Przegląd metod MD i innych wieloskalowych metod modelowania

Bogdan Lesyng

Zakład Biofizyki oraz Centrum Doskonałości BioExploratorium

Wydział Fizyki, Uniwersytet Warszawski

wykład 3.12.2009

t

fs ps ns μs ms s

Rozmiar

nm μm mm m

•fizyka
•chemia
•biologia molekularna

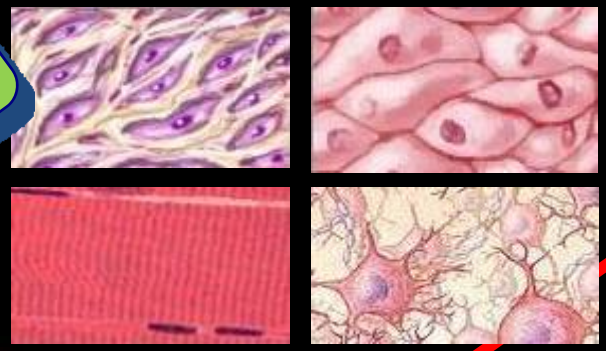
•bionformatyka i biologia systemów
(fizyka systemów)

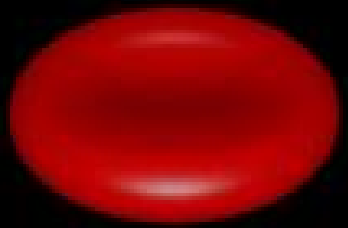
• informatyki stosowane
• modele matematyczne szlaków sygnałowych i metabolicznych

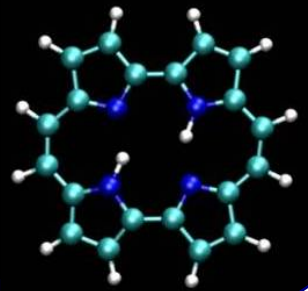
• MM, MD
• PB, GB
• oddziaływania hydrofobowe
• metody VR

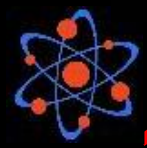
• QM, DFT, AVB
• QD, MD, QCMD

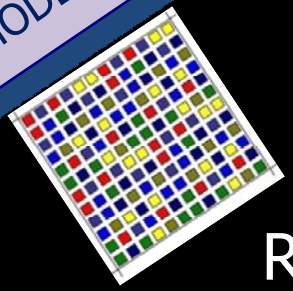
• METODY WIELOSKALOWEGO MODELOWANIA I BIOINFORMATYKI







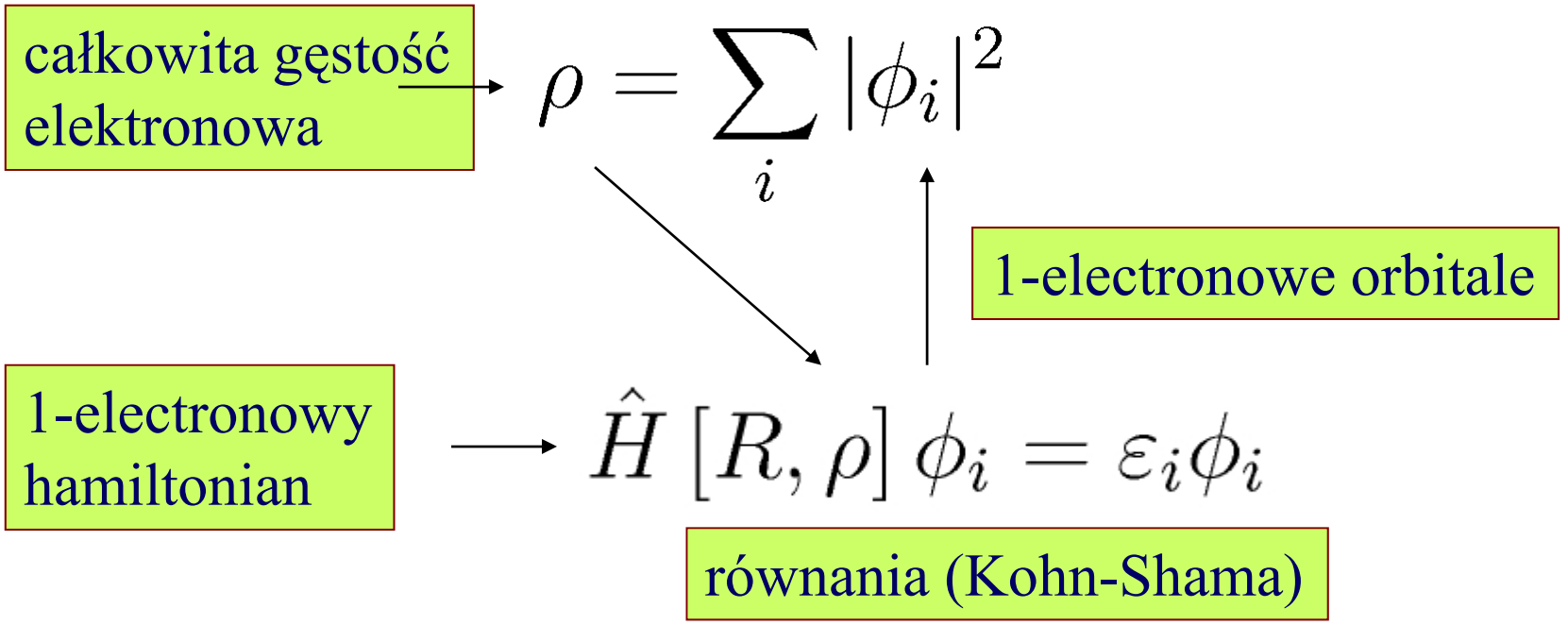




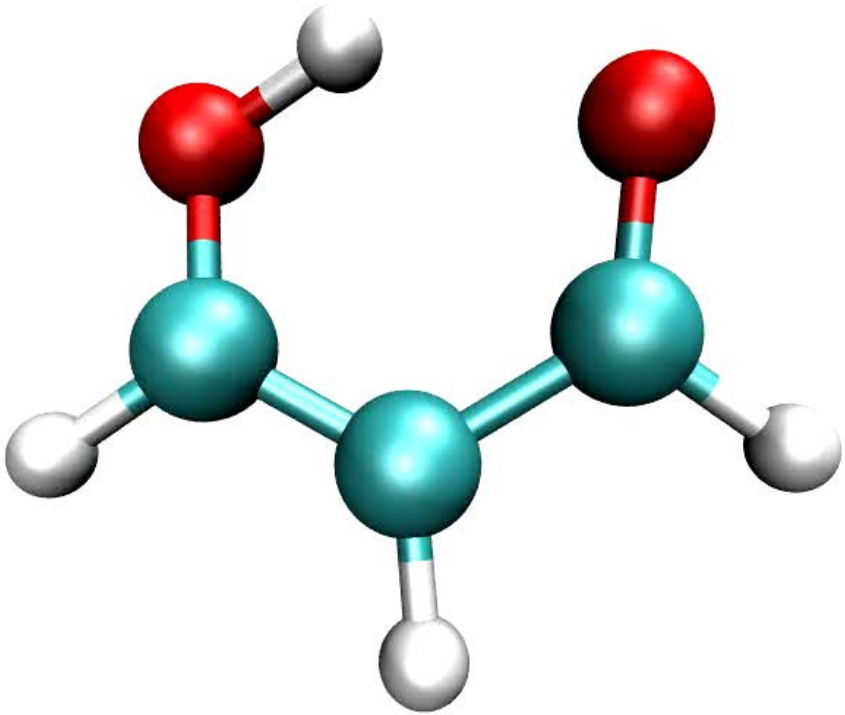
Metoda SCC-DFTB

(Self Consistent Charge Density Functional Based Tight Binding method, SCC DFTB, Frauenheim et al. Phys Stat. Sol. **217**, 41, 2000)

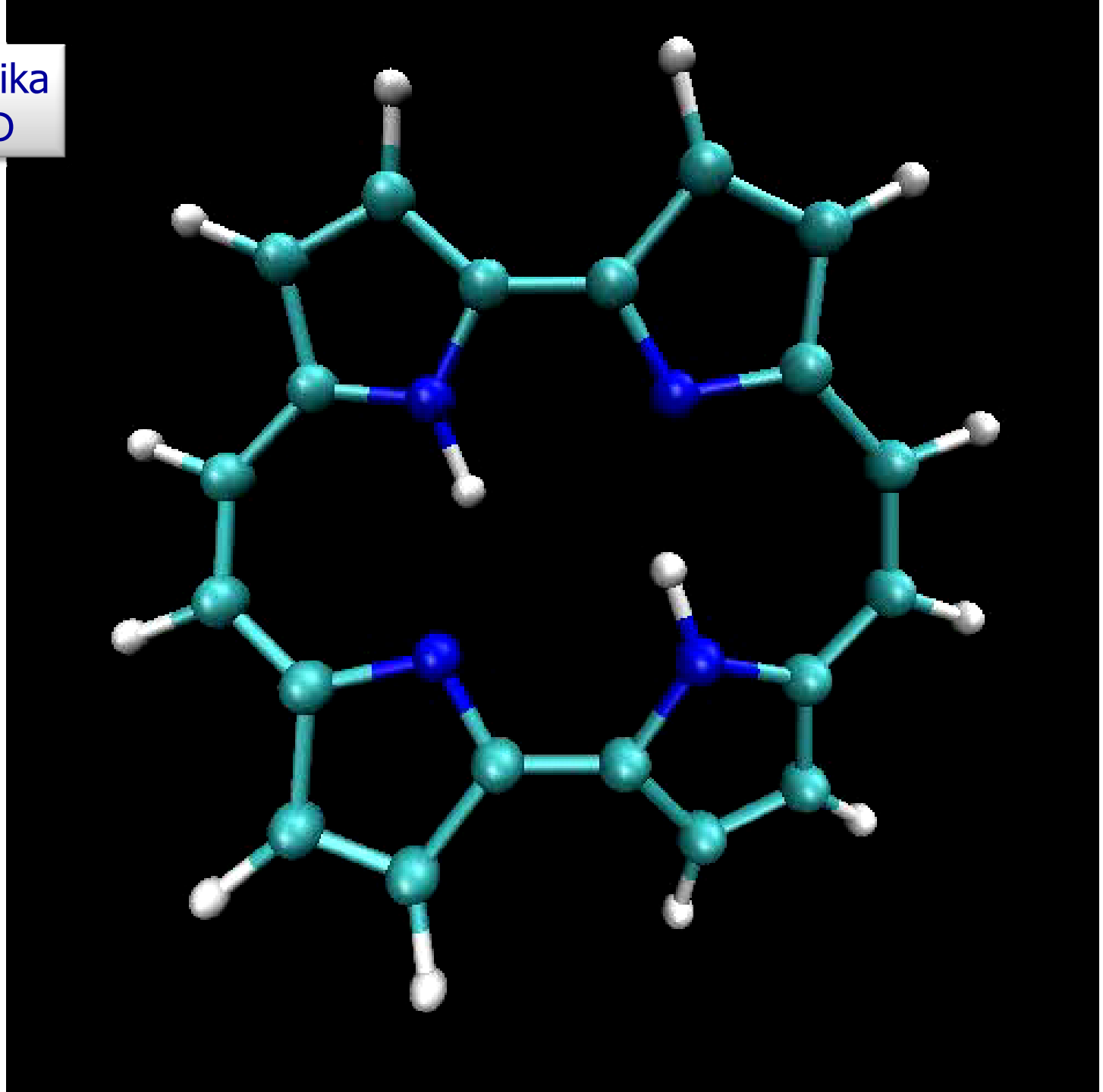
podstawy DFT :



