

**ANALIZA TRAJEKTORIJA RAČUNALNIH  
(SIMULACIJSKIH) METODA U ISTRAŽIVANJU  
BIOPOLIMERA (PROTEINA I NUKLEINSKIH KISELINA)**

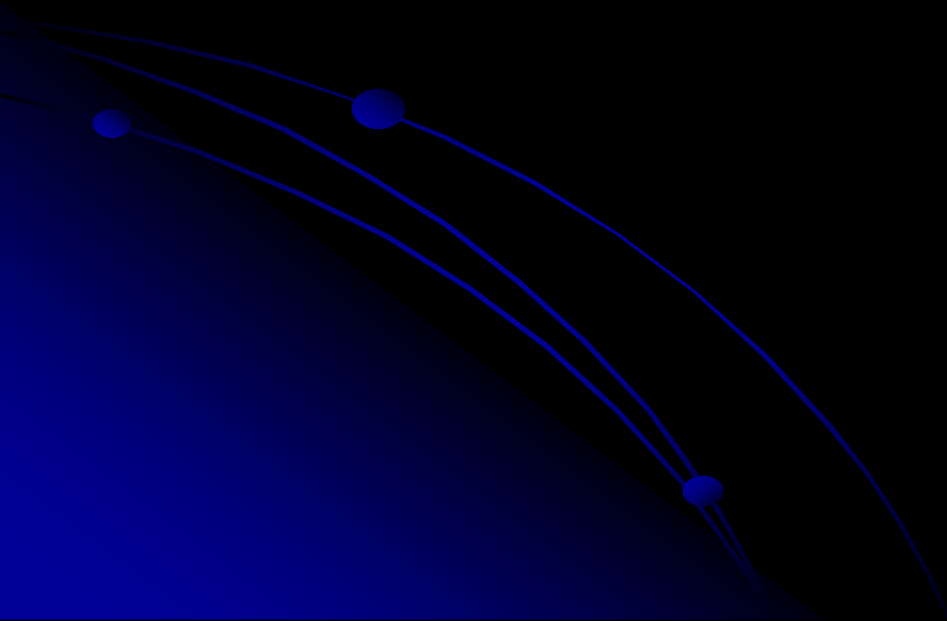
Kolegij:

Strukturalna računalna biofizika



# ANALIZA TRAJEKTORIJA

- VIZUALIZACIJA TRAJEKTORIJE
- KVANTITATIVNA ANALIZA TRAJEKTORIJE



# Vizualizacija

## Programi za vizualizaciju trajektorija:

- Visual Molecular Dynamics (**VMD**) (<http://www.ks.uiuc.edu/Research/vmd/>)
- **Chimera** (<http://www.cgl.ucsf.edu/chimera/>)

izvršni za vizualizaciju, ali ne Gromacs trajektorija:

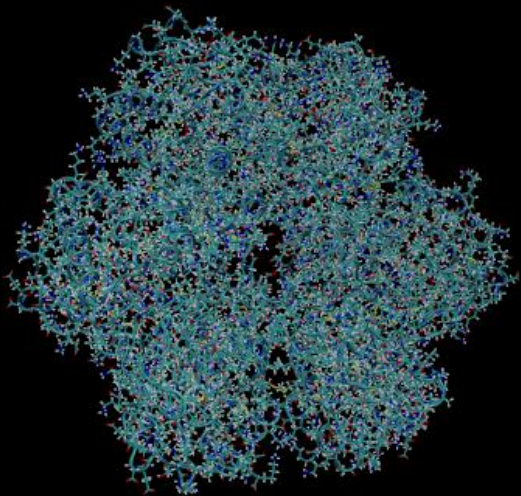
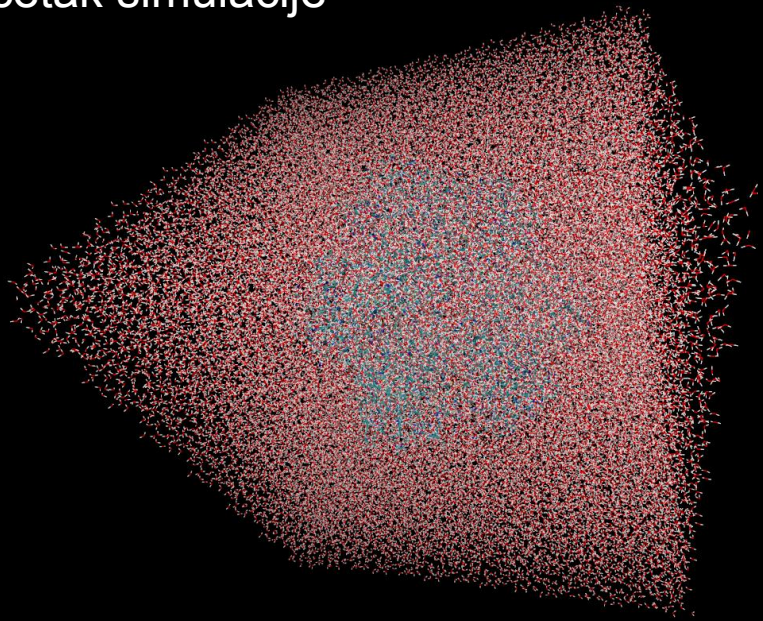
- **Maestro** (<http://www.schrodinger.com/Maestro/>)
- **Pymol** (<http://www.pymol.org/>)
- **Accelrys** (<http://accelrys.com>)

# Vizualizacija

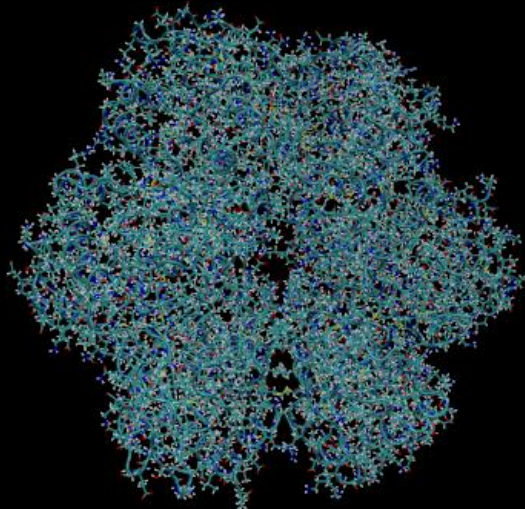
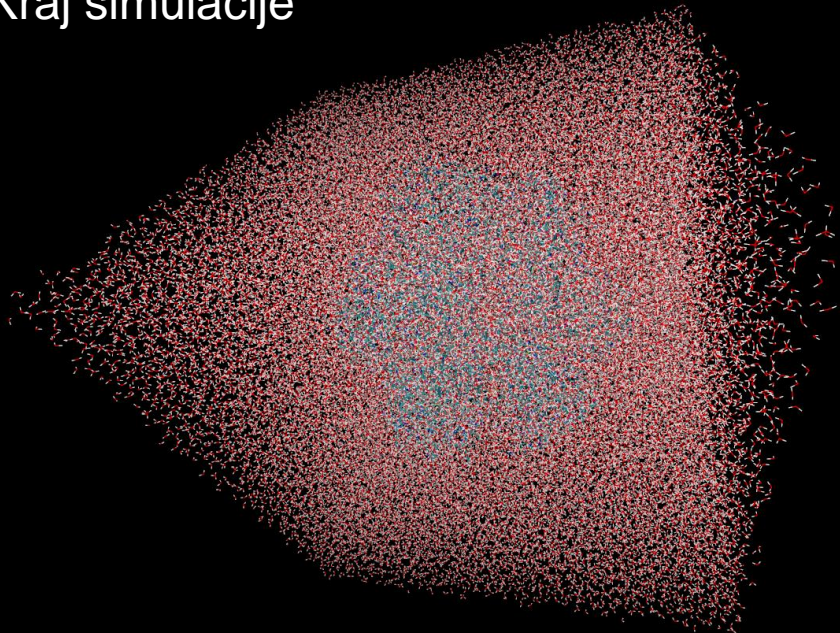
- najvažniji dio analize simulacija
- sve rezultate vizualizacije potrebno je potvrditi „statistikom”
- radi preglednije vizualizacije neophodno je koristiti različite prikaze te ponekad selektivno prikazati samo pojedine dijelove strukture

