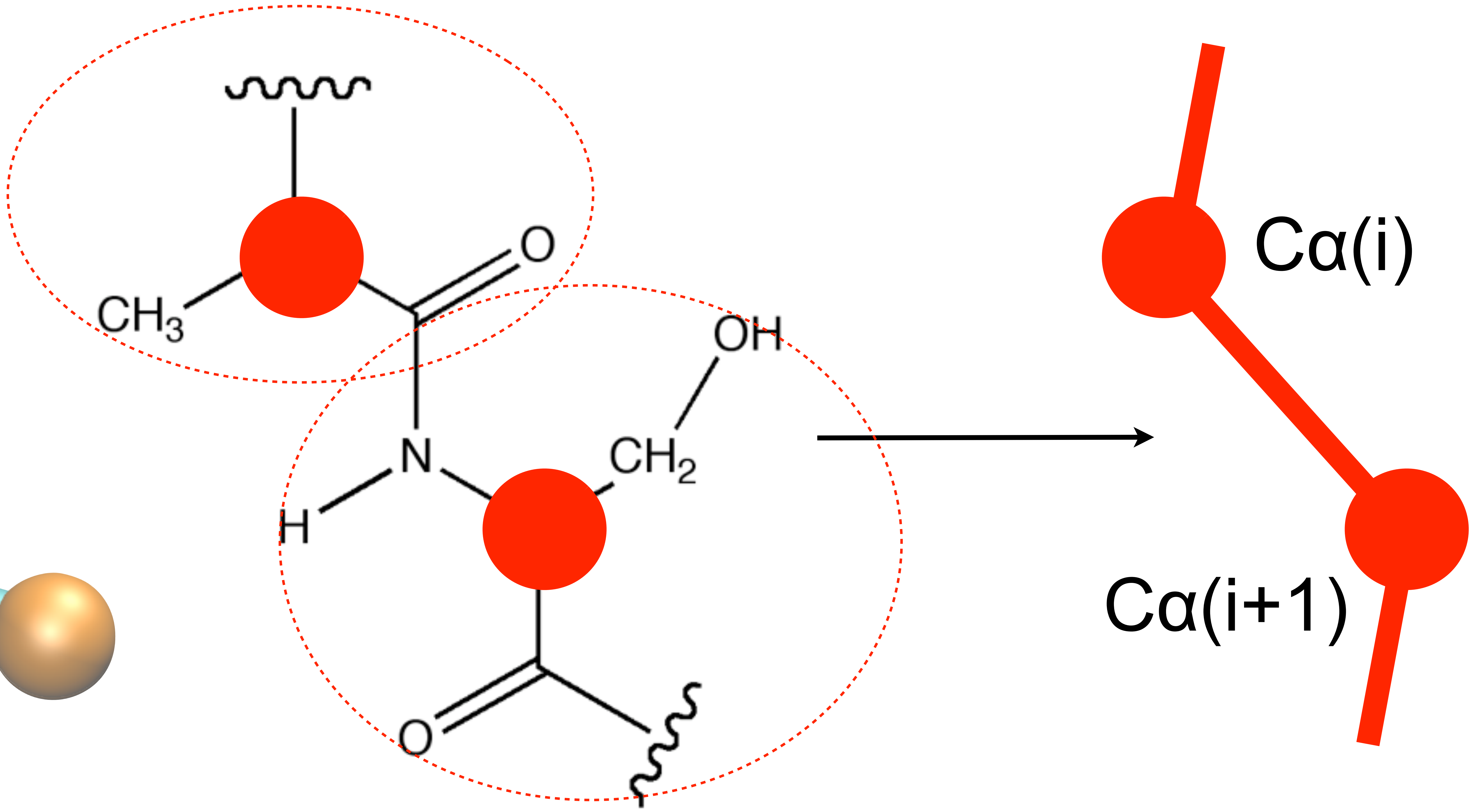
$$V_{QM} \rightarrow V_{MM} \rightarrow V_{CG-MM} \rightarrow V_{mesoscopic}$$

- the CG process implies a **simplification** of your potential that is not always rigorous and includes **approximations**

- what you obtain is an **effective** potentials which is parametrized to reproduce given properties