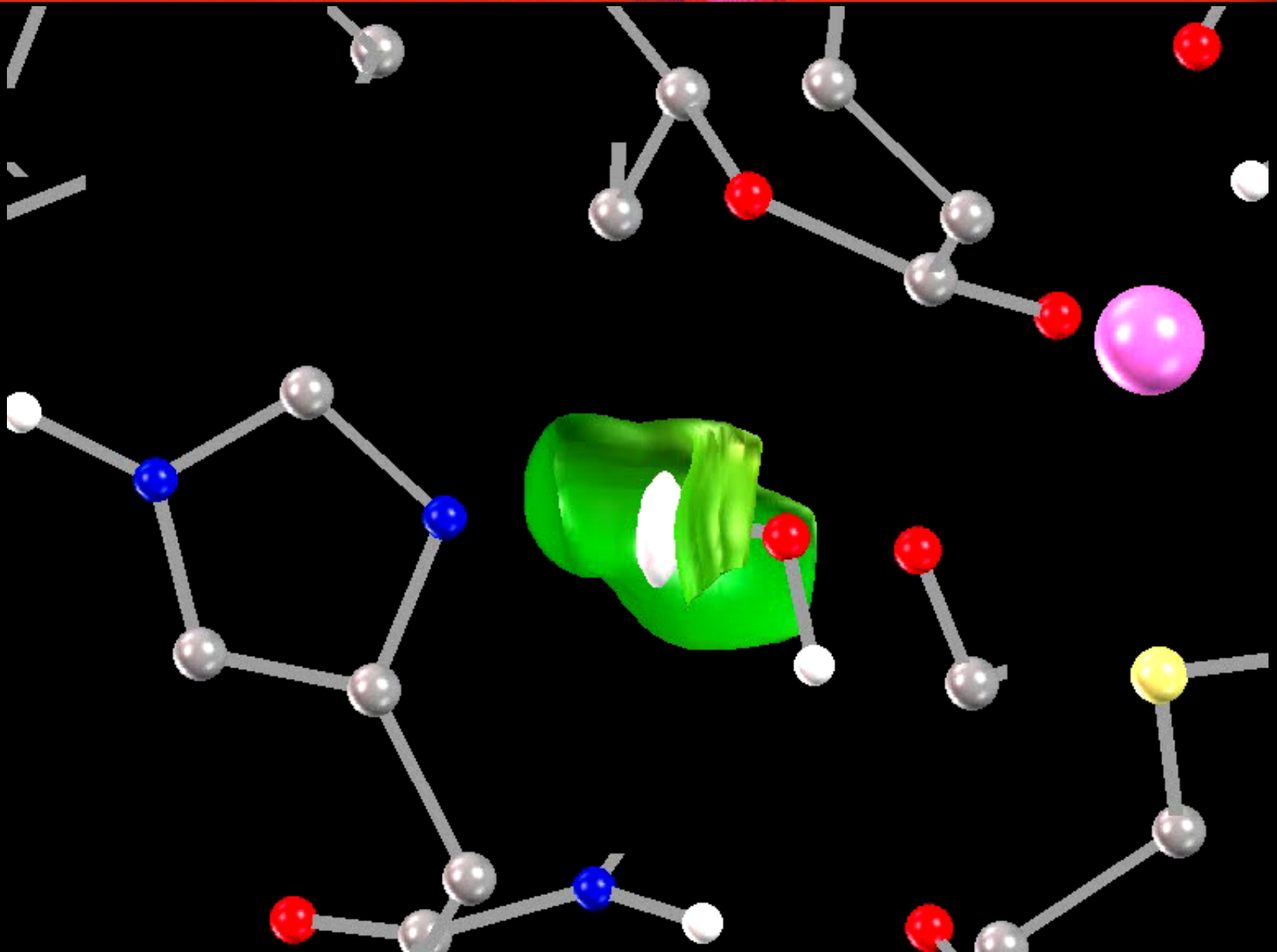
Malonoaldehyd
SCC-DFTB/MD



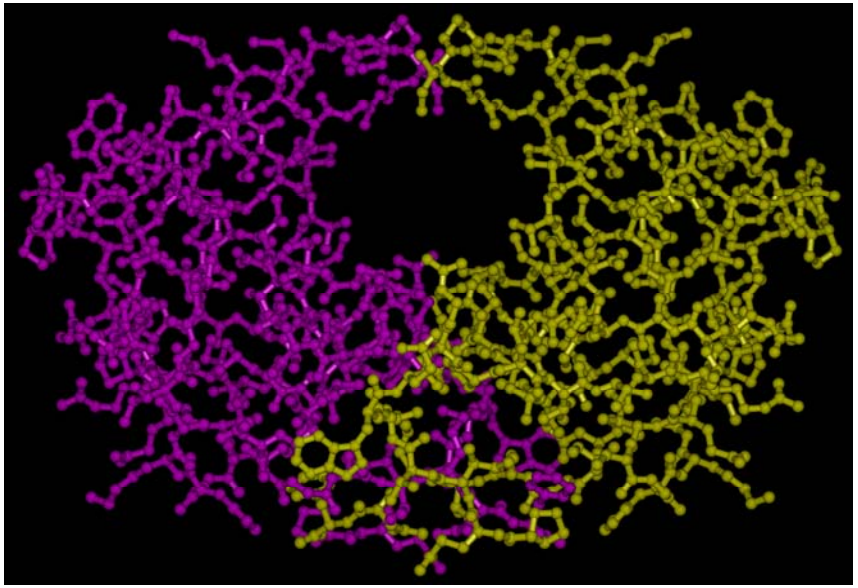
Dynamika
CPMD

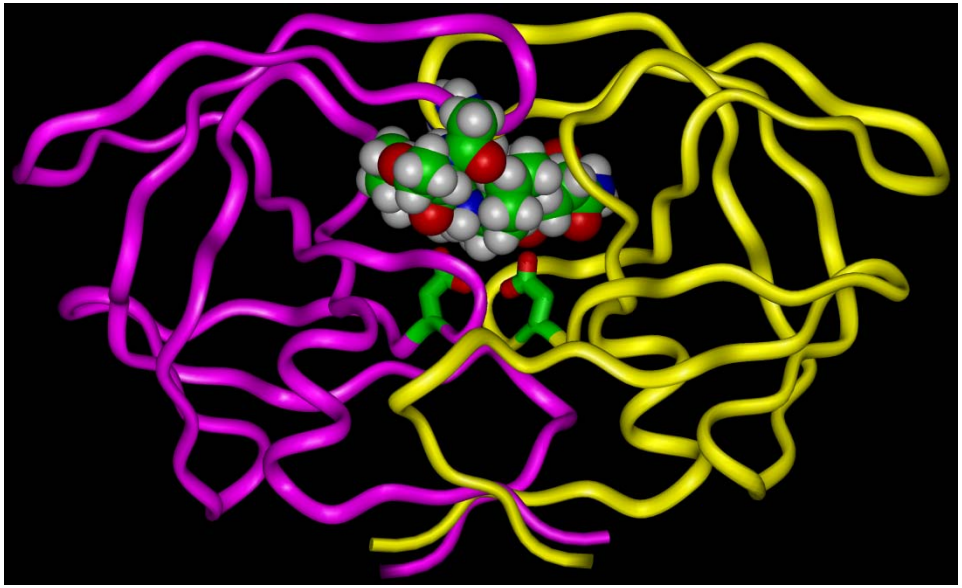


QCMD

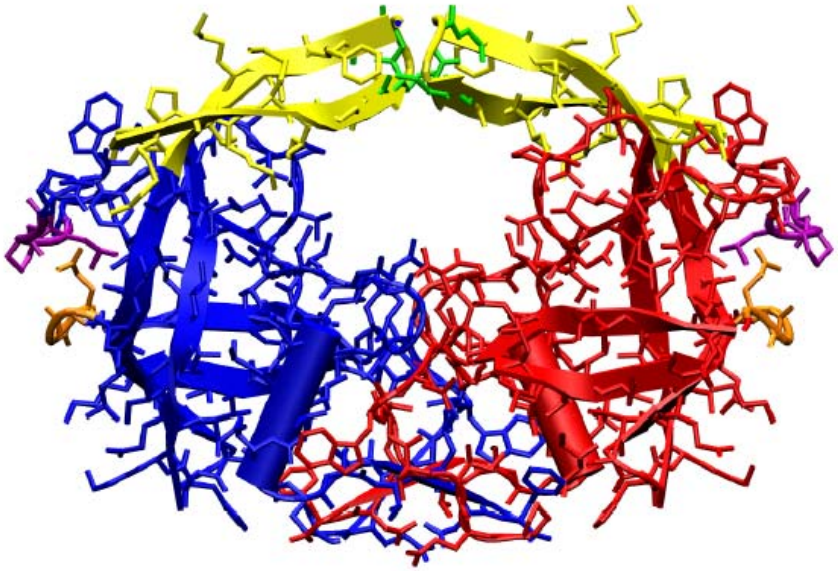


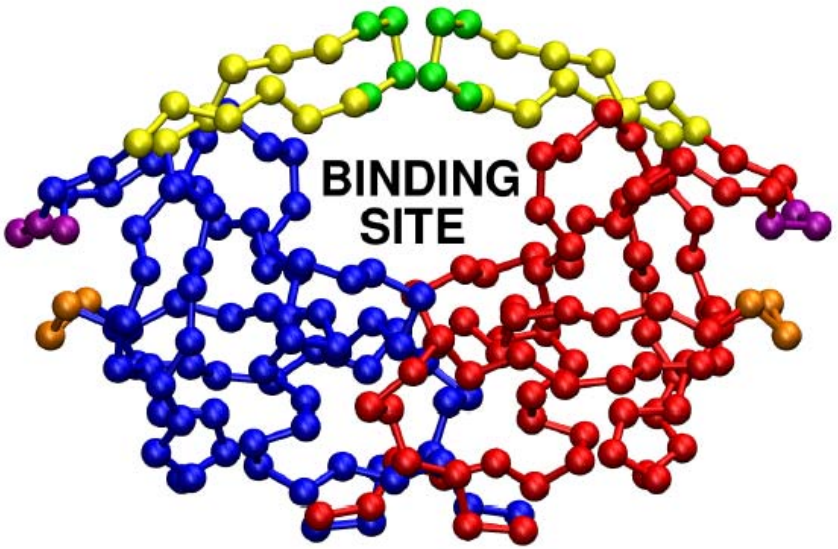
HIV-1 protease.
Flap opening dynamics

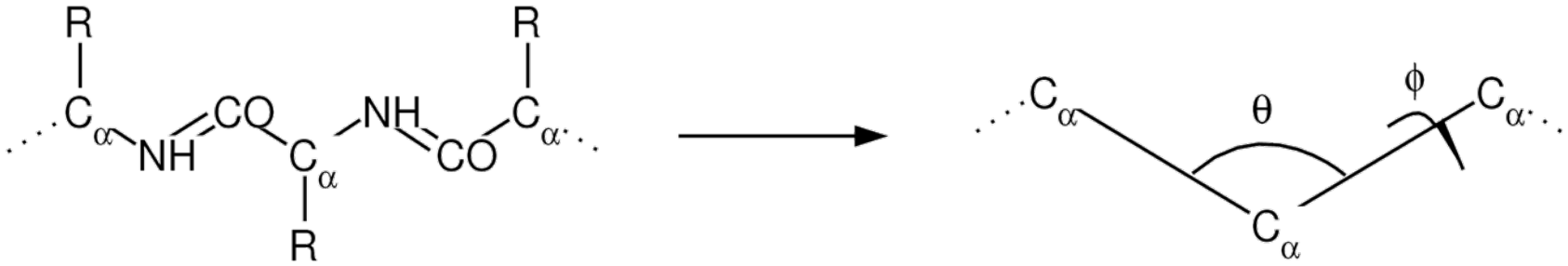




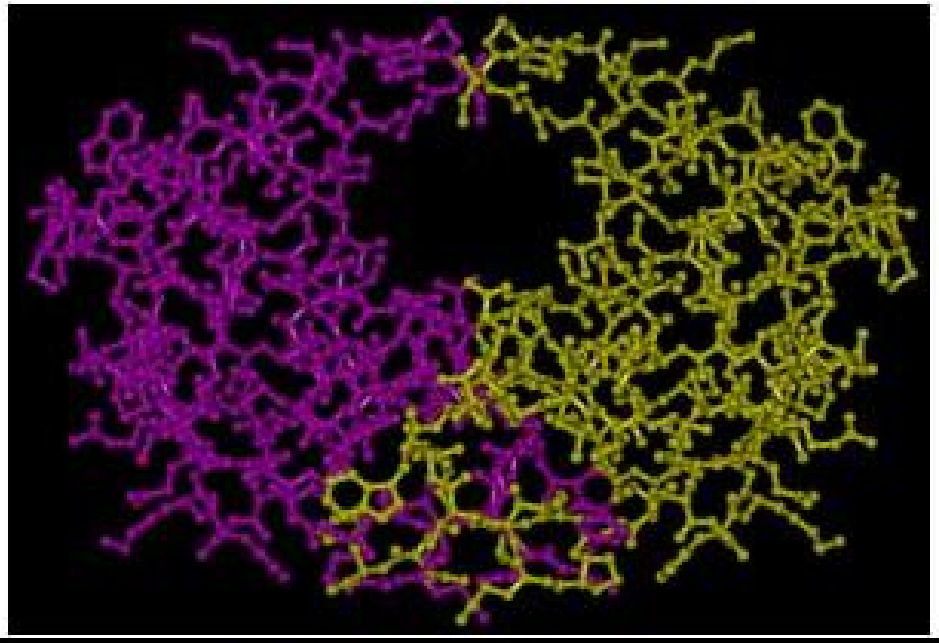
Data for the analysis taken taken from microsecond time-scale dynamics:
V.Tozzini, J.Trylska, C.Chang, J. A.McCammon, J.Struct.Biol., 157,606-615, 2006