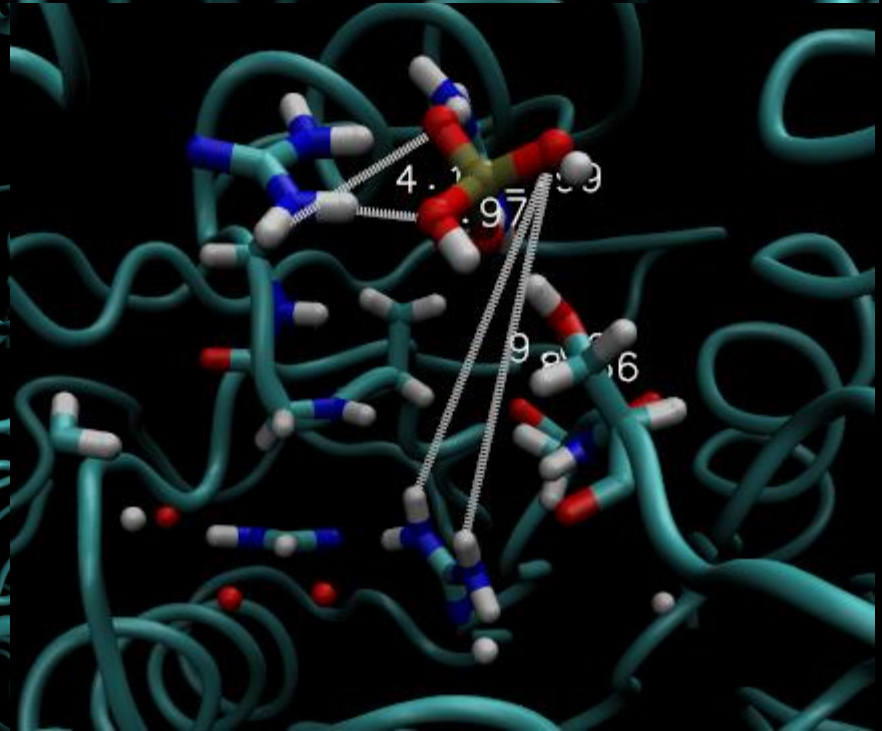
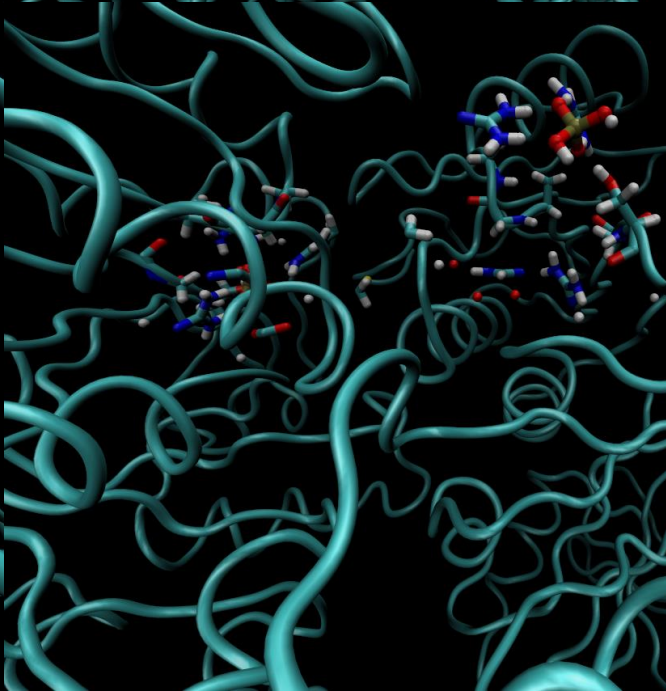
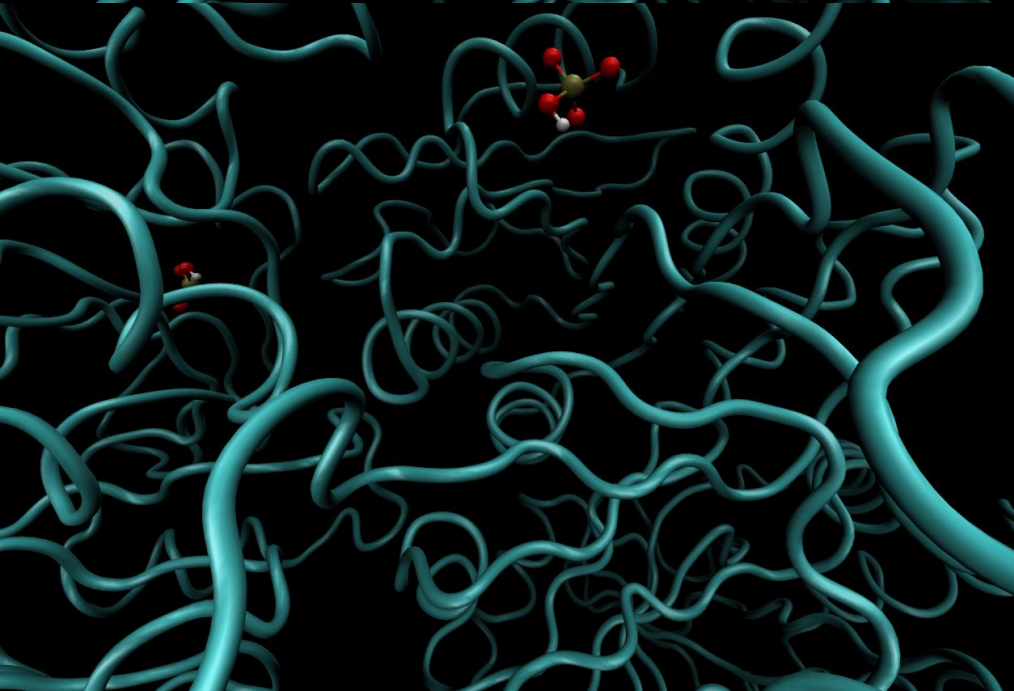
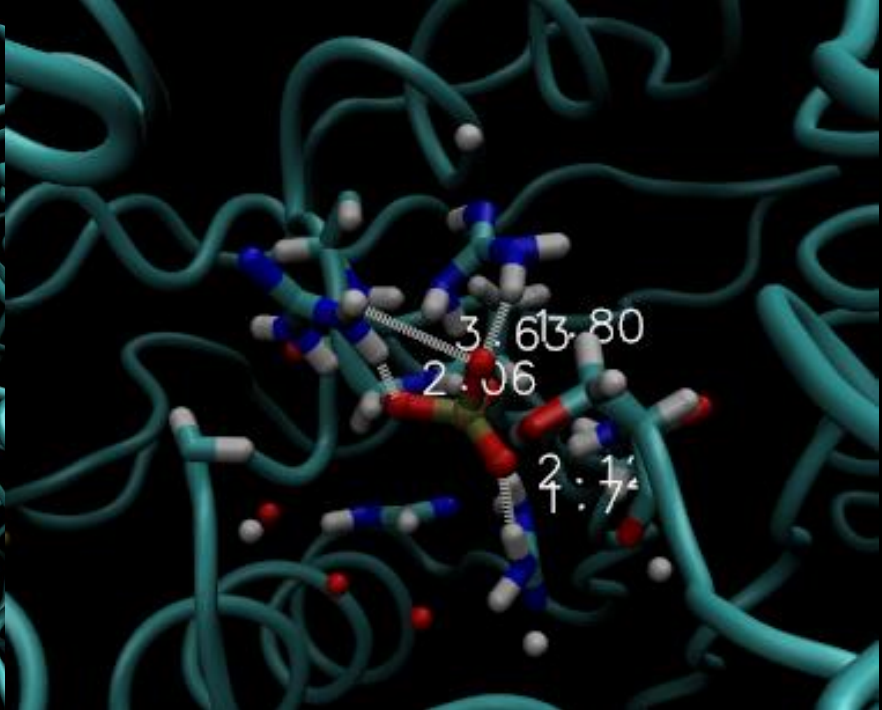
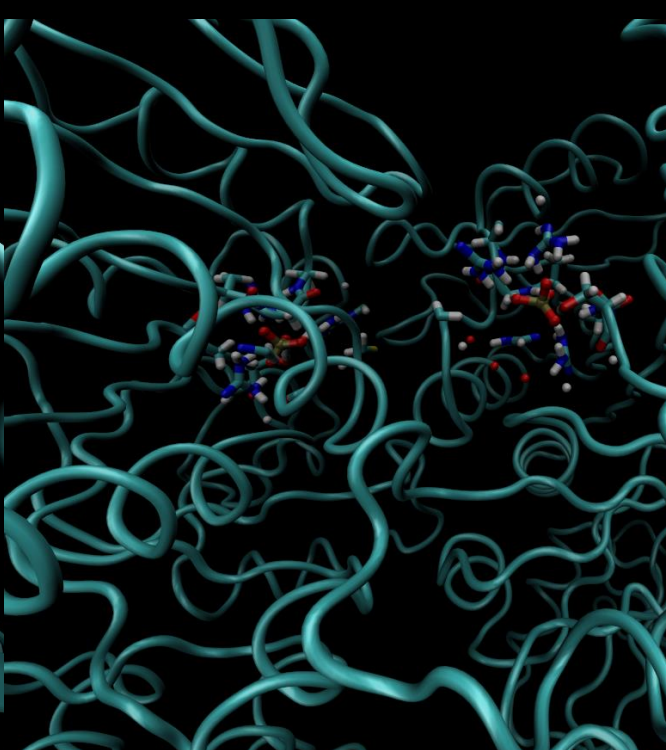
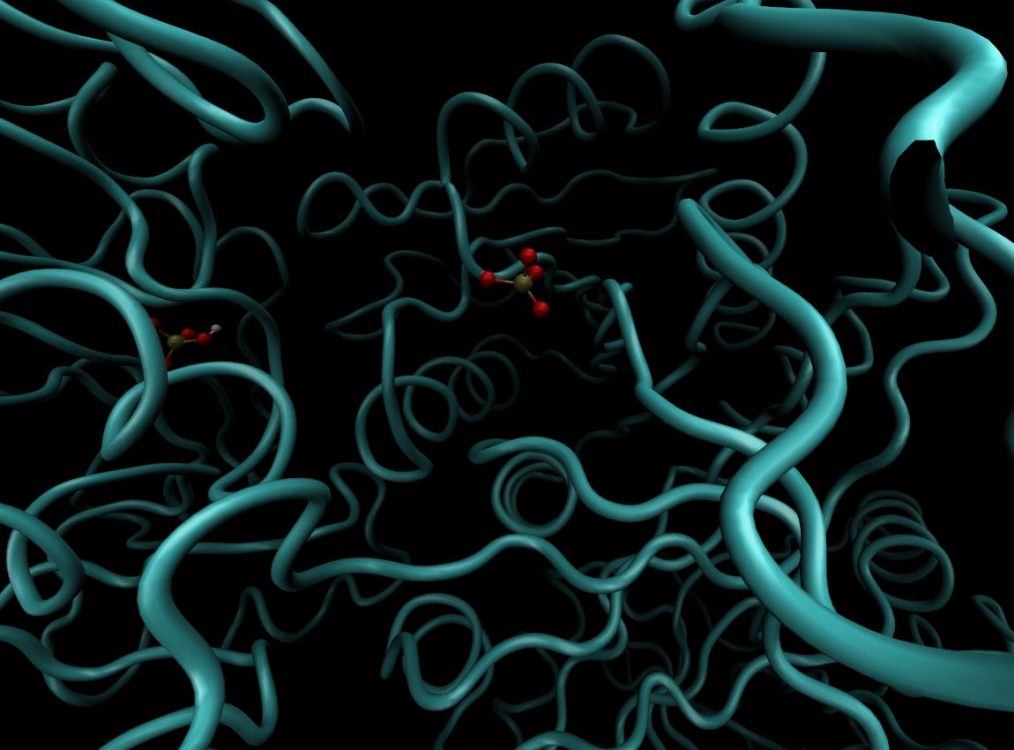
- Primjer – PNP – pomak fosfatnog iona:

Početak simulacije



Kraj simulacije





# RMSD

RMSD - *root-mean-square deviation* (srednji korijen kvadrata udaljenosti) – je mjera prosječne udaljenosti ekvivalentnih atoma dvije strukture, a time i mjera sličnosti dviju struktura. Najčešće se koristi kao statistički pokazatelj sličnosti dvije konformacije iste molekule, ali može se upotrijebiti i za uspoređivanje različitih molekula. Manja RMSD vrijednost predstavlja veću sličnost (iste konformacije imaju RMSD vrijednost nula).

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

$N$  – ukupan broj atoma koje uspoređujemo,  $\delta$  – udaljenost dva ekvivalentna atoma

# RMSD

RMSD vrijednost (najčešće, ali ne nužno, proteinske okosnice) se prati tijekom trajektorije na način da se sukcesivne konformacije dobivene tijekom simulacije uspoređuju s početnom. Na takav način se dolazi do informacije o mijenjanju proteinske strukture tijekom simulacije.