Coarse-graining degrees of freedom



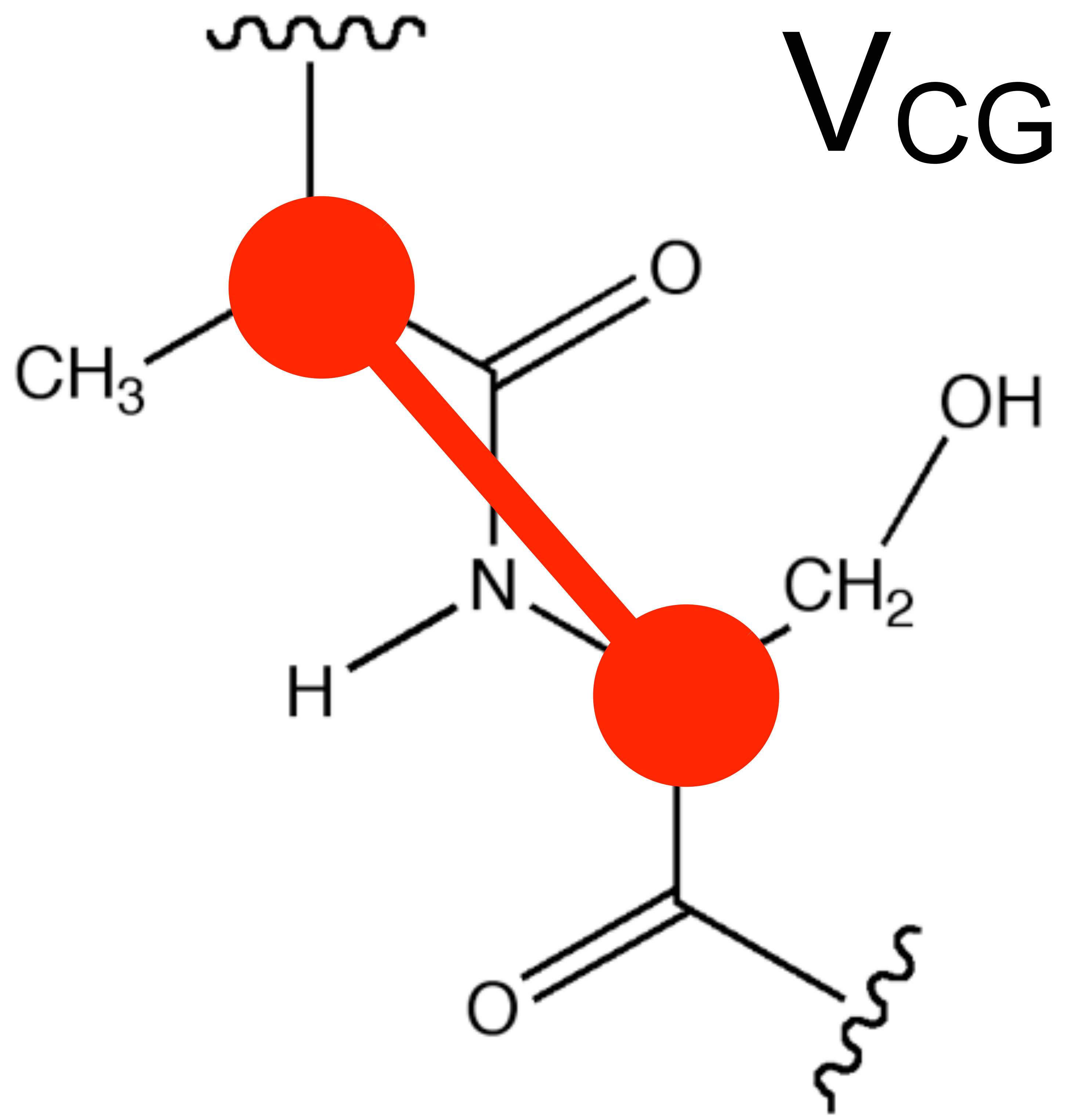


U_{MM}

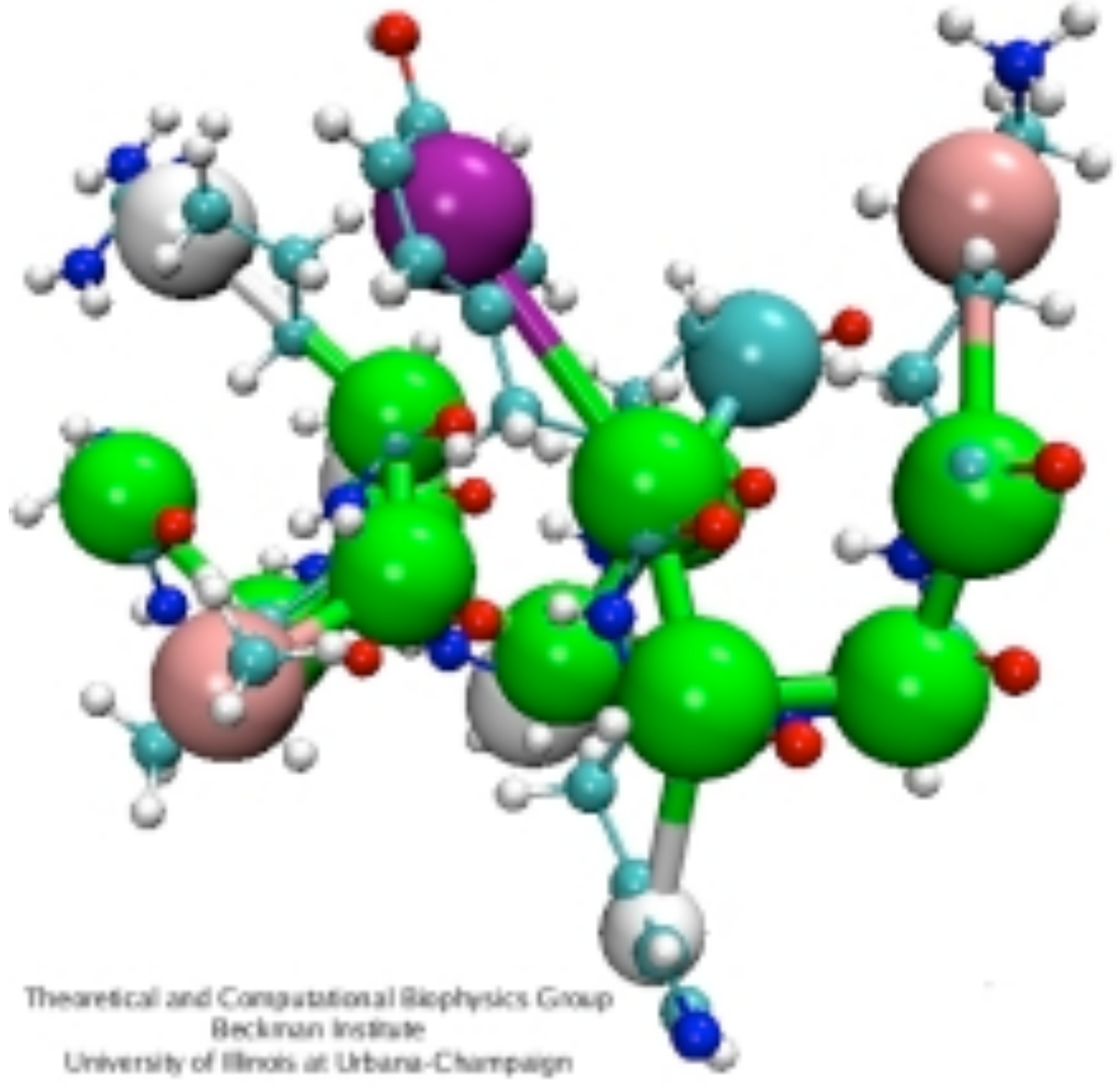
U_{CG}

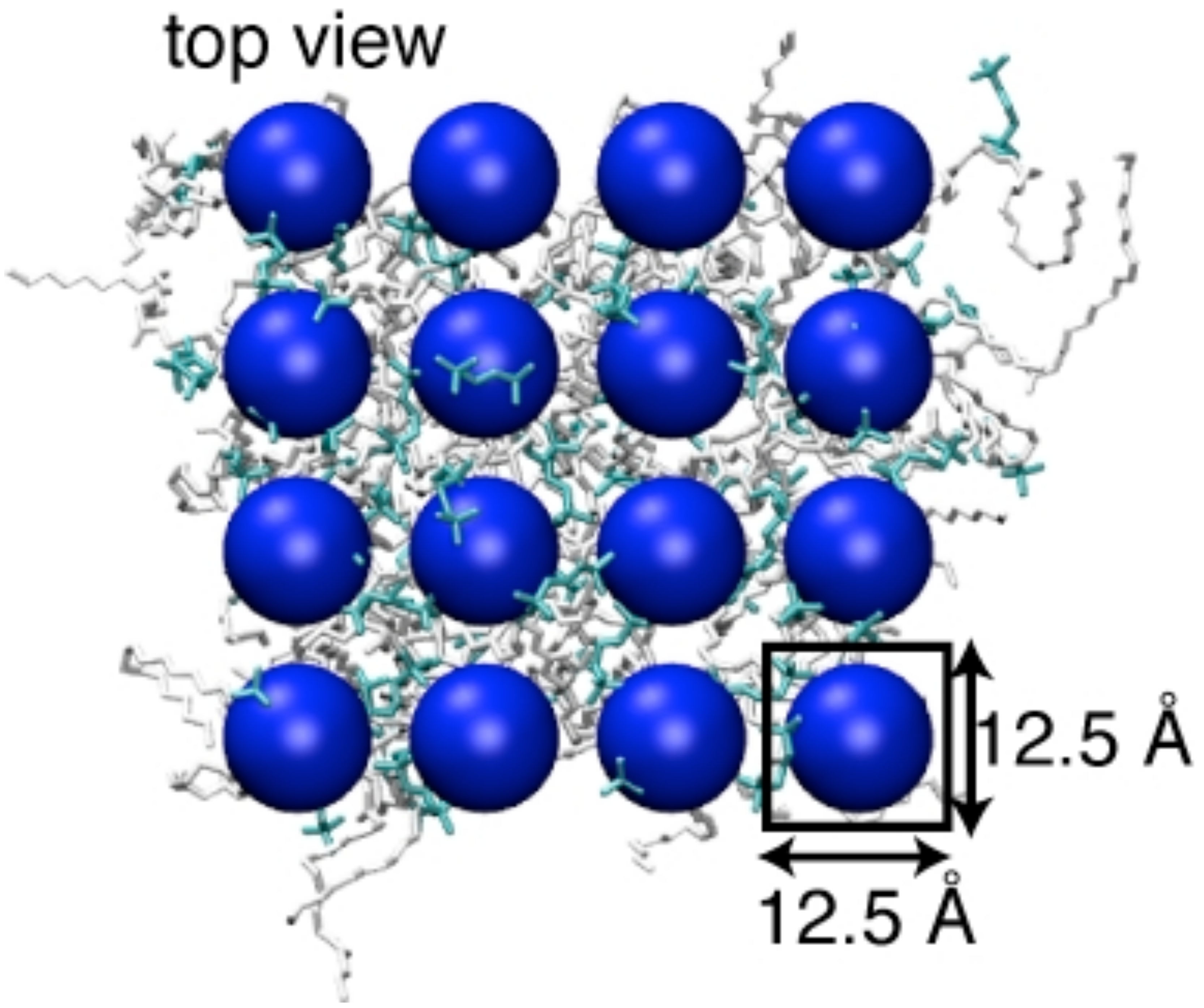
$$U_{MM}(r) = \boxed{\sum_{bonds} \frac{k_b}{2}(r-r_0)^2 + \sum_{angles} \frac{k_\theta}{2}(\theta-\theta_0)^2 + \sum_{torsions,n} \frac{k_{\phi,n}}{2}[1+cos(n\phi-\delta)]} + \\ + \boxed{\sum_{i>j}^{N} \left(\frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{6}}\right) + \sum_{i>j}^{N} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}}$$