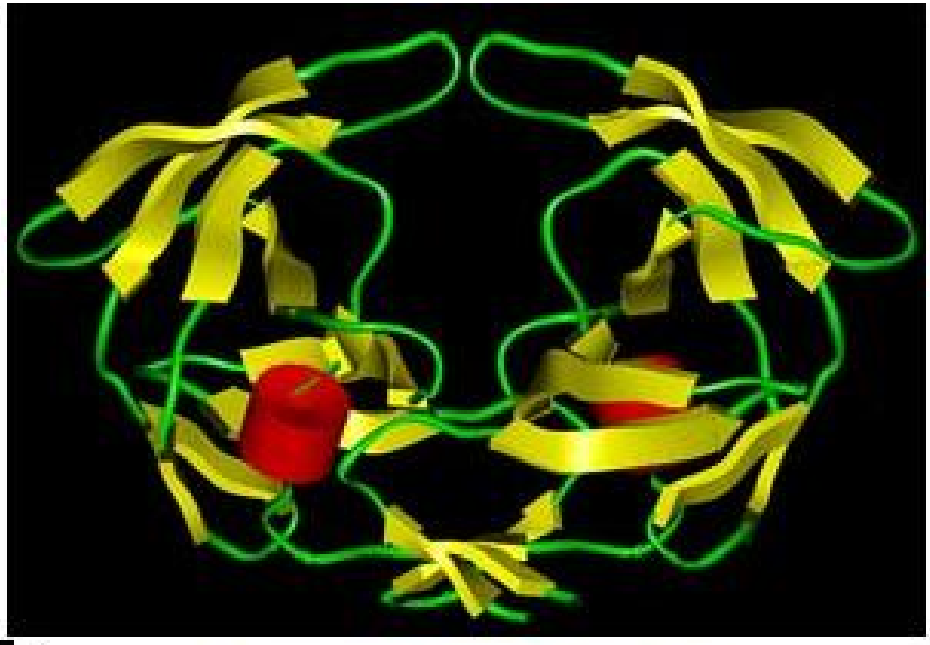


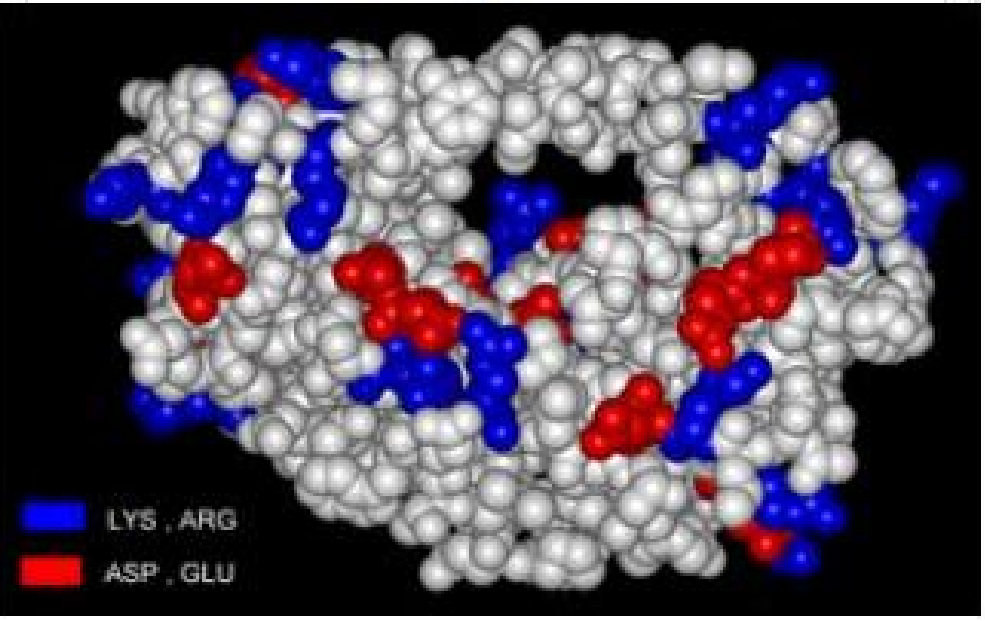


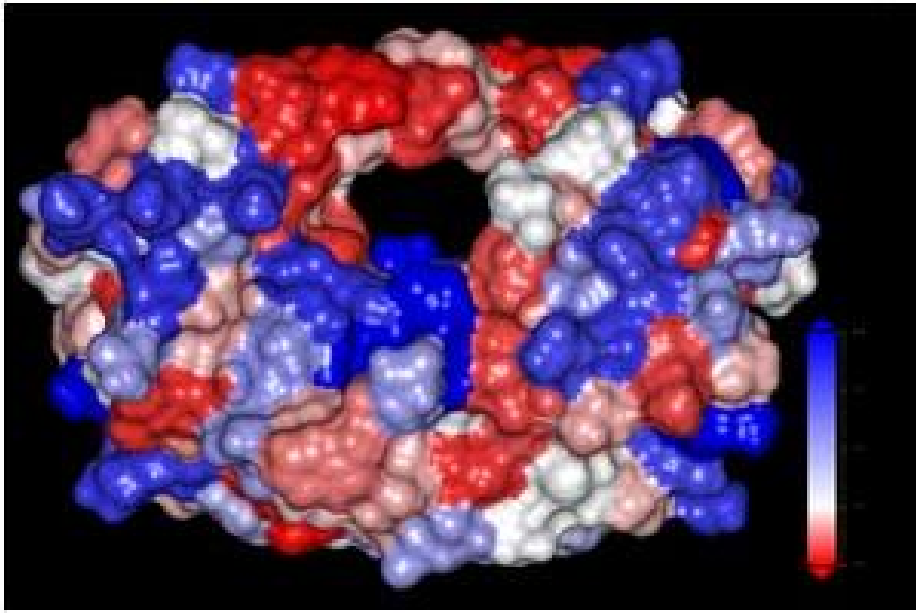


Simulation method:
DL_POLY, package, W.Smith and T.R.Forester, J. Mol. Graph.,14, 136-141, 1996