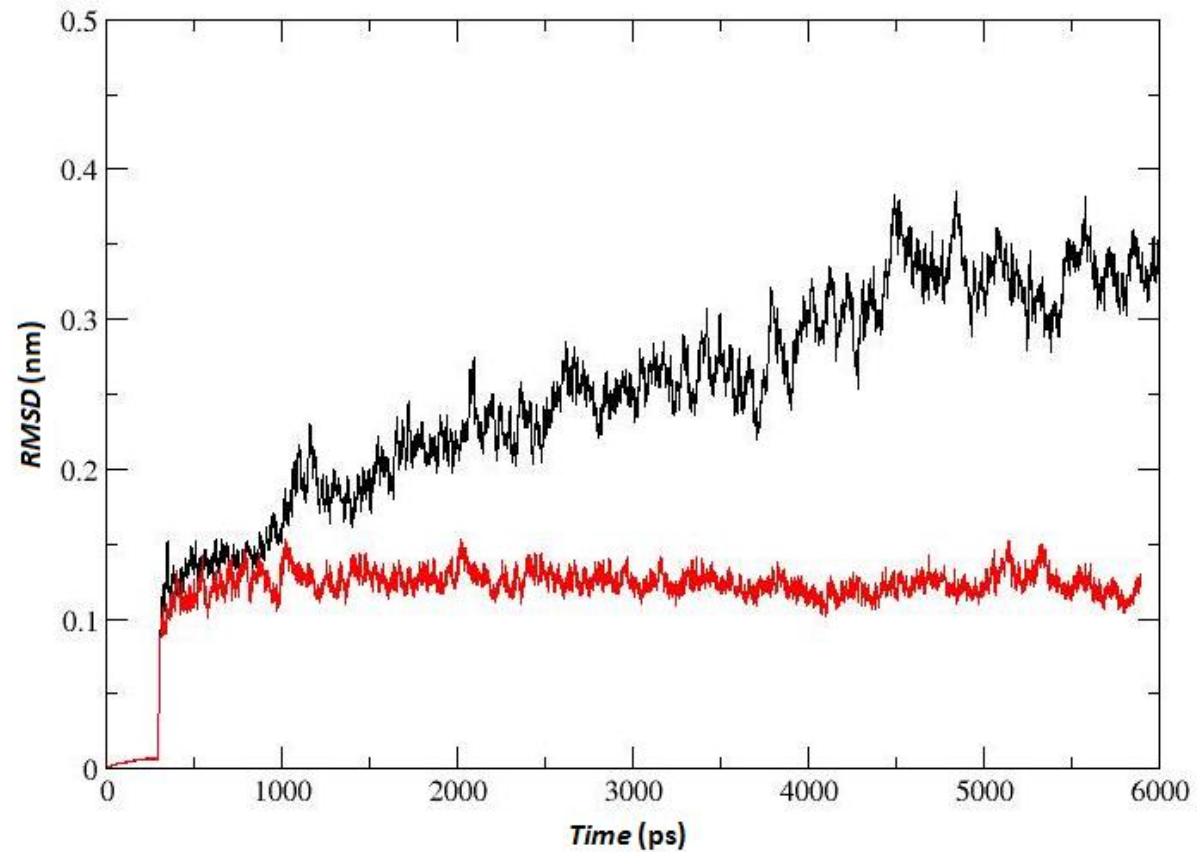
$$RMSD(t_1, t_2) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \| \mathbf{r}_i(t_1) - \mathbf{r}_i(t_2) \|^2 \right]^{\frac{1}{2}}$$

$$M = \sum_{i=1}^N m_i$$

$r(t)$  – položaj atoma u vremenu  $t$ .

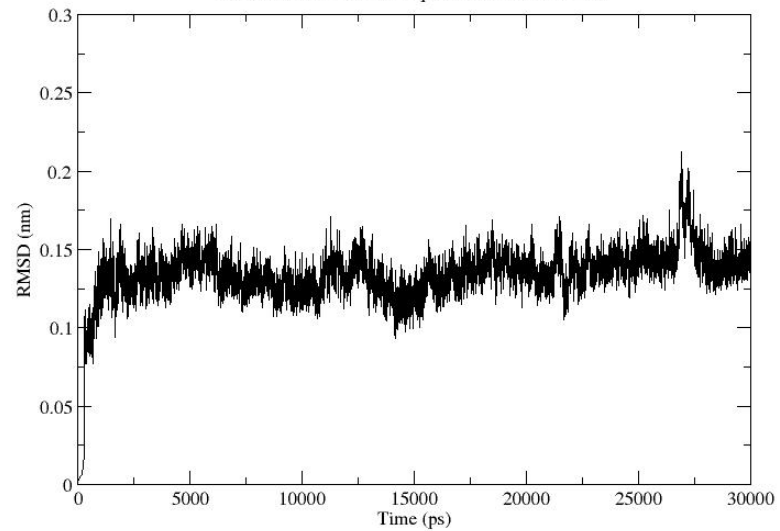


# RMSD - primjeri



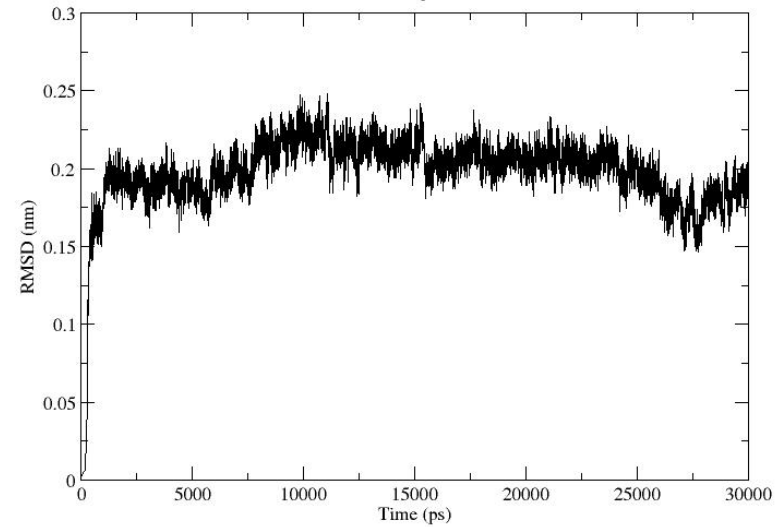
### RMSD APO

BACKBONE\_CA after lsq fit to BACKBONE\_CA



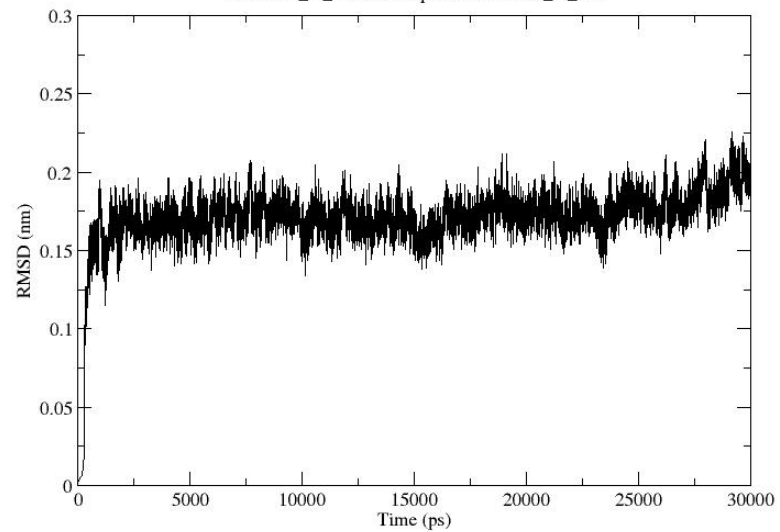
### RMSD SDP-TAP-HIP

SUSTAV\_&\_CA after lsq fit to SUSTAV\_&\_CA



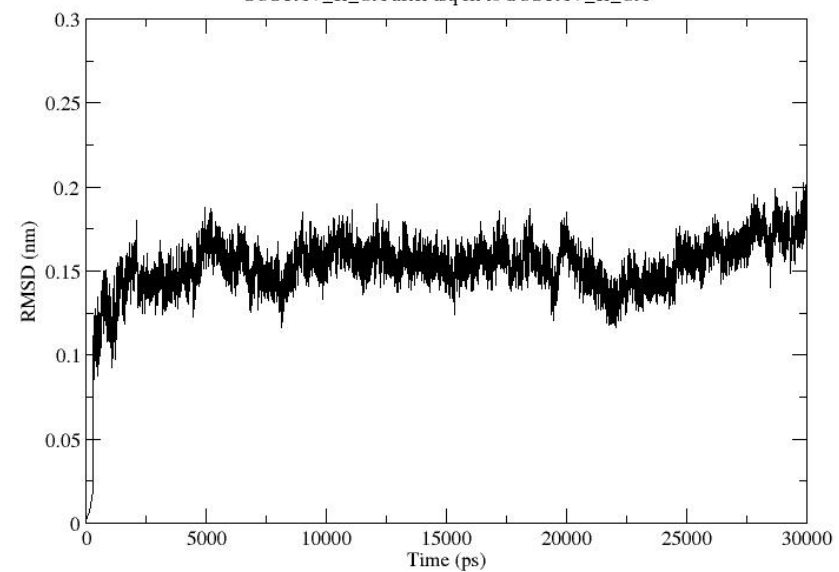
### RMSD OCT-TAP

SUSTAV\_&\_CA after lsq fit to SUSTAV\_&\_CA



### RMSD COA12-TAP

SUSTAV\_&\_CA after lsq fit to SUSTAV\_&\_CA



# RMSF

RMSF - *root-mean-square fluctuation* (srednji korijen kvadrata fluktuacija) – je mjera prosječnog fluktuiranja pojedinog atoma, aminokiseline ili dijela proteina tijekom simulacije. Predstavlja mjeru fleksibilnosti pojedinog dijela strukture makromolekule. Iz tih vrijednosti se lako mogu izračunati B-faktori koji se koriste u kristalografiji.

$$RMSF = \sqrt{\frac{1}{T} \sum_{t_j=1}^T (x_i(t_j) - \tilde{x}_i)^2}$$

referentni položaj atoma  
(najčešće u početnoj  
strukturi trajektorije)

položaj atoma u vremenu  $t_j$   
tijekom simulacije

$T$  – ukupno vrijeme trajanja trajektorije,  $x$  – položaj atoma

# RMSF

RMSF vrijednosti nam pokazuju koliko neki atom, aminokiselina ili proteinska regija fluktuiraju tijekom simulacije što nam omogućava istraživanje njihovih dinamičkih svojstava, odnosno omogućava nam identifikaciju fleksibilnih i rigidnih dijelova strukture makromolekule.