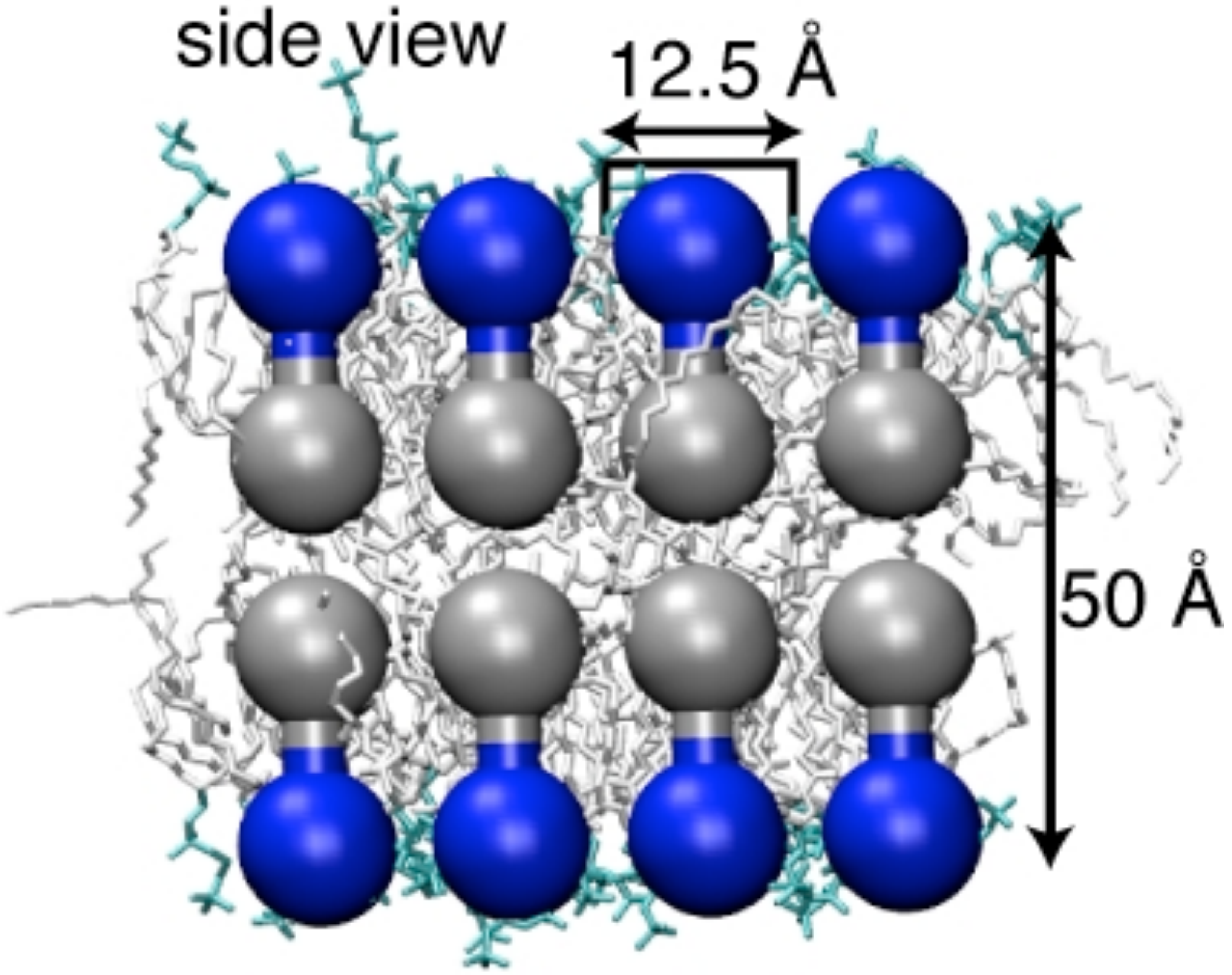
Coarse-grained force fields



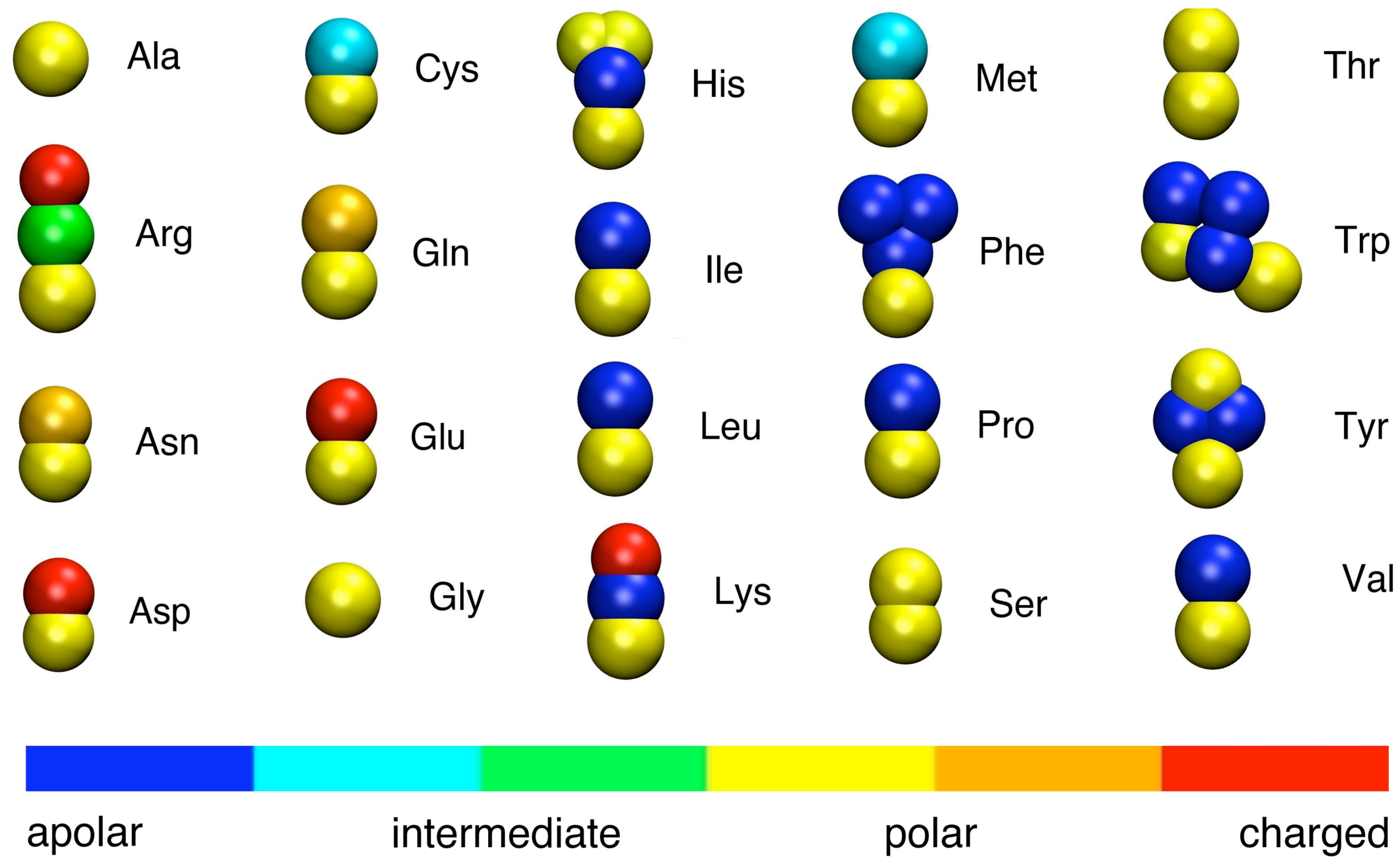
- CG FF models are not topologically biased on the native structure
- softer interactions allow for **longer** timestep in MD simulations
- sampling on the **millisecond** timescale
- accuracy can be a problem (e.g. **no explicit electrostatic** contribution)
- biases on the secondary structures