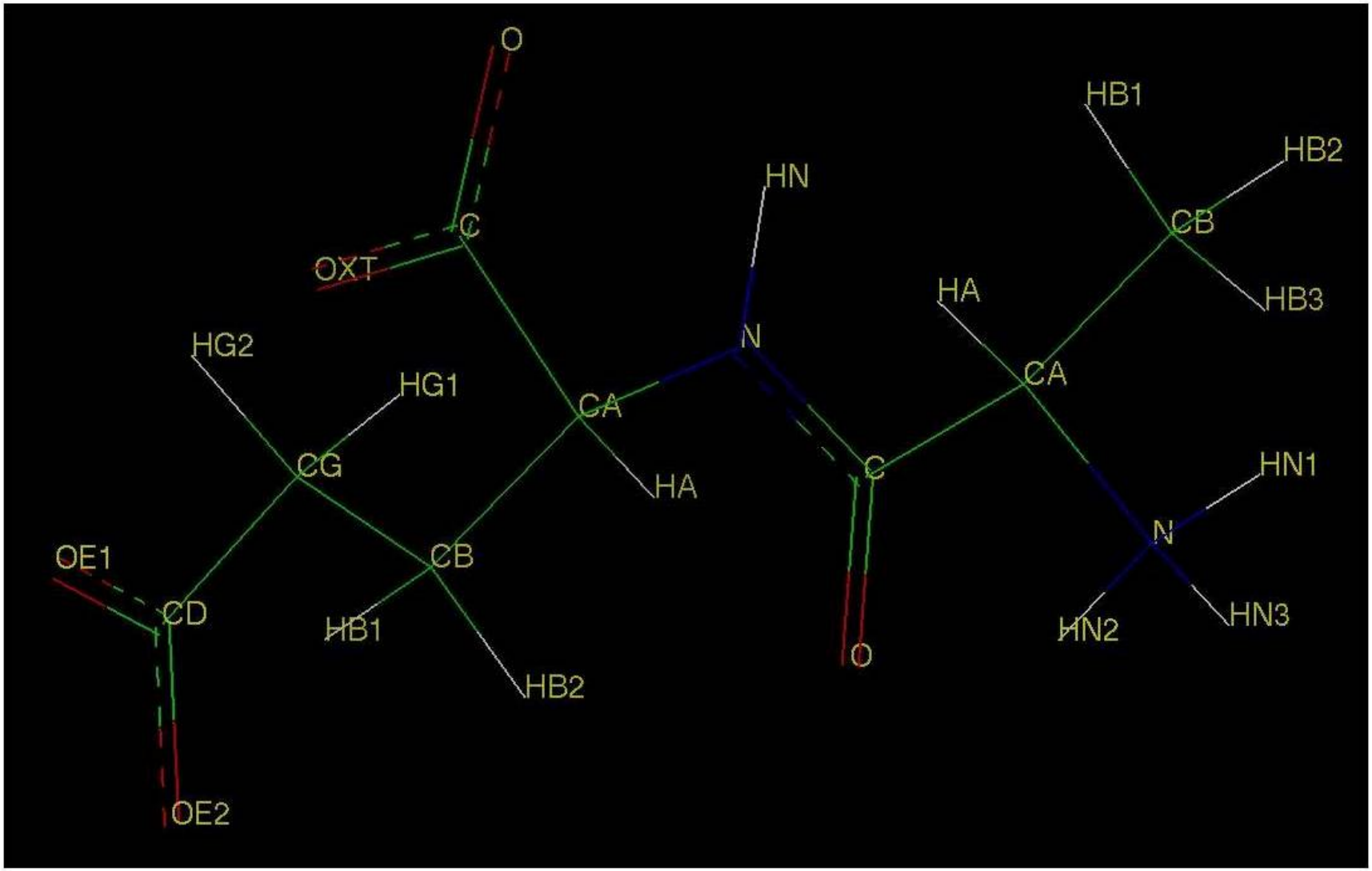




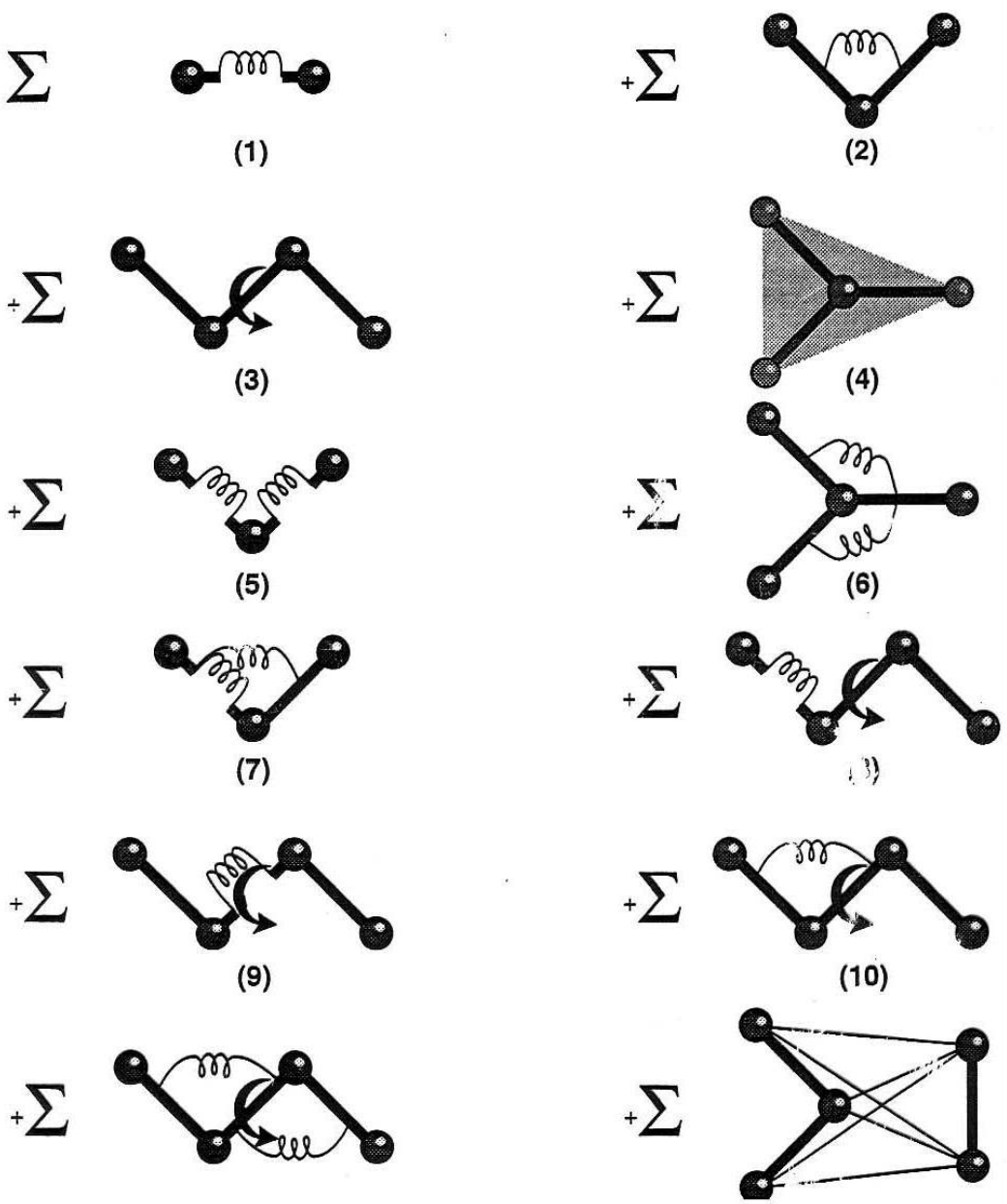

LYS , ARG
ASP , GLU



- Molecular dynamics
- Monte Carlo methods



Glutamate-Alanine


Σ
(1)
+Σ
(2)
+Σ
(3)
+Σ
(4)
+Σ
(5)
+Σ
(6)
+Σ
(7)
+Σ
(9)
+Σ
(10)
+Σ

Potential energy function

$$E(\mathbf{R}) = E_{bonded} + E_{non-bonded}$$

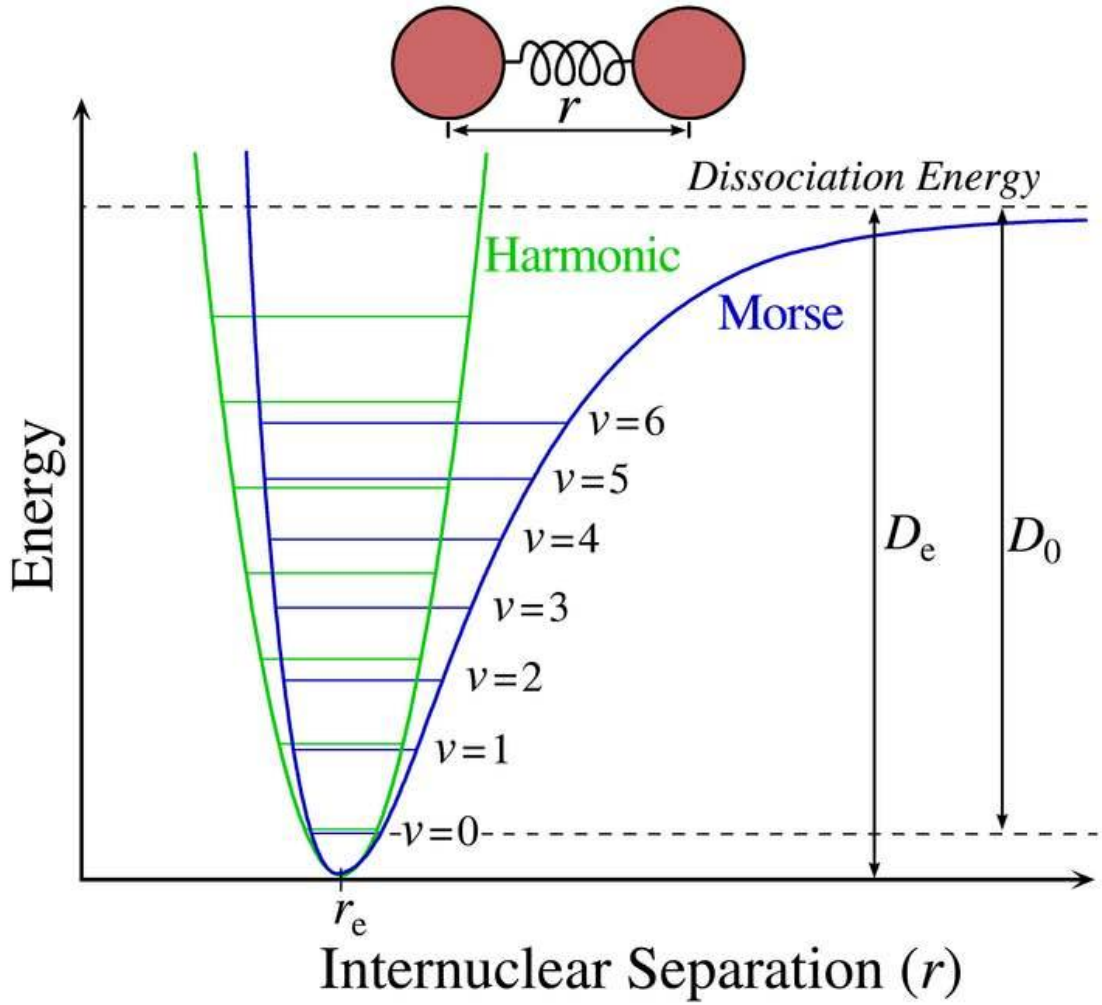
$$E_{bond-stretch} = \sum_{pairs} \frac{1}{2} k_{bond}(r - r_0)^2 \quad (1)$$

$$E_{angle} = \sum_{angles} \frac{1}{2} k_{ang}(\theta - \theta_0)^2 \quad (2)$$

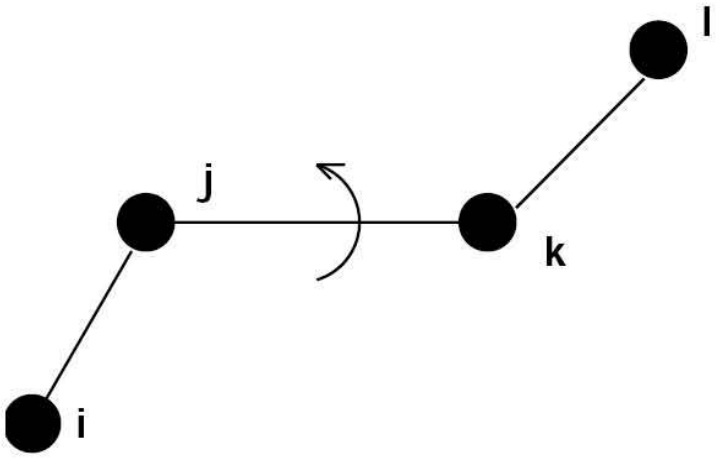
$$E_{dihedral} = \sum_{dihedrals} \sum_{n=1}^{3} \frac{1}{2} k_{dih}(1 + cos(n\phi - \phi_0)) \quad (3)$$

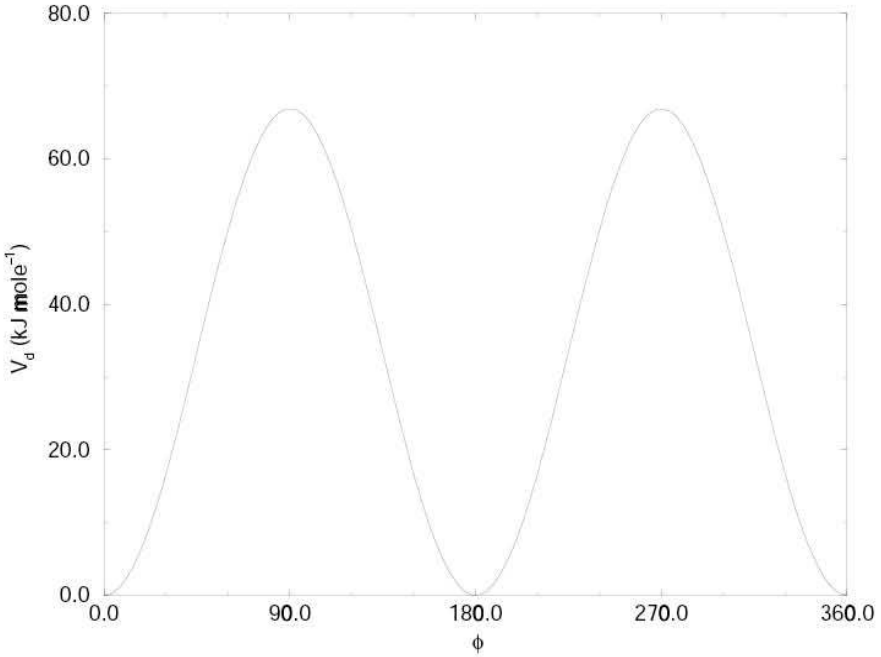
$$E_{coulomb} = \sum_{i>j} \frac{q_i q_j}{\epsilon r_{ij}}$$