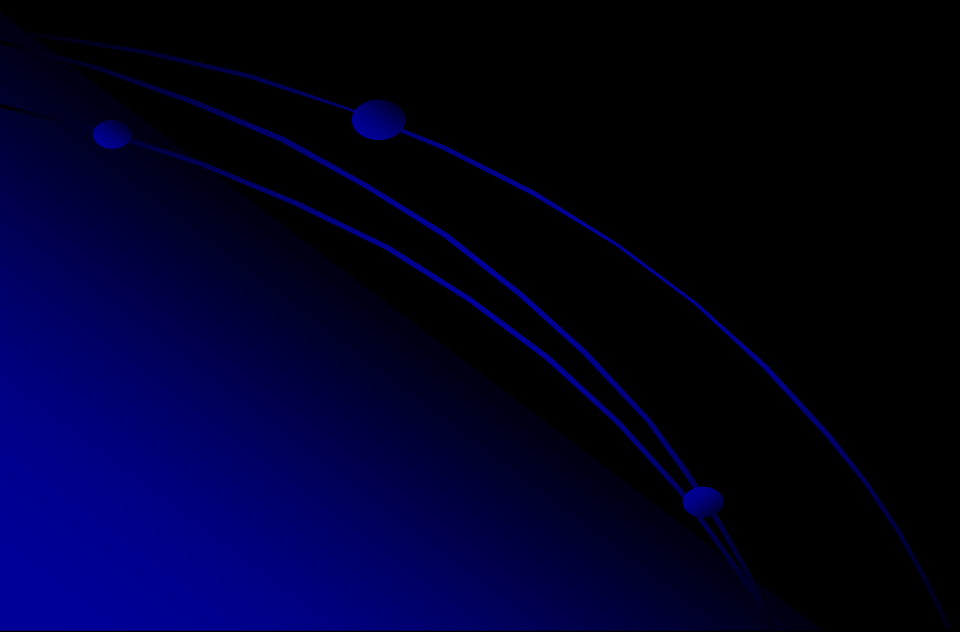
$$RMSD(t_1, t_2) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \| \mathbf{r}_i(t_1) - \mathbf{r}_i(t_2) \|^2 \right]^{\frac{1}{2}}$$

$$M = \sum_{i=1}^N m_i$$

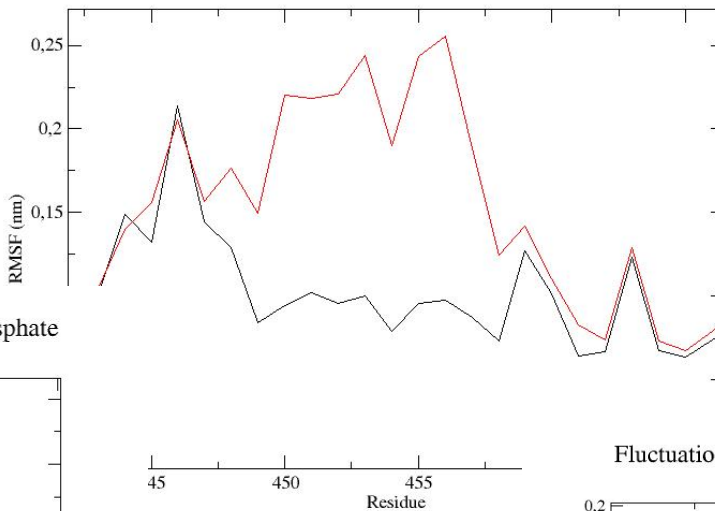
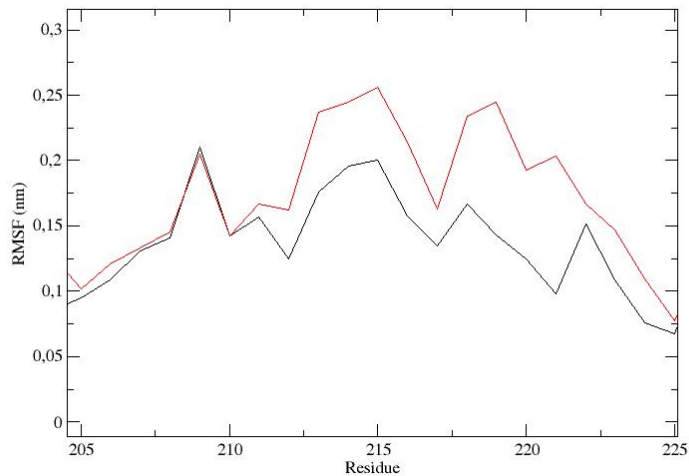
$r(t)$  – položaj atoma u vremenu  $t$ .

Za razliku od RMSD, RMSF je usrednjena vrijednost tijekom vremena što daje srednju vrijednost fluktuiranja za svaku česticu (atom, amino kiselinu ili dio proteina) tijekom simulacije. Promatranjem RMSF vrijednosti možemo detektirati fleksibilnije i rigidnije regije u makromolekuli, što nam uvelike olakšava vizualizaciju, ali i cjelokupnu analizu.

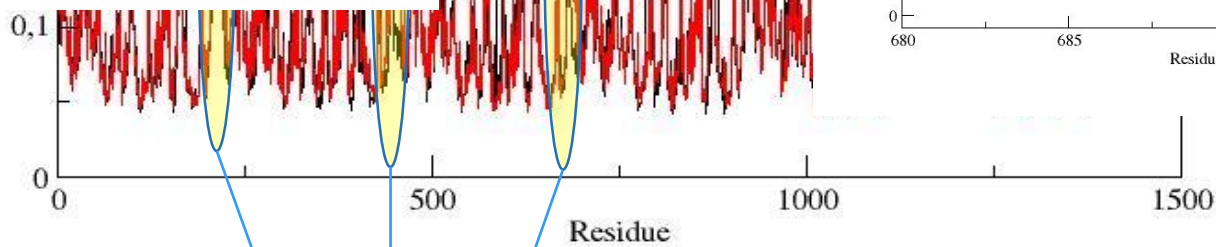
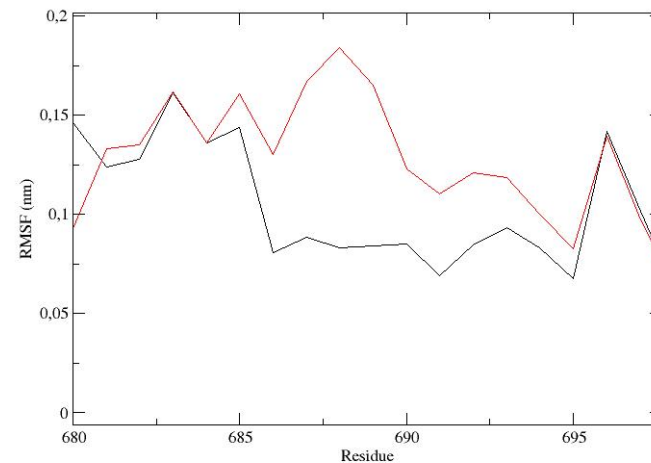
# RMSF - primjeri



Fluctuations of R24A PNP binary complex with phosphate



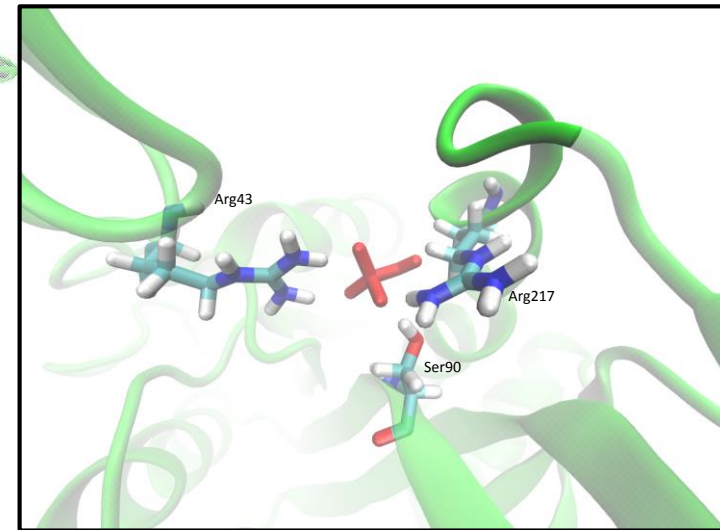
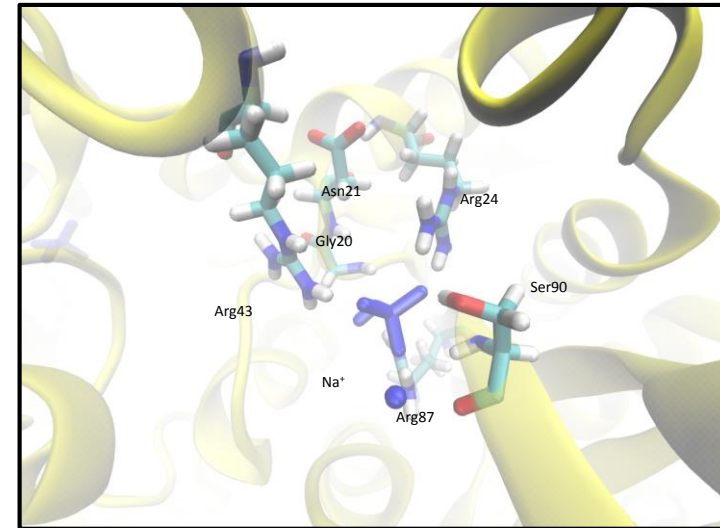
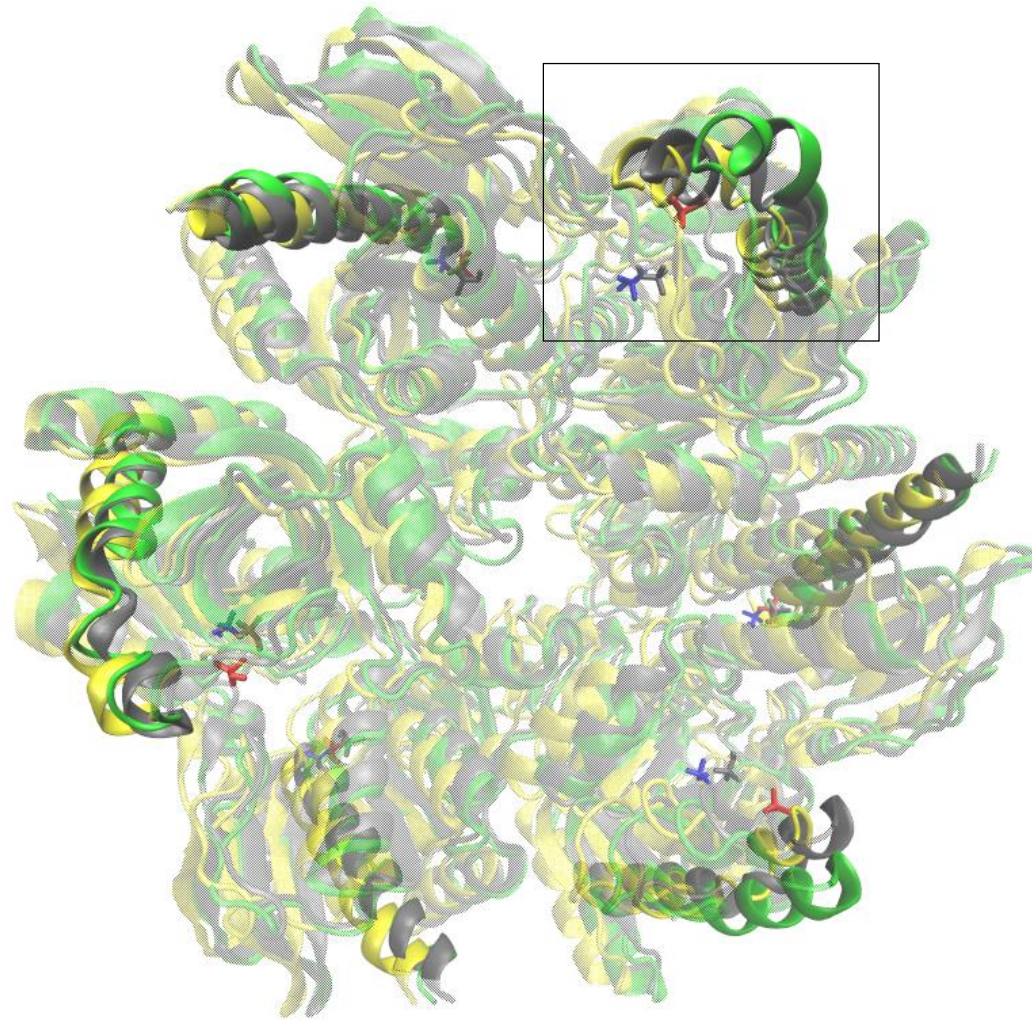
Fluctuations of R24A PNP binary complex with phosphate