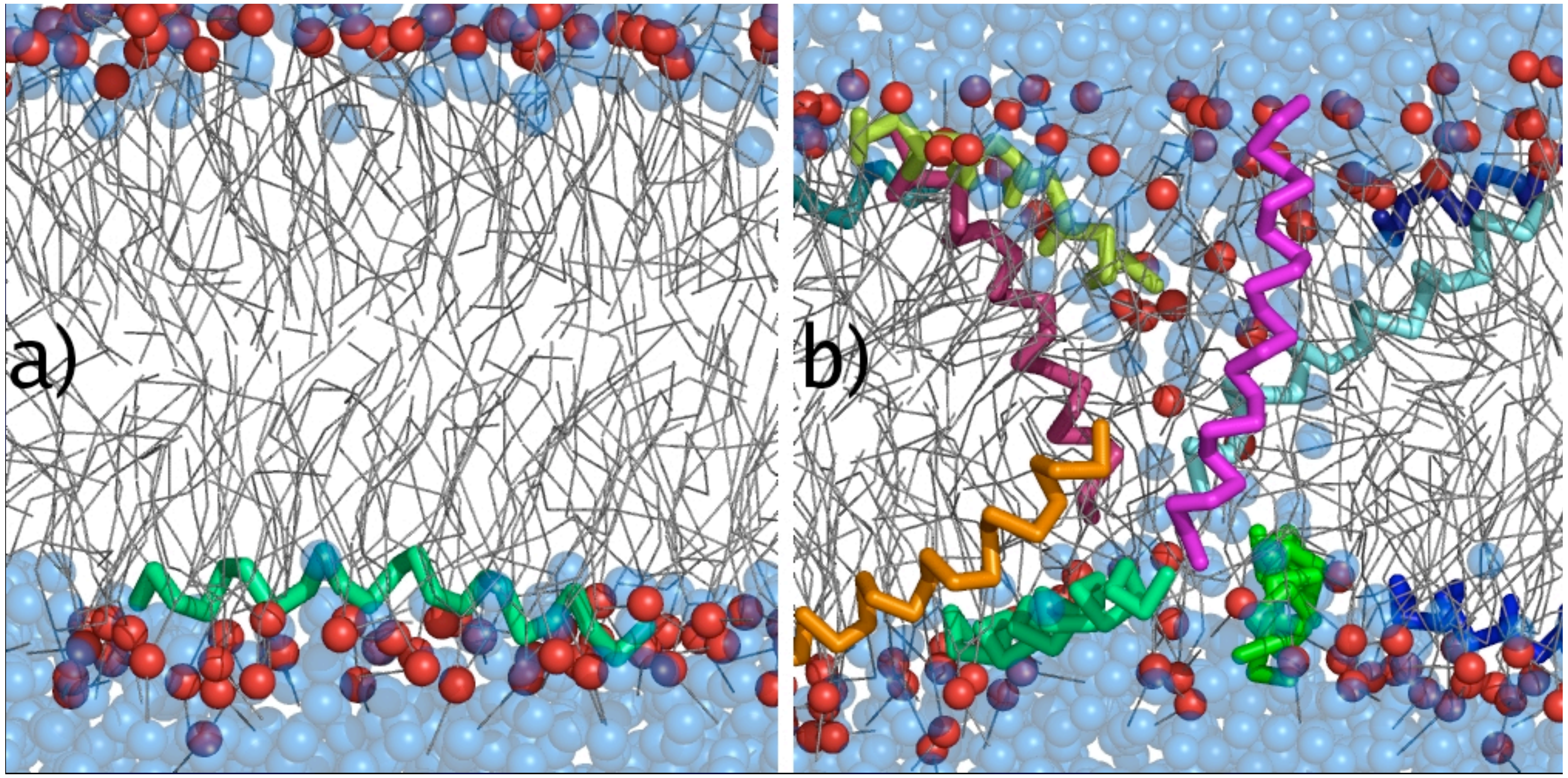


Coarse-grained MARTINI FF

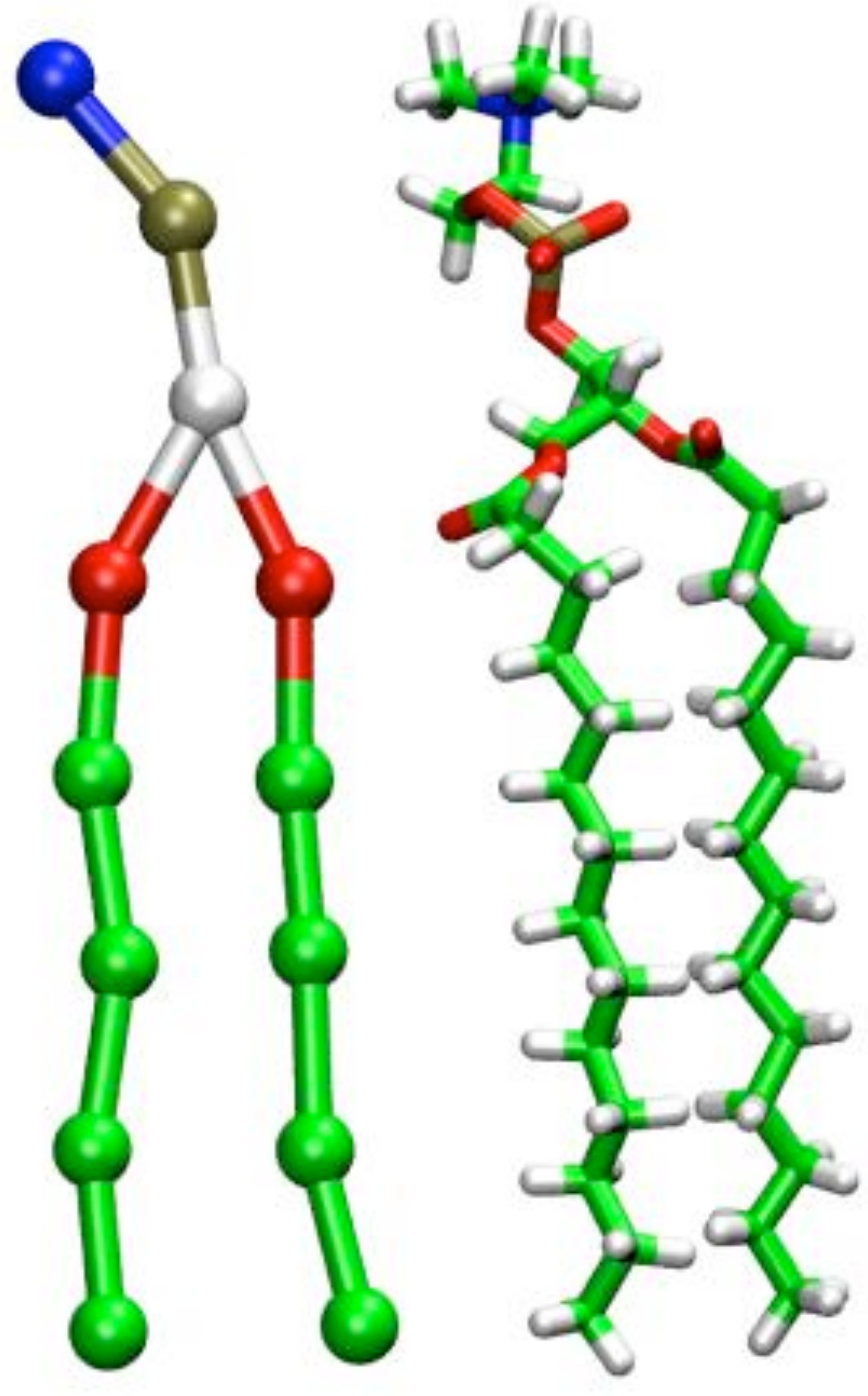


- MARTINI CG FF has functional form similar to MM FF
- 4-to-1 mapping from MM to CG
- very convenient for membranes and peptide-membrane interactions