$$E_{vdw} = \sum_{i>j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

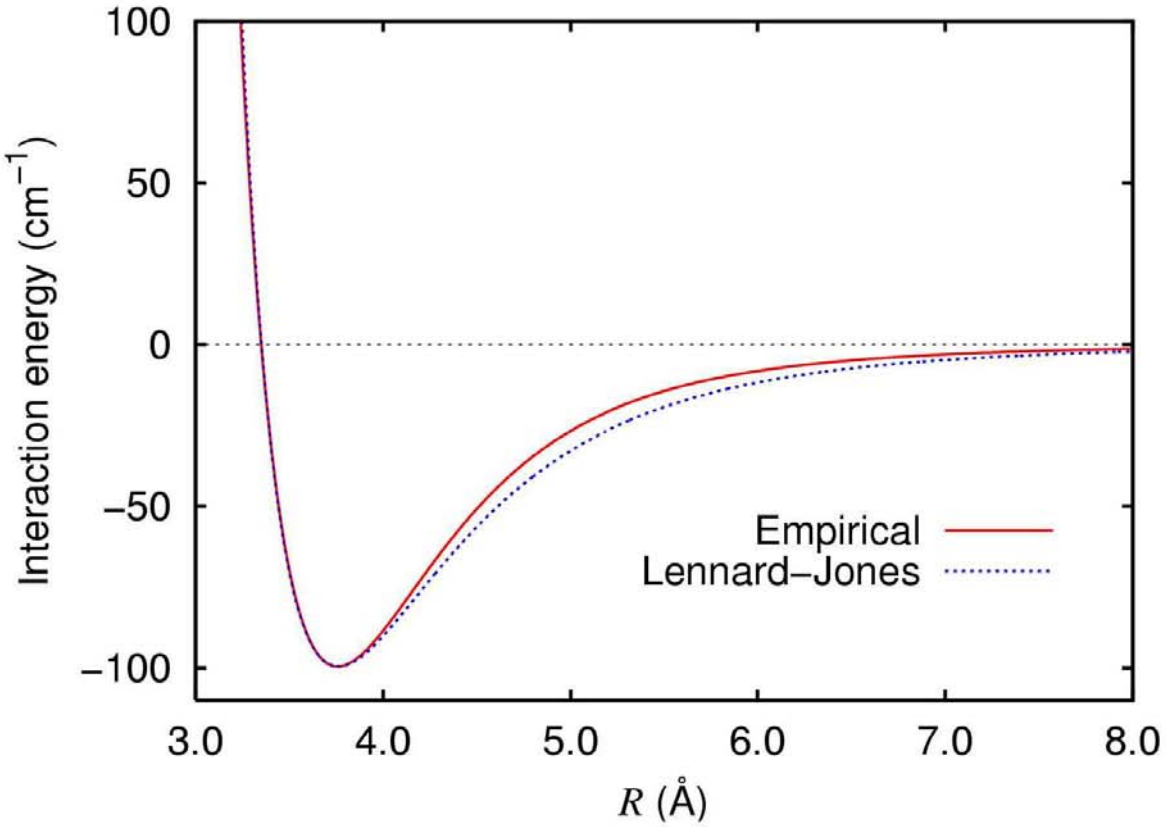


Source: Wikipedia



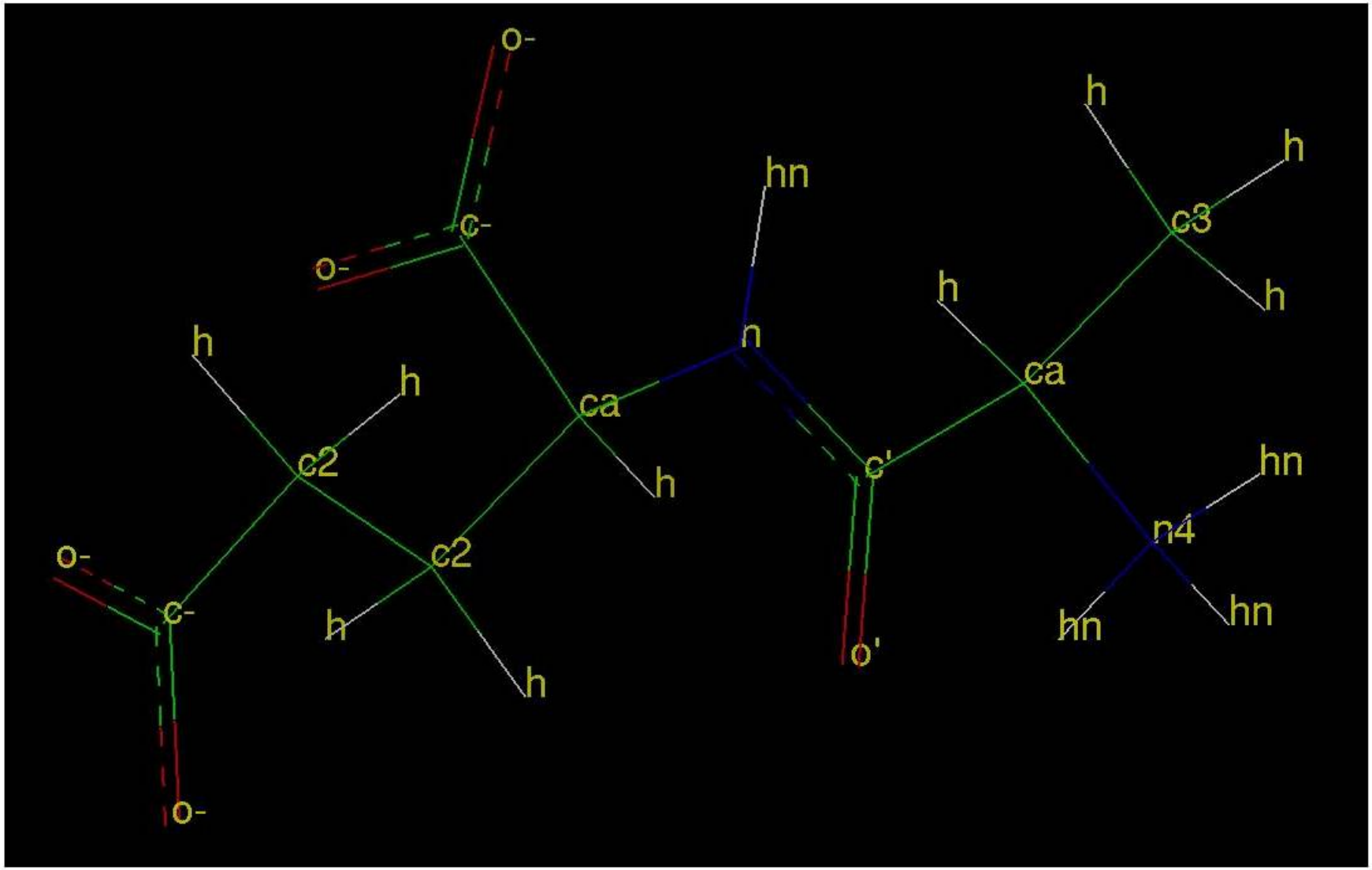


Source: Gromacs manual



Source: Wikipedia

Atom Types



Glutamate-Alanine

Bond Parameters

FORCE_FIE	A	B	C	D
1	Bond		k	R0
2	ca	n4	356.599	1.47
3	hn	n4	457.459	1.026
4	ca	h	340.618	1.105
5	c'	ca	283.092	1.52
6	c3	ca	322.716	1.526
7	c'	o'	615.322	1.23
8	c'	n	388.0	1.32
9	c3	h	340.618	1.105
10	ca	n	377.575	1.46
11	hn	n	483.451	1.026
12	c-	ca	283.092	1.52
13	c2	ca	322.716	1.526
14	c-	o-	540.0	1.25
15	c2	h	340.618	1.105
16	c2	c2	322.716	1.526
17	c-	c2	283.092	1.52
18				

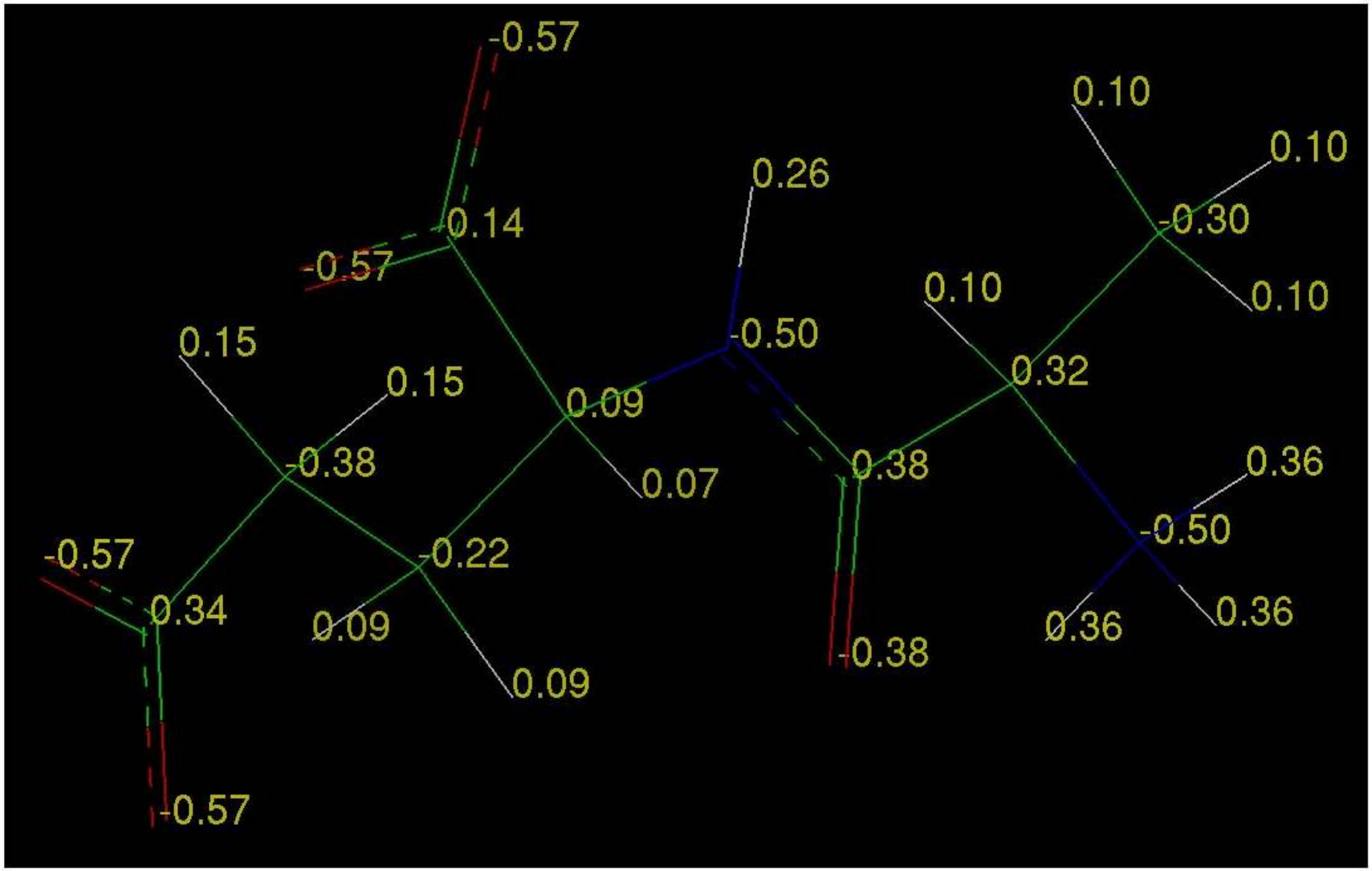
Angle Parameters

FORCE_FIE	A	B	C	D	E
1	Angle			k	R0
2	ca	n4	hn	41.6	110.0
3	hn	n4	hn	36.0	105.5
4	h	ca	n4	57.3	109.5
5	c'	ca	n4	50.0	109.5
6	c3	ca	n4	50.0	109.5
7	c'	ca	h	45.0	109.5
8	c3	ca	h	44.4	110.0
9	c'	ca	c3	46.6	110.5
10	ca	c'	o'	68.0	120.0
11	ca	c'	n	53.5	114.1
12	n	c'	o'	68.0	120.0
13	ca	c3	h	44.4	110.0
14	h	c3	h	39.5	106.4
15	ca	n	hn	35.0	122.0
16	c'	n	ca	111.0	118.0
17	c'	n	hn	37.5	115.0
18	h	ca	n	51.5	109.5
19	c-	ca	n	50.0	109.5
20	c2	ca	n	50.0	109.5
21	c-	ca	h	45.0	109.5
22	c2	ca	h	44.4	110.0
23	c-	ca	c2	46.6	110.5
24	ca	c-	o-	68.0	120.0
25	o-	c-	o-	145.0	123.0
26	ca	c2	h	44.4	110.0
27	c2	c2	ca	46.6	110.5
28	h	c2	h	39.5	106.4
29	c2	c2	h	44.4	110.0
30	c-	c2	c2	46.6	110.5
31	c-	c2	h	45.0	109.5
32	c2	c-	o-	68.0	120.0

Dihedral Parameters

FORCE_FIE	A	B	C	D	E	F	G
1	Torsion				k	n	RO
2	h	ca	n4	hn	0.8	3.0	0.0
3	c'	ca	n4	hn	0.8	3.0	0.0
4	c3	ca	n4	hn	0.8	3.0	0.0
5	o'	c'	ca	n4	0.0	0.0	0.0
6	n	c'	ca	n4	0.0	0.0	0.0
7	h	c3	ca	n4	1.4225	3.0	0.0
8	o'	c'	ca	h	0.0	0.0	0.0
9	n	c'	ca	h	0.0	0.0	0.0
10	h	c3	ca	h	1.4225	3.0	0.0
11	h	c3	ca	c'	1.4225	3.0	0.0
12	o'	c'	ca	c3	0.0	0.0	0.0
13	n	c'	ca	c3	0.0	0.0	0.0
14	ca	c'	n	ca	3.2	2.0	180.0
15	ca	c'	n	hn	1.2	2.0	180.0
16	o'	c'	n	ca	3.8	2.0	180.0
17	o'	c'	n	hn	1.8	2.0	180.0
18	h	ca	n	hn	0.0	0.0	0.0
19	c-	ca	n	hn	0.0	0.0	0.0
20	c2	ca	n	hn	0.0	0.0	0.0
21	h	ca	n	c'	0.0	0.0	0.0
22	c-	ca	n	c'	0.0	0.0	0.0
23	c2	ca	n	c'	0.0	0.0	0.0
24	o-	c-	ca	n	0.0	0.0	0.0
25	h	c2	ca	n	1.4225	3.0	0.0
26	c2	c2	ca	n	1.4225	3.0	0.0
27	o-	c-	ca	h	0.0	0.0	0.0
28	h	c2	ca	h	1.4225	3.0	0.0
29	c2	c2	ca	h	1.4225	3.0	0.0
30	h	c2	ca	c-	1.4225	3.0	0.0

Partial Charges



Glutamate-Alanine

Force Fields

- a set of parameters and functions
- transferable
- (semi-)empirical
- functional form and parameters may differ among force fields (Amber, Charmm, OPLS)

Optimization of the geometry (methods for nearest local minimum):

1. Steepest Descents
2. Conjugate Gradients
3. Newton-Raphson

Steepest Descents

- moving down the steepest gradient (motion is always downhill)
- steps in the direction of the negative gradient of the pot. energy (force)
- does not remember previous steps
- the gradient becomes smaller as the minimum is approached (poor convergence)

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \lambda \nabla V(\mathbf{r}_i)$$

λ - step size parameter (how far to shift coordinates at each step)
If energy drops we increase λ by 20%, if energy rises λ is halved.