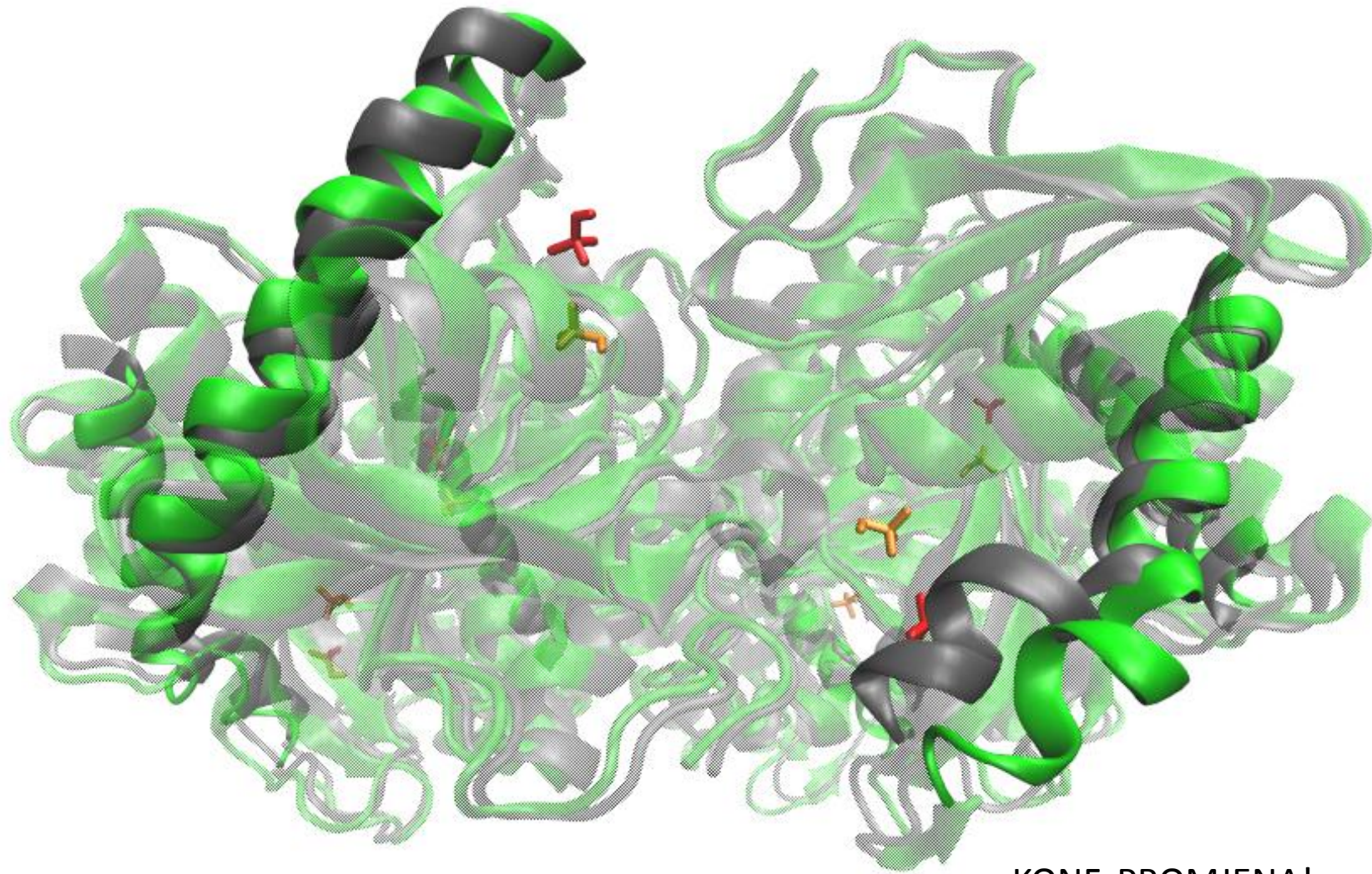
Amino kiseline - 212-222  
(212-222, 449-459, 686-696)

WT – crno  
R24A - crveno

BINARNI KOMPLEKS (zeleno R24A, žuto WT, sivo kristalna struktura)



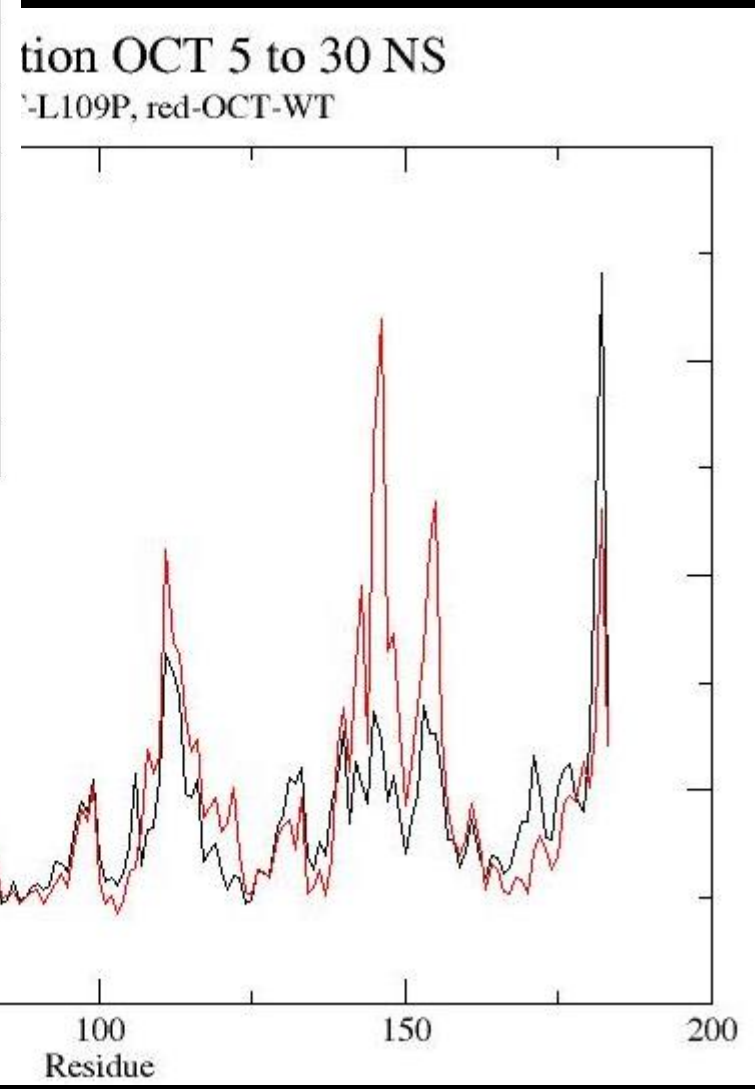
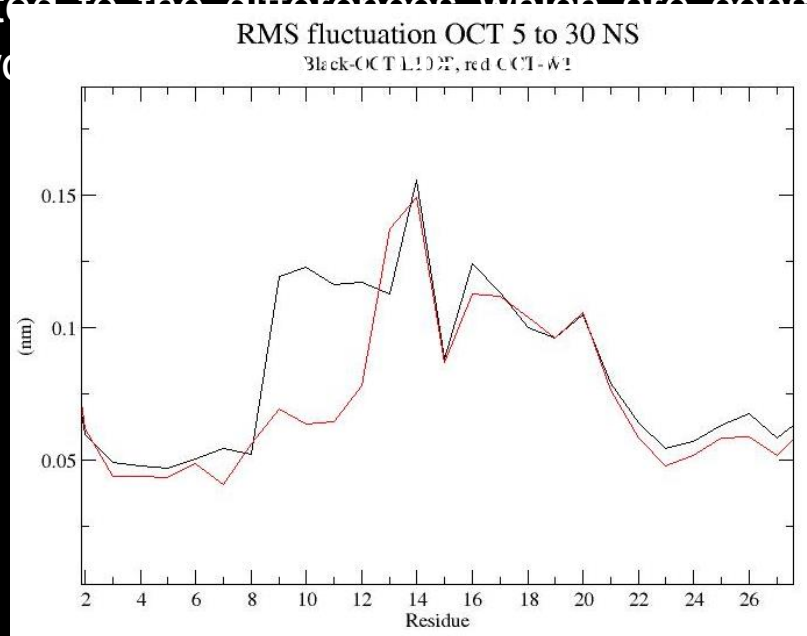
BINARNI KOMPLEKS (zeleno R24A, sivo kristalna struktura)



KONF. PROMJENA!