Monticelli et al, JCTC 2008
Klein and coworkers



Magainin H2 in a DPPC bilayer, at low concentration (a) and high concentration



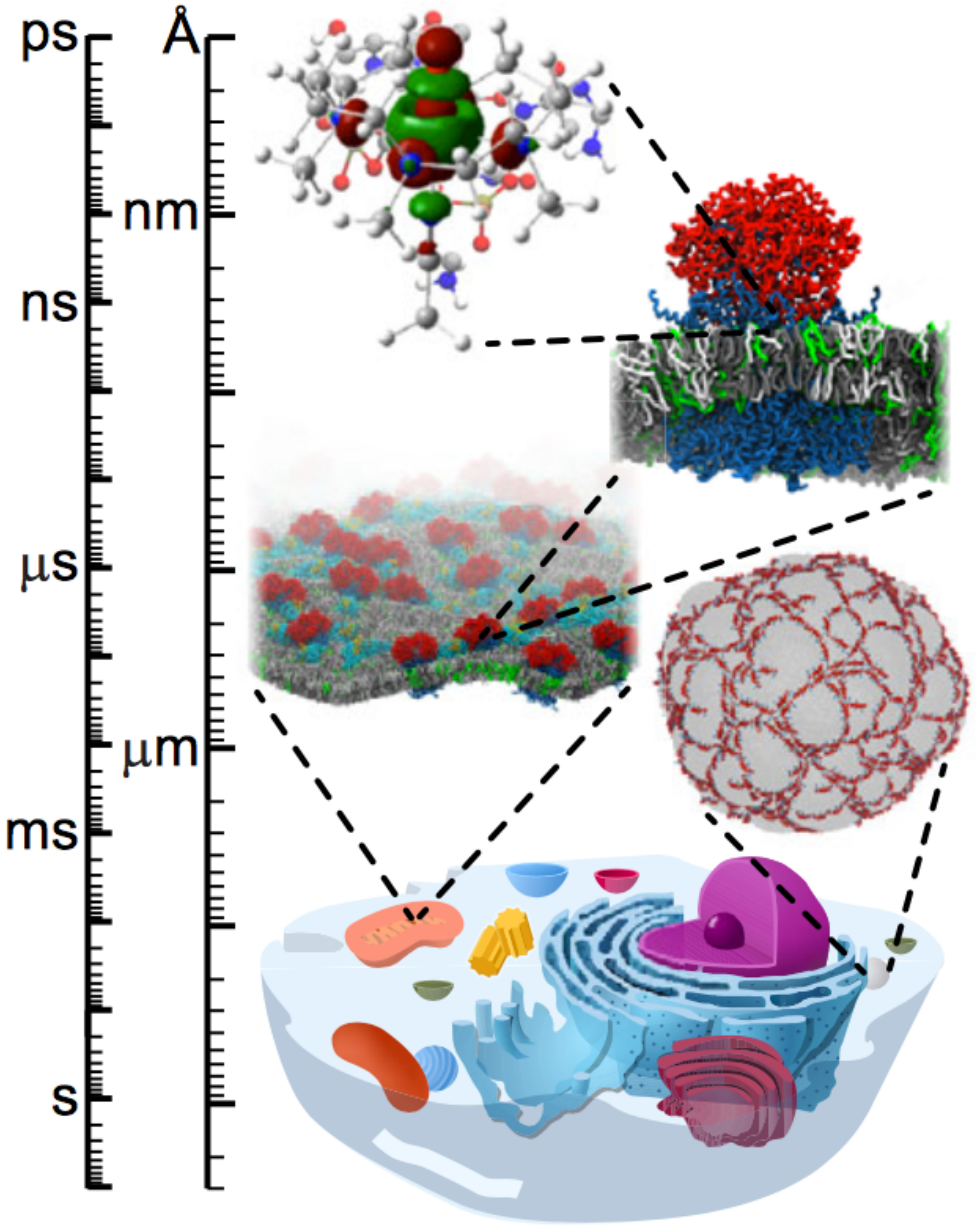
© 2016. Published by The Company of Biologists Ltd | Journal of Cell Science (2016) 129, 257-268 doi:10.1242/jcs.176040

COMMENTARY

ARTICLE SERIES: IMAGING

Computational ‘microscopy’ of cellular membranes

Helgi I. Ingólfsson, Clément Arnarez, Xavier Periole and Siewert J. Marrink*



Quantum
- atoms, electrons and electron clouds included
- explicit solvent
- quantum mechanics

All-atom
- all or most atoms present
- explicit solvent
- molecular dynamics

Coarse-grained
- beads comprising a few atoms
- explicit or implicit solvent
- molecular dynamics

Supra-coarse-grained
- interaction sites comprising many atoms, protein parts or proteins
- implicit solvent
- stochastic dynamics

Continuum
- materials as a continuous mass
- implicit solvent
- continuum mechanics