Newton-Raphson:

Mimimum $\rightarrow f'(x) = 0$

$$f'(x) = f'(x_0) + f''(x_0)(x - x_0) + \ldots = 0$$

$$-f'(x_0) = f''(x_0)(x - x_0)$$

$$(x - x_0) = -f''(x_0)^{-1} f'(x_0)$$

$$x = x_0 - f''(x_0)^{-1} f'(x_0)$$

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \mathbf{H}^{-1}(\mathbf{r}_i)\nabla V(\mathbf{r}_i)$$

Hamiltonian dynamics

$$H = H(p_1, \ldots, p_N, q_1, \ldots, q_N, t)$$

$$\begin{cases} \dot{p}_i = -\frac{\partial H}{\partial q_i} \\ \dot{q}_i = \frac{\partial H}{\partial p_i} \end{cases} i = 1, \ldots, N$$

N - number of degrees of freedom

$$H = \frac{\mathbf{p}^2}{2m} + U(\mathbf{r})$$

$$\dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i} = -\frac{\partial U}{\partial \mathbf{r}_i} = \mathbf{F}_i = m\ddot{\mathbf{r}}_i$$

Based on the potential energy function V, we can find components of the force acting on an atom as:

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i}$$

Newton's equation of motion:

$$\mathbf{F}_i = m_i \mathbf{a}_i$$

$$-\frac{dV}{d\mathbf{r_i}} = m_i \frac{d^2\mathbf{r}}{dt^2}$$

Change of the potential energy is related to the changes in position as a function of time

Euler integration (first order method)

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \mathbf{a}(t)\Delta t$$

$$\mathbf{a}(t) = -1/m \frac{\partial V}{\partial \mathbf{r}_i}$$

$$\mathbf{v}(t) = \frac{d\mathbf{r}(t)}{dt}$$

To calculate a trajectory we need:

- initial positions of the atoms (e.g. X-ray or NMR structures)
- an initial distribution of velocities
- the acceleration, which is determined by the gradient of the potential energy function

Euler method for one dimensional harmonic oscillator:

$$x(t + \Delta t) = x(t) + v(t)\Delta t$$

$$v(t + \Delta t) = v(t) + a(t)\Delta t$$

$$a(t) = -\frac{k}{m}x(t) = -\omega^2 x(t)$$

Time integration algorithms are based on finite difference methods (from positions and velocities at time t we derive the same quantities at time t+Δt)

Errors:

- Truncation errors - methods are based on a Taylor series expansion truncated at some term

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt}\Delta t + 1/2\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 + ...$$

  where t - current time and Δt - time step

- Round-off errors - related to a particular implementation of the algorithm

Verlet algorithm:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + 1/2\mathbf{a}(t)\Delta t^2 + 1/6\mathbf{b}(t)\Delta t^3 + O((\Delta t)^4)$$

$$\mathbf{r}(t-\Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + 1/2\mathbf{a}(t)\Delta t^2 - 1/6\mathbf{b}(t)\Delta t^3 + O((\Delta t)^4)$$

$$\mathbf{r}(t+\Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \mathbf{a}(t)\Delta t^2 + O(\Delta t^4)$$

Velocities can be calculated from

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t)}{2\Delta t}$$

Trajectory - coordinates as a function of time

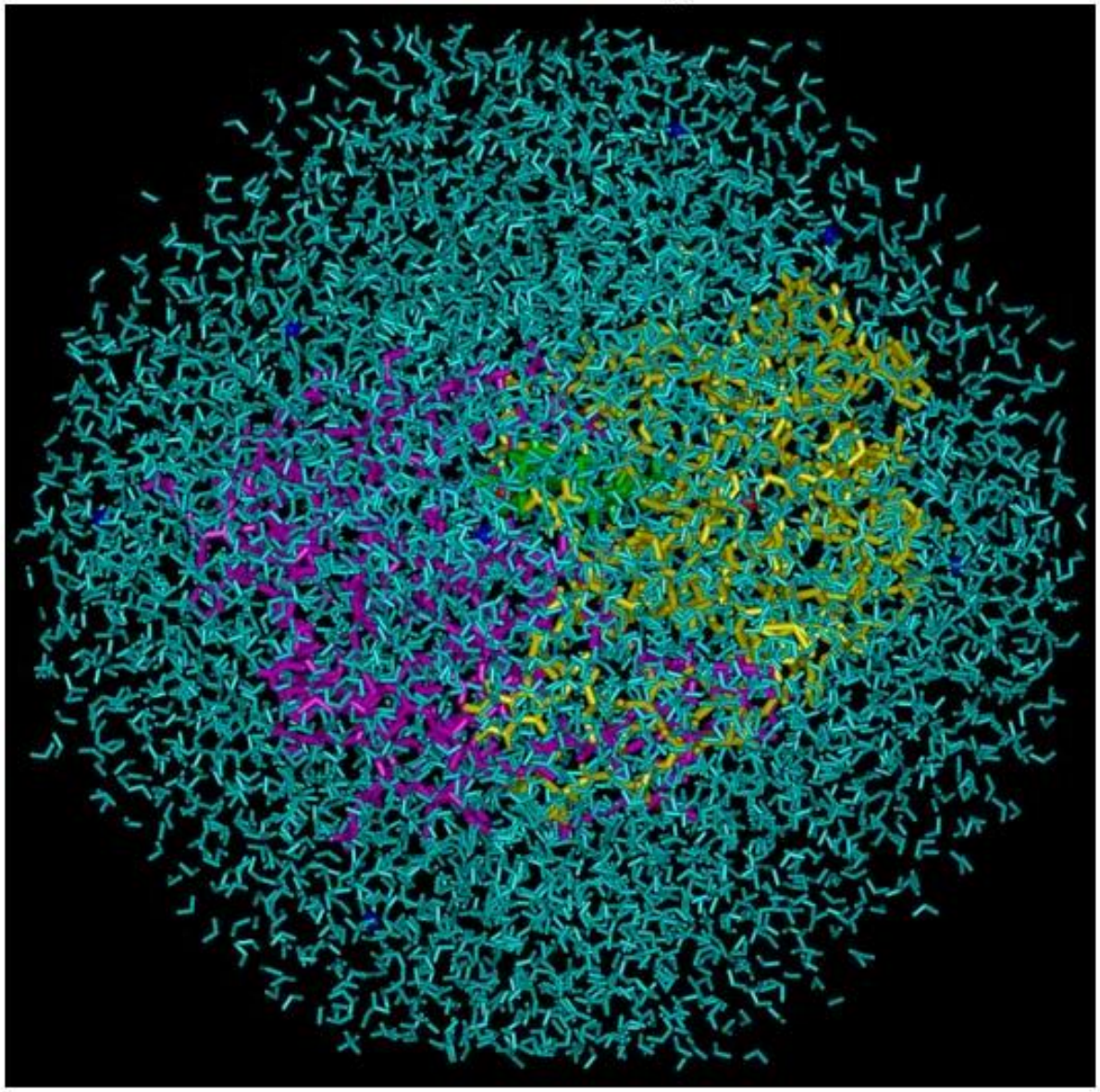
How to choose a time step?

- for all-atom description of the system Δt=0.5-2 fs
- atomic vibrations - characteristic times 10^{-14}-10^{-11}s

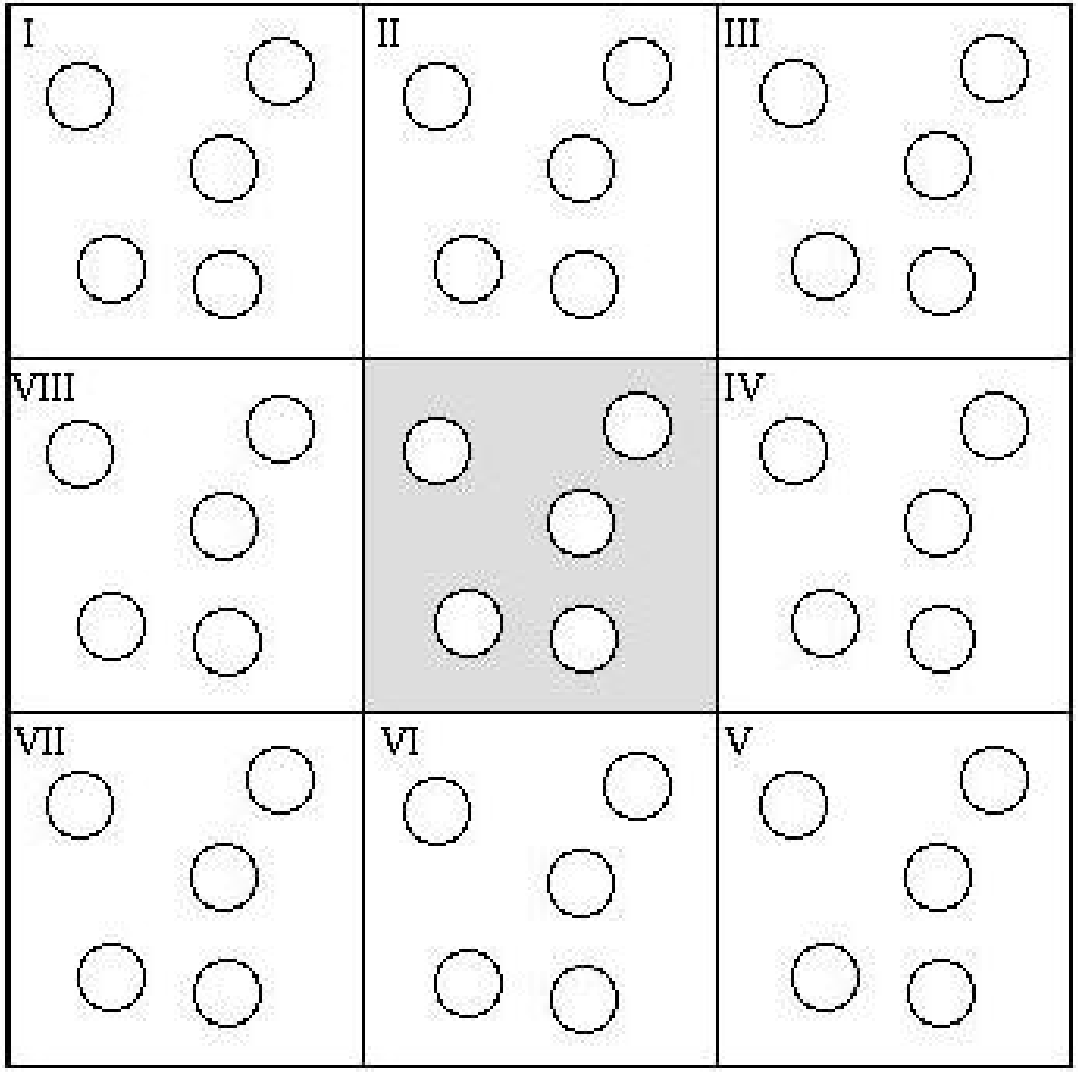
Limitations:

- potential energy function (force field) is approximate
- force field is pair-additive
- long range interactions are truncated
- periodic boundary conditions