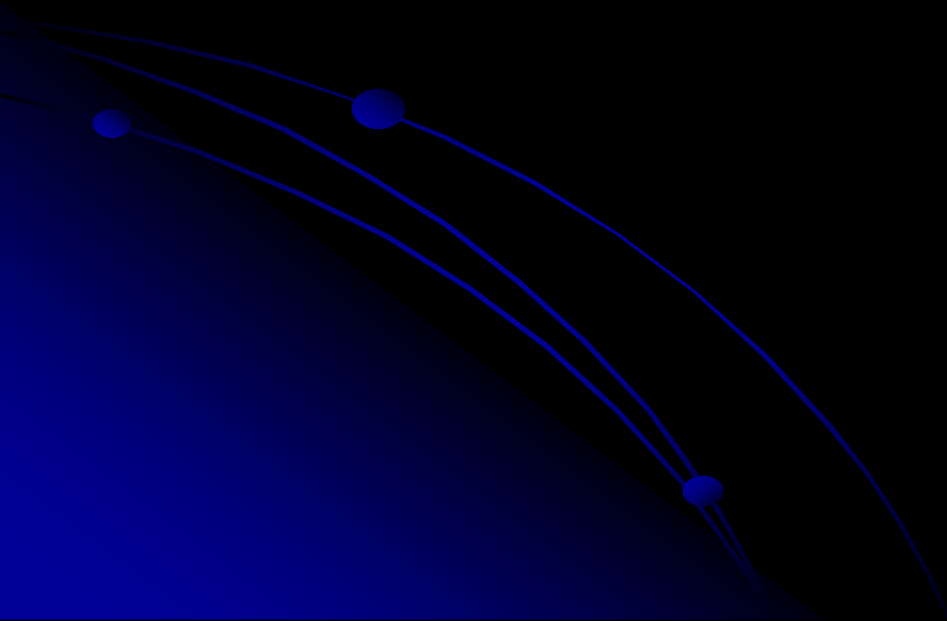
Comparison of RMSF of these two systems (OCT-L109P-black, OCT-WT-red) pointed to the differences which are consequence of differences in Trp23 H-bond network



PRIMJER: RMSF je statistički potvrdio, odnosno kvantizirao rezultate vizualizacije (vizualno je uočena razlika koja je fluktuacijama i potvrđena, odnosno kvantizirana.)

# MJERENJE UDALJENOSTI

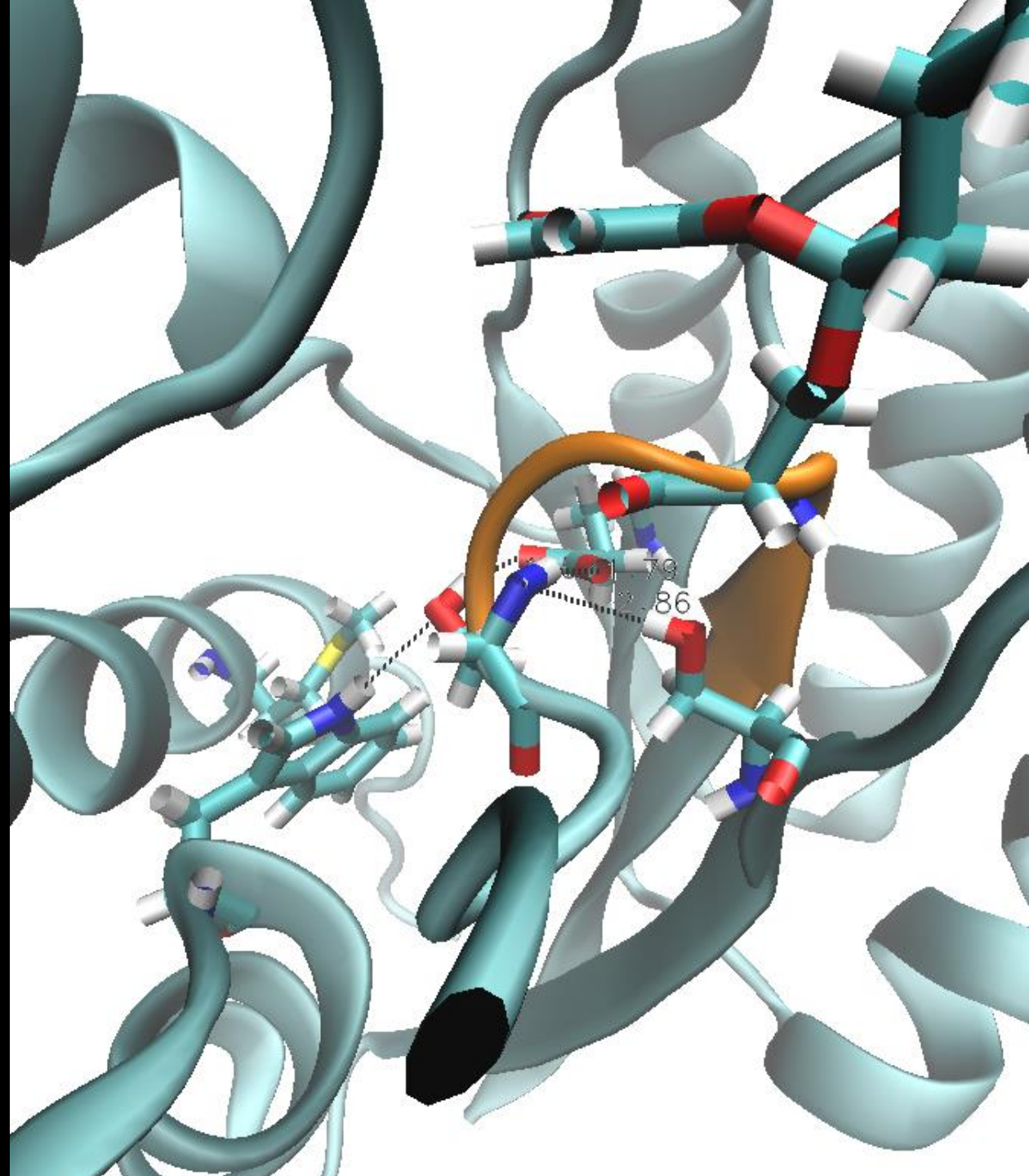
Mjerenje udaljenosti između dva atoma ili skupina atoma te praćenje te udaljenosti tijekom simulacije, često je prvi oblik kvantizacije rezultata vizualizacije. U većini programa za vizualizaciju jednostavno je izmjeriti udaljenost dva atoma i pratiti vrijednost te udaljenosti tijekom trajektorije. Za mjerenje udaljenosti centara masa dviju skupina atoma, koristi se naredba `g_dist` u programskom paketu Gromacs.

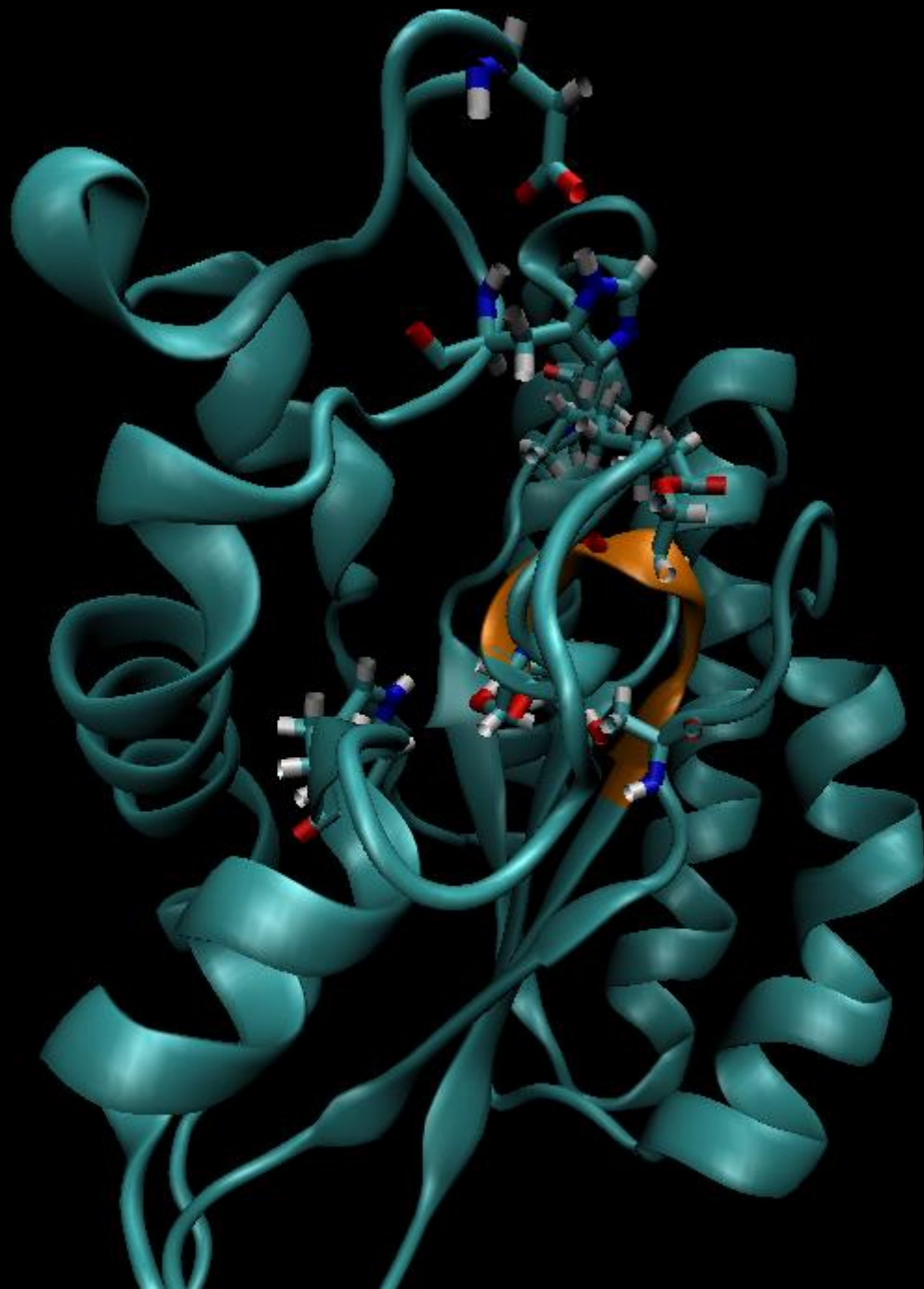


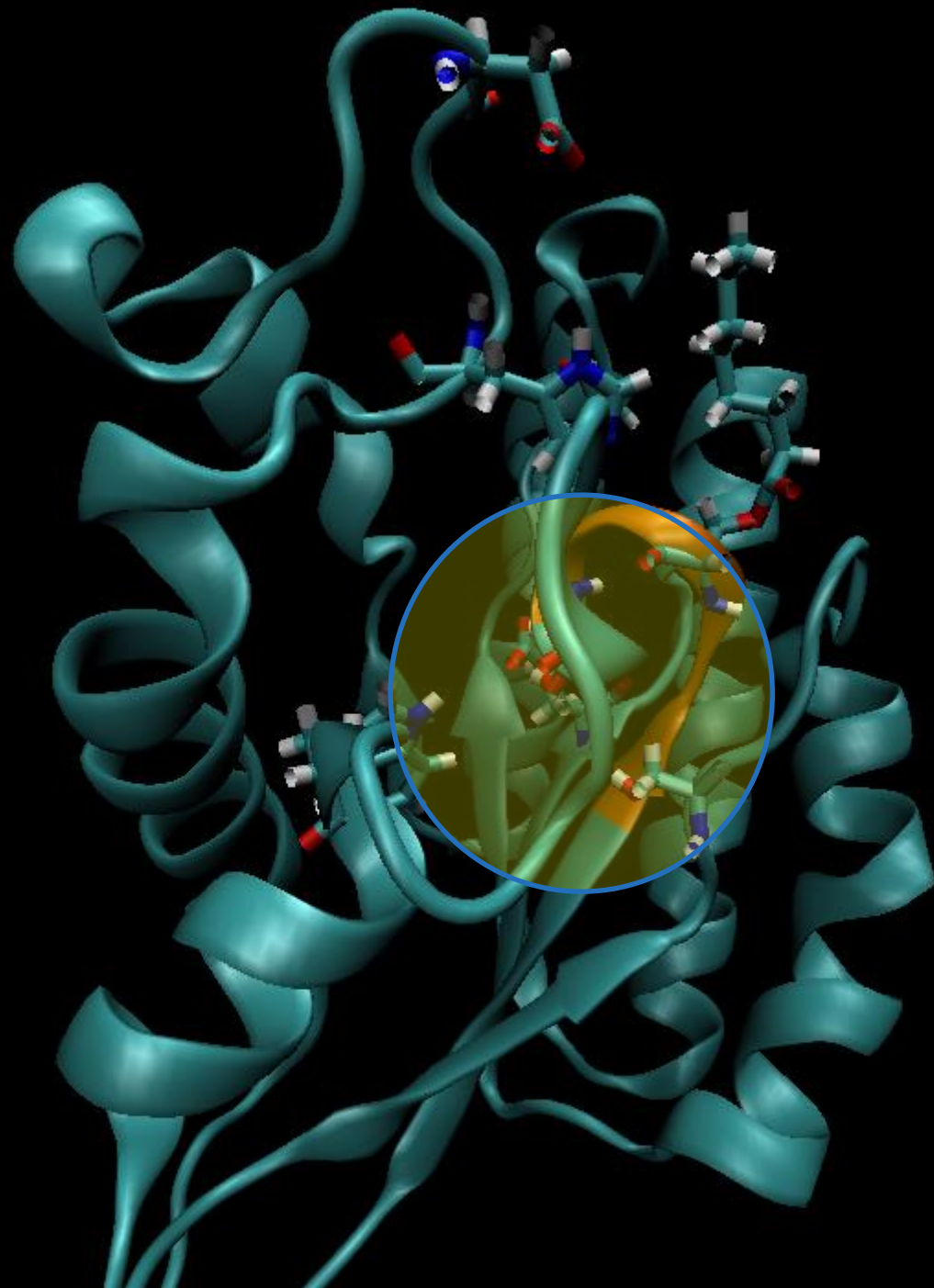
## PRIMJER KVANTIZACIJE REZULTATA VIZUALIZACIJE

### MJERENJEM UDALJENOSTI:

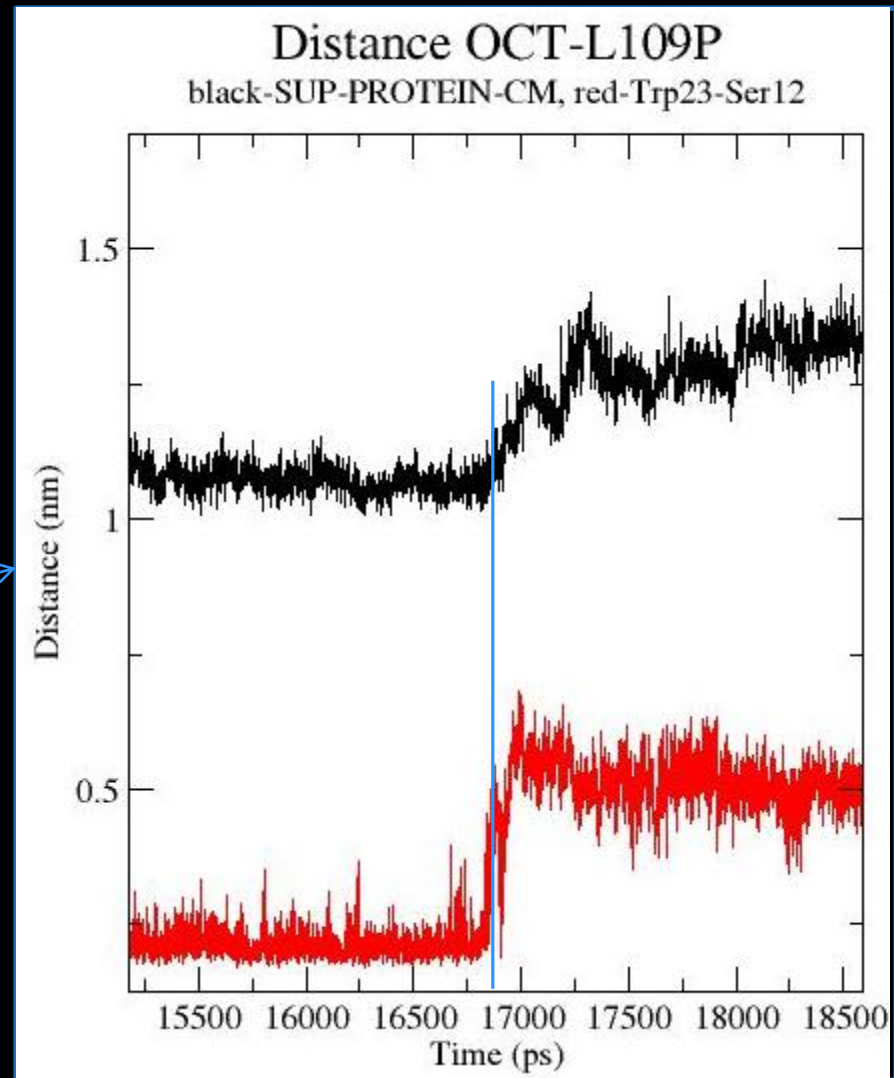
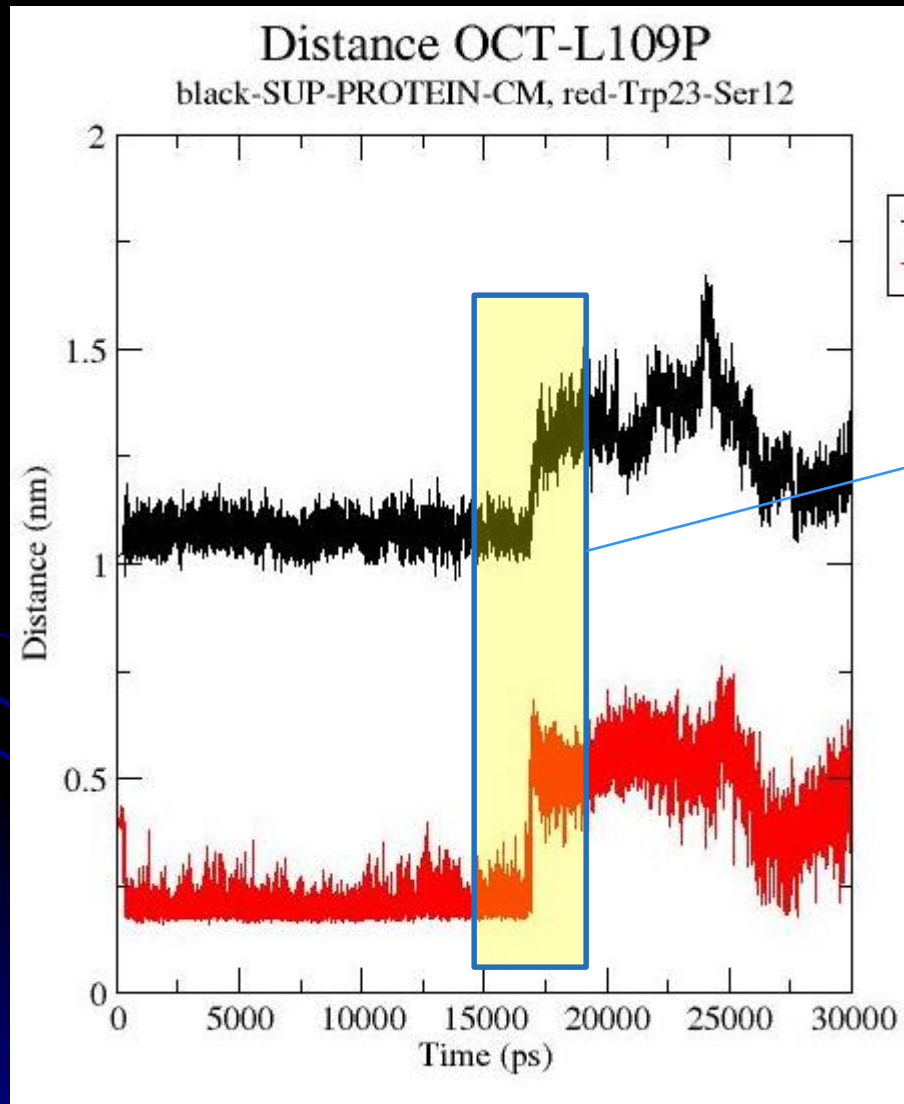
- uočeno je da izlaz supstrata iz veznog mjesta prati i pucanje mreže vodikovih veza
- postavilo se pitanje: je li pucanje mreže vodikovih veza uzrok ili posljedica izlaska supstrata iz veznog mjesta
- odgovor je nađen mjerenjem udaljenosti: 1) centra mase supstrata i centra mase proteina, i 2) ključne vodikove veze u spomenutoj mreži vodikovih veza







movement of the substrate occurs **immediately AFTER** the Trp23 H-bond network breaks!



# RADIJUS GIRACIJE

## RADIJUS GIRACIJE (RG)

- matematički: srednji kvadrat udaljenosti dijelova nekog objekta od centra mase tog objekta (ili, ponekad, osi rotacije)
- u biofizici: srednji kvadrat sume udaljenosti pojedinih monomera (aminokiselina ili nukleotida) od centra mase makromolekule

$$R_g = \left( \frac{\sum_i \|\mathbf{r}_i\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}$$

$r_i$  – udaljenost i-tog monomera (atoma, aminokiseline ili nukleotida) od centra mase makromolekule,  $m_i$  – masa monomera (atoma, aminokiseline ili nukleotida),  $\sum m_i$  – ukupna masa makromolekule

- vrlo korisna procjena volumena makromolekule ili dijela makromolekule u otopini

### • Molecular radius of gyration

$$R_{gyr} = \left[ \frac{1}{N_a} \sum_{i=1}^{N_a} [\vec{r}_i - \vec{R}_{cm}]^2 \right]^{\frac{1}{2}}$$

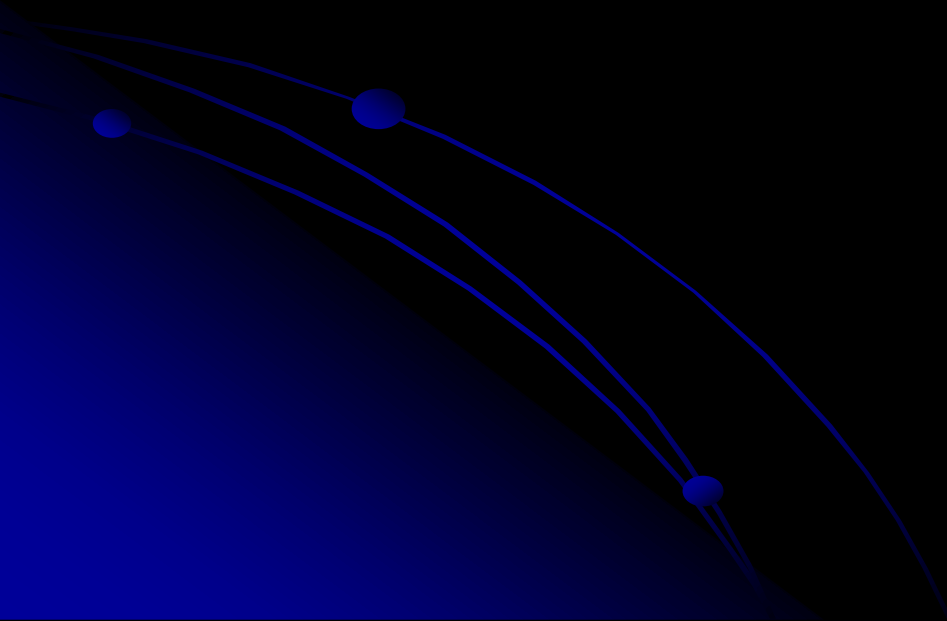
$N_a$  = number of atoms in the molecule

$\vec{R}_{cm}$  = centre of mass of molecule

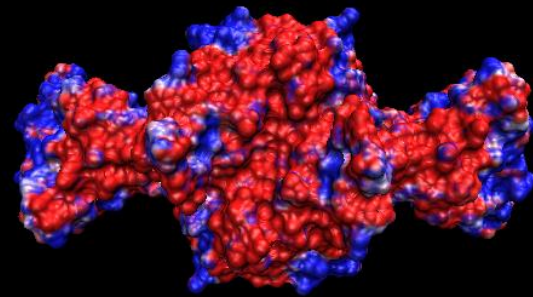
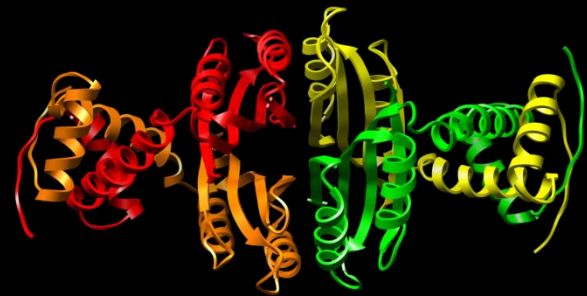
$$= \frac{1}{M} \sum_{i=1}^{N_a} m_i \vec{r}_i$$

$$M = \text{mass of molecule} = \sum_{i=1}^{N_a} m_i$$

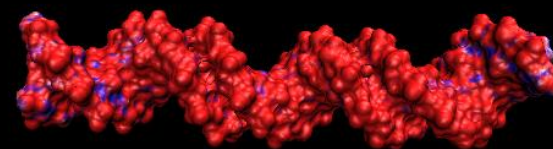
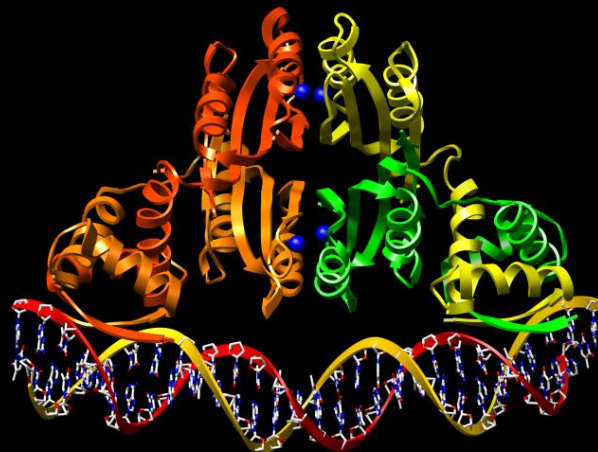
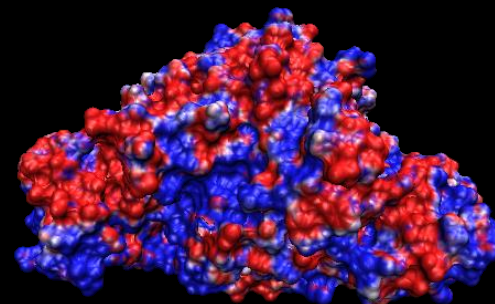
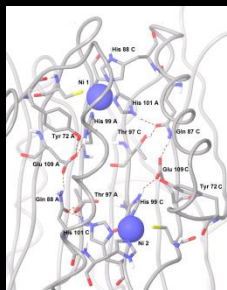
# RADIJUS GIRACIJE - primjer

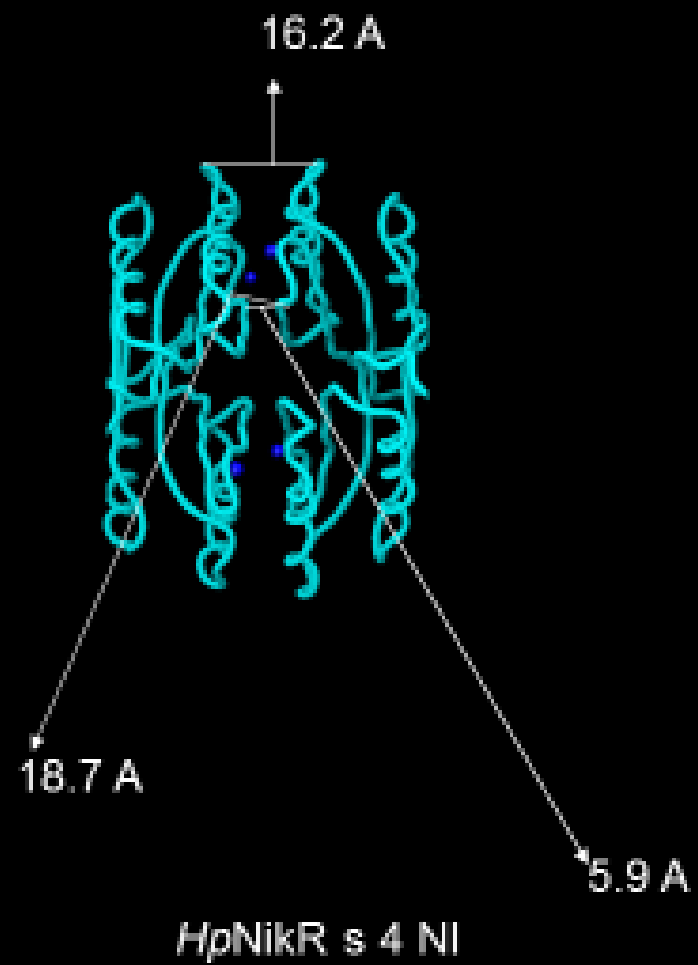
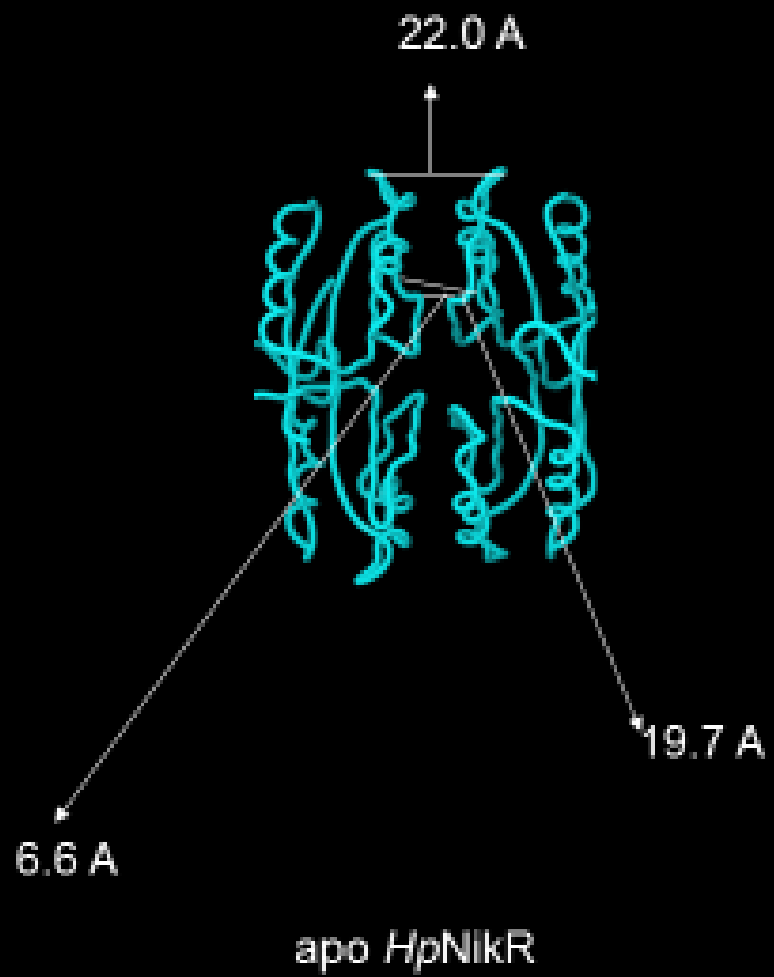