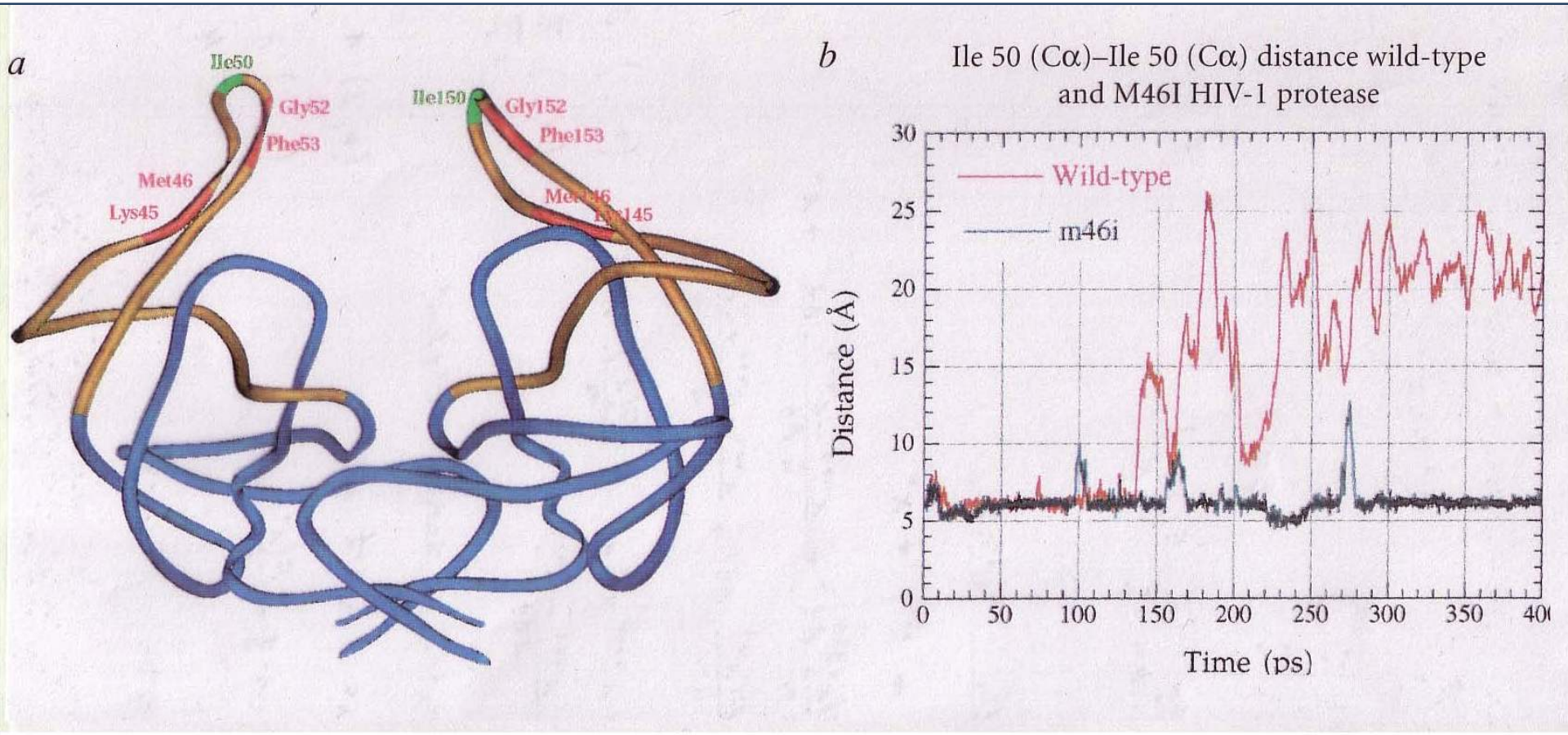
MD setup



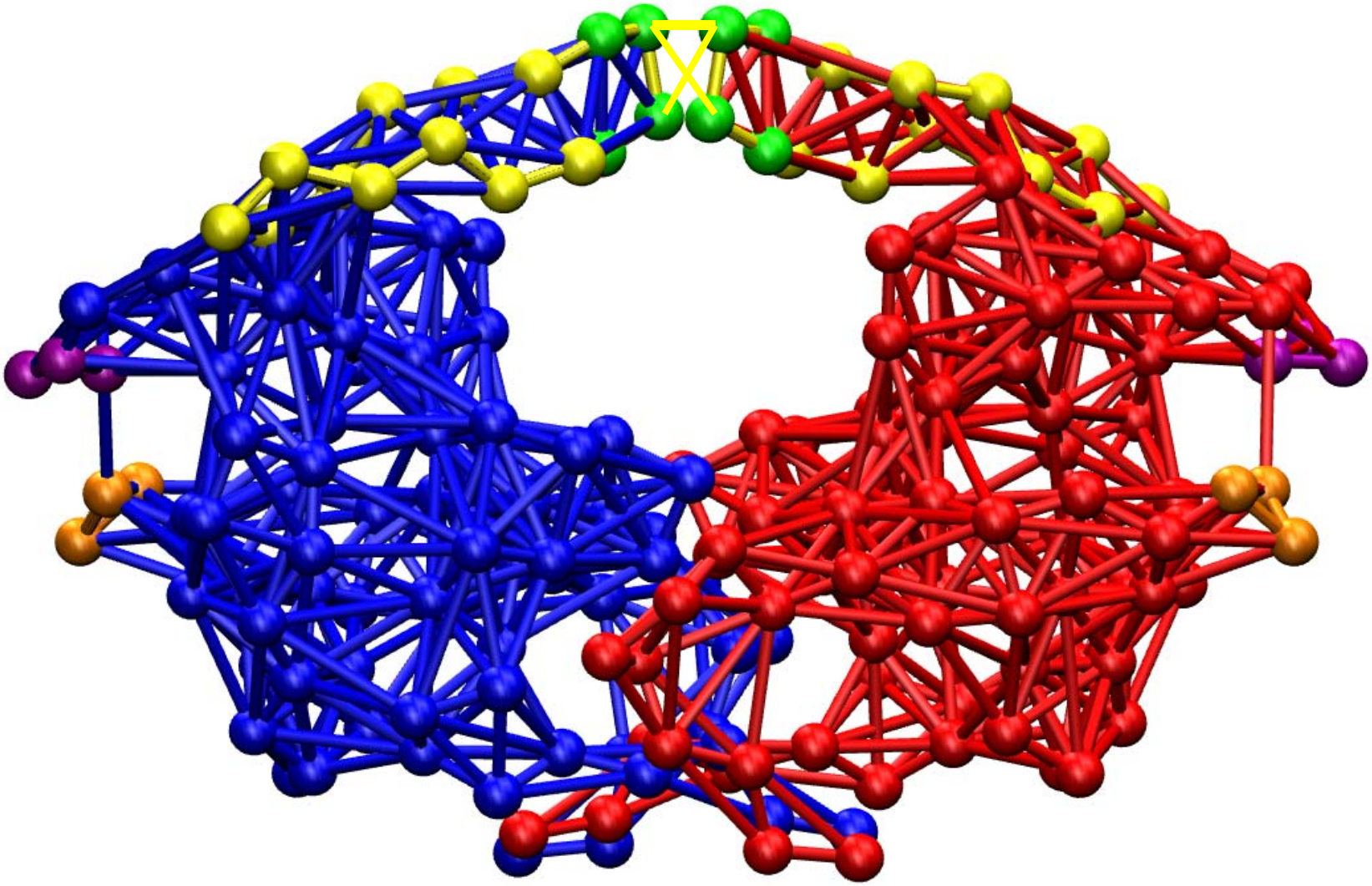
Periodic boundary conditions



http://www.chembnet.org/

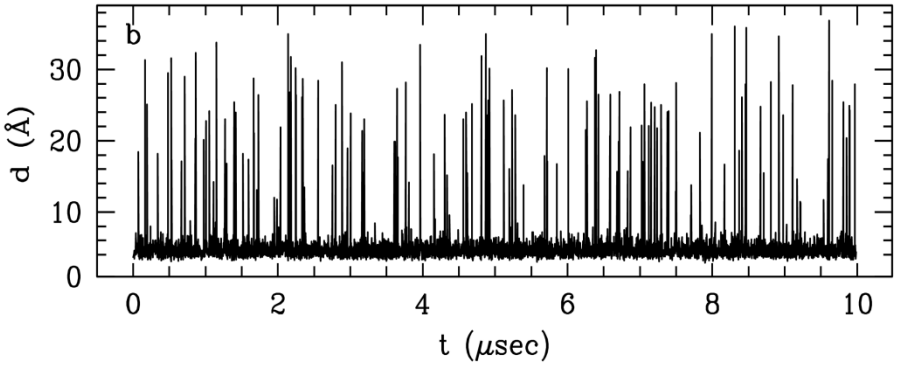

a
Ile50
Gly52
Phe53
Met46
Lys45
Ile150
Gly152
Phe153
b
Ile 50 (Cα)–Ile 50 (Cα) distance wild-type and M46I HIV-1 protease
Wild-type
m46i
Distance (Å)
Time (ps)
0
5
10
15
20
25
30
0
50
100
150
200
250
300
350

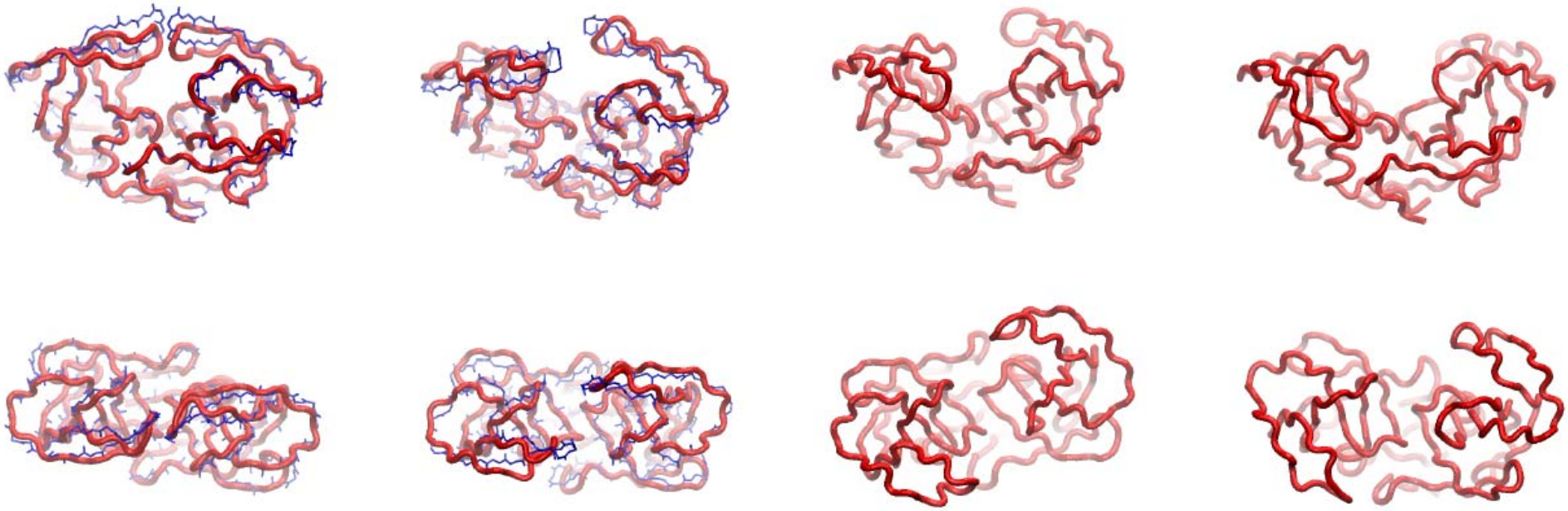
Elastic network model



Flap opening dynamics

"Bunches" of opening events separated by longer closed intervals: Tossini et al., J.Struct.Biol., 157, 606 (2007)





closed
~ 5Å

semi-open
~ 12Å

open
~ 20Å

wide-open
~ 35Å

Principal component analysis – (Essential Dynamics)

- It serves to identify significant directions of global correlated motions.

- One solves the eigenvalue problem for the symmetric 3Nx3N covariance matrix derived from the MD trajectory.

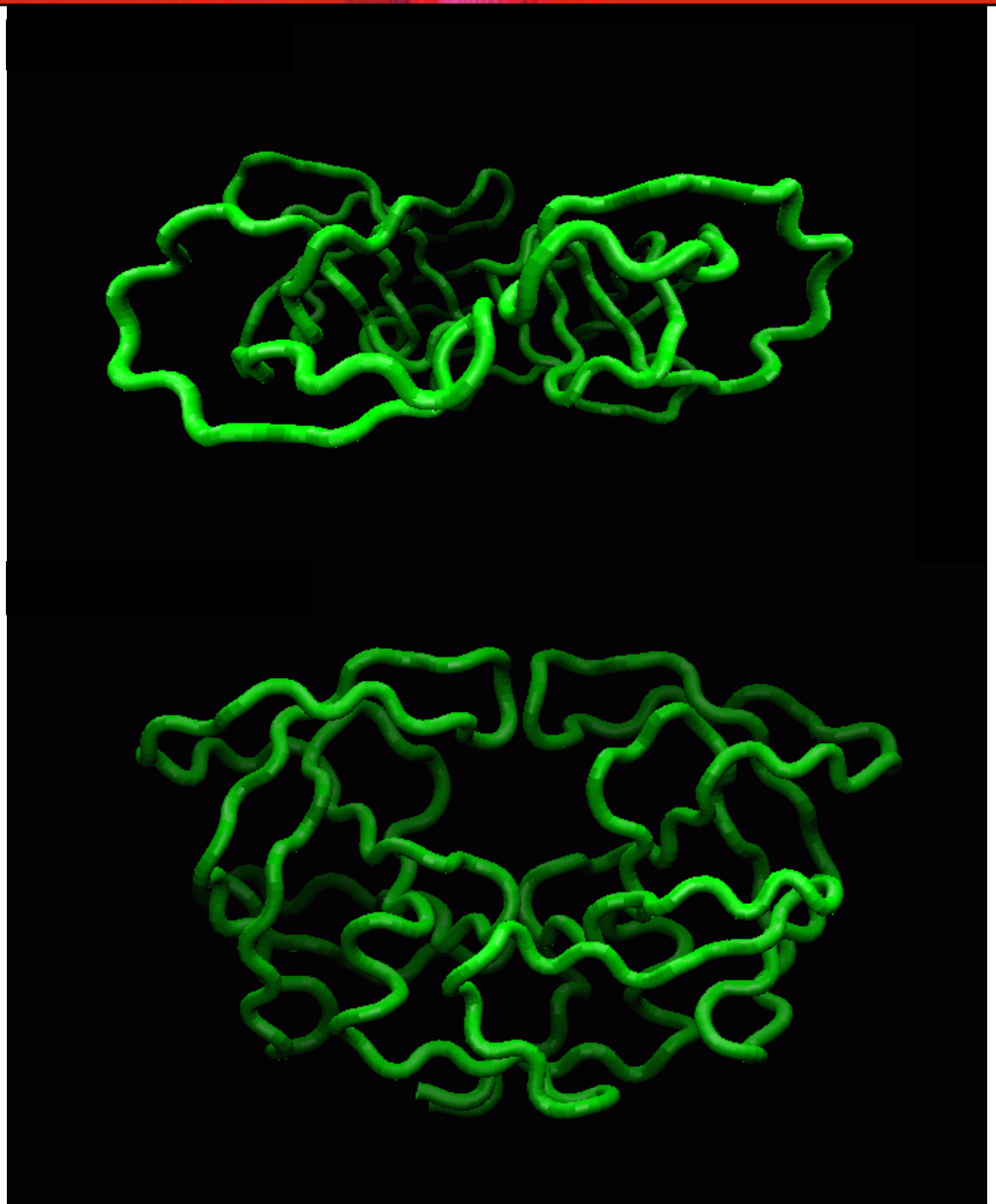
$$C_{ij} = \langle (x_j - \langle x_j \rangle)(x_i - \langle x_i \rangle) \rangle$$

$$dim(C_{ij}) = 3N \times 3N$$

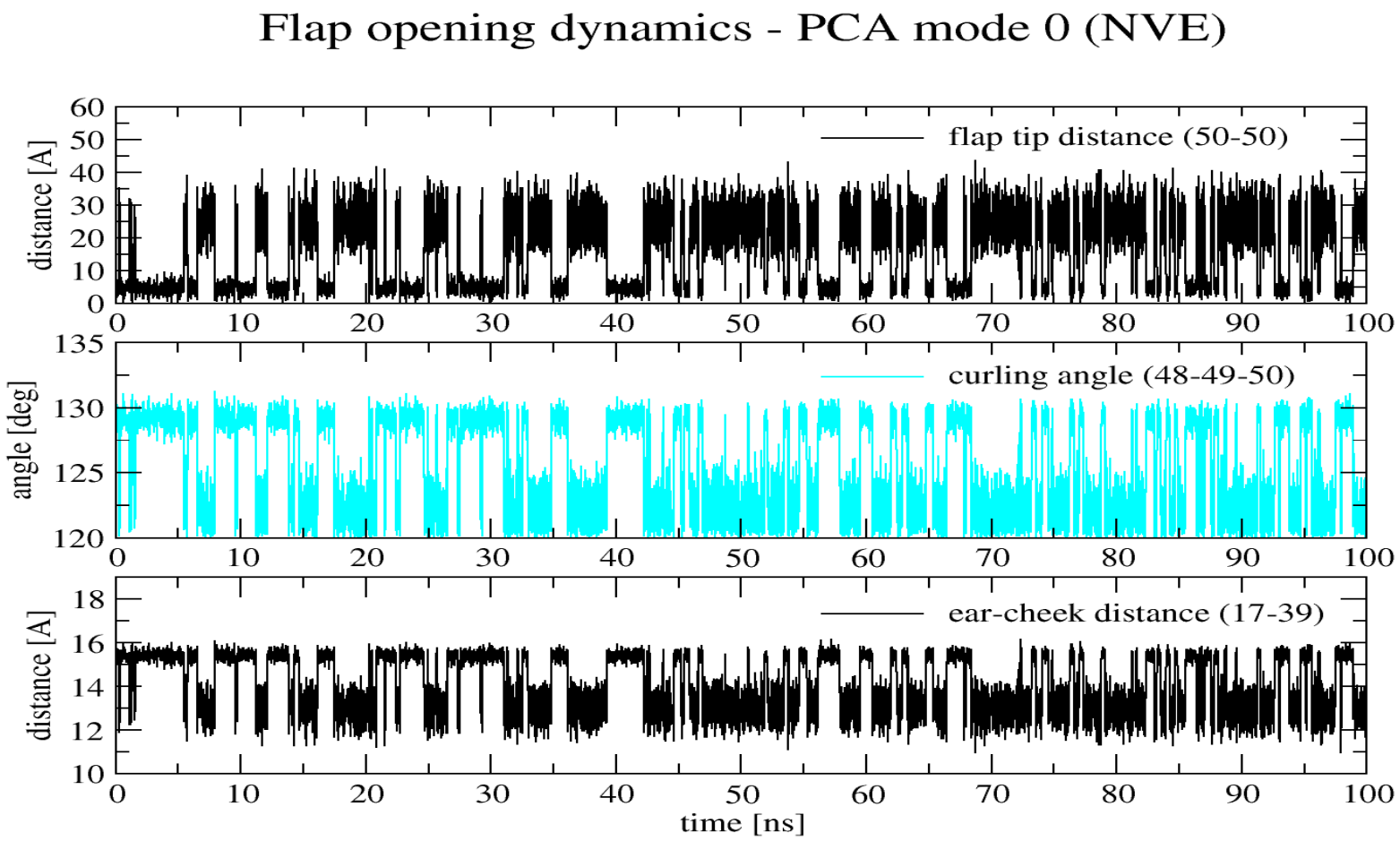
- The brackets represent the time averages over all trajectories.
- The eigenvectors give directions, and eigenvalues quantify the magitude of the fluctuations.
- The trajectory can be projected onto the eigenvectors - the first several represent most of the functional motions of the system.

See: A. Amadei,A.B.M. Linsenn, H.J.C. Berendsen, Proteins: Struct. Funct.Genet., 17, 412-425(1993)

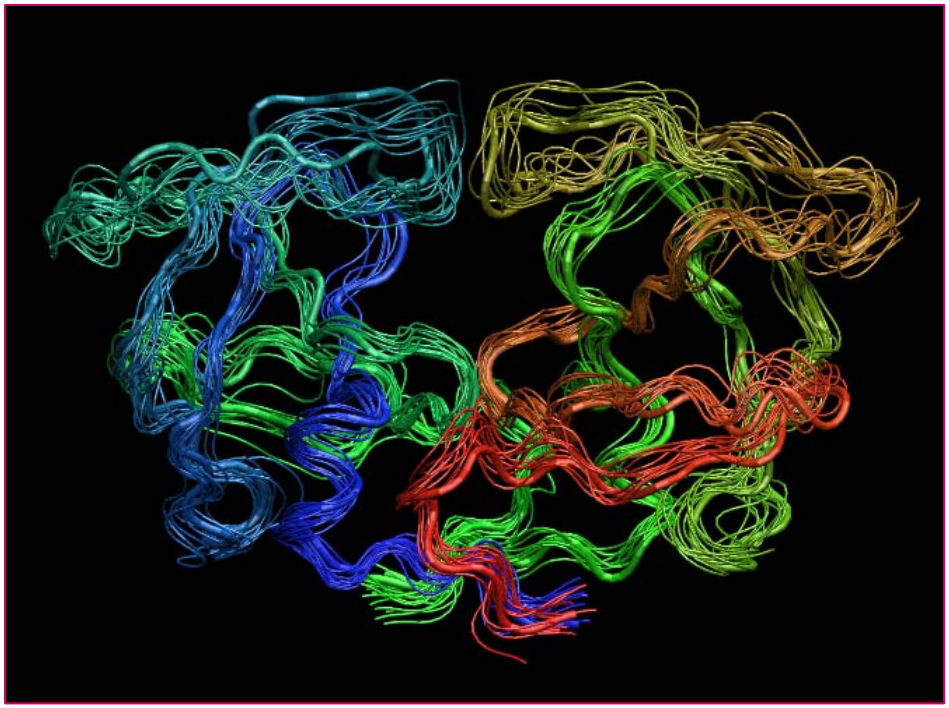
The first normal mode of HIV-1

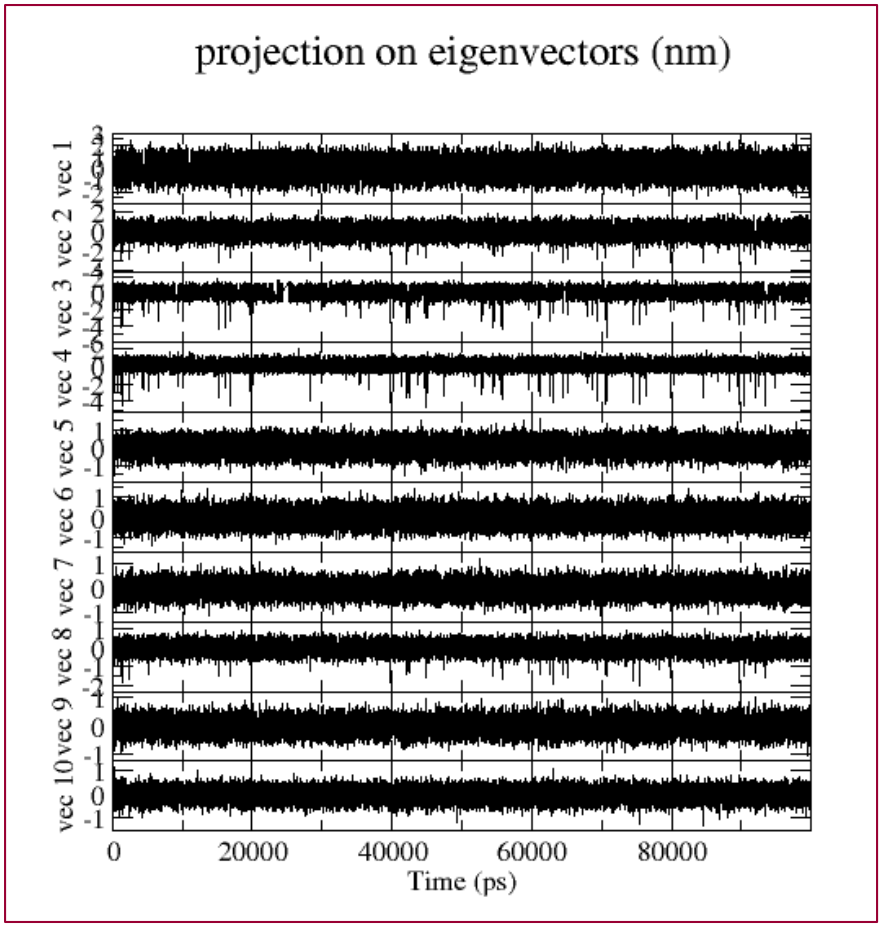


Sygnals: MD trajectories, their combinations or projections on N-modes



Decomposition of atomic motions normal modes





Wnioski końcowe

- Dynamika molekularna włącznie z kwantowymi metodami *ab initio* są bardzo efektywnymi metodami badawczymi, pozwalającym na opis różnorodnych mechanizmów przemian strukturalnych w układach (bio)molekularnych, jak również np. nanostruktur.
- Przy ustalonej topologii (bez zrywania wiązań chemicznych) badanego układu metodami MD można symulować różnorodne przemiany konformacyjne.
- Metody QD oraz QCMD pozwalają również na symulacje reakcji chemicznych oraz biochemicznych, w tym enzymatycznych.
- Wszystkie metody w konfrontacji z dostępnymi danymi doświadczalnymi pozwalają na pogłębione zrozumienie badanych mechanizmów i procesów.
- Rozwija się nowa dziedzina badająca relacje przyczynowości w symulacjach MD/QD/QCMD tzw. „causality analysis”

Podziękowania:

W prezentacji zostały wykorzystane niektóre slajdy dr hab. Joanny Trylskiej, m.in. te dotyczące proteazy kodowanej przez genom wirusa HIV-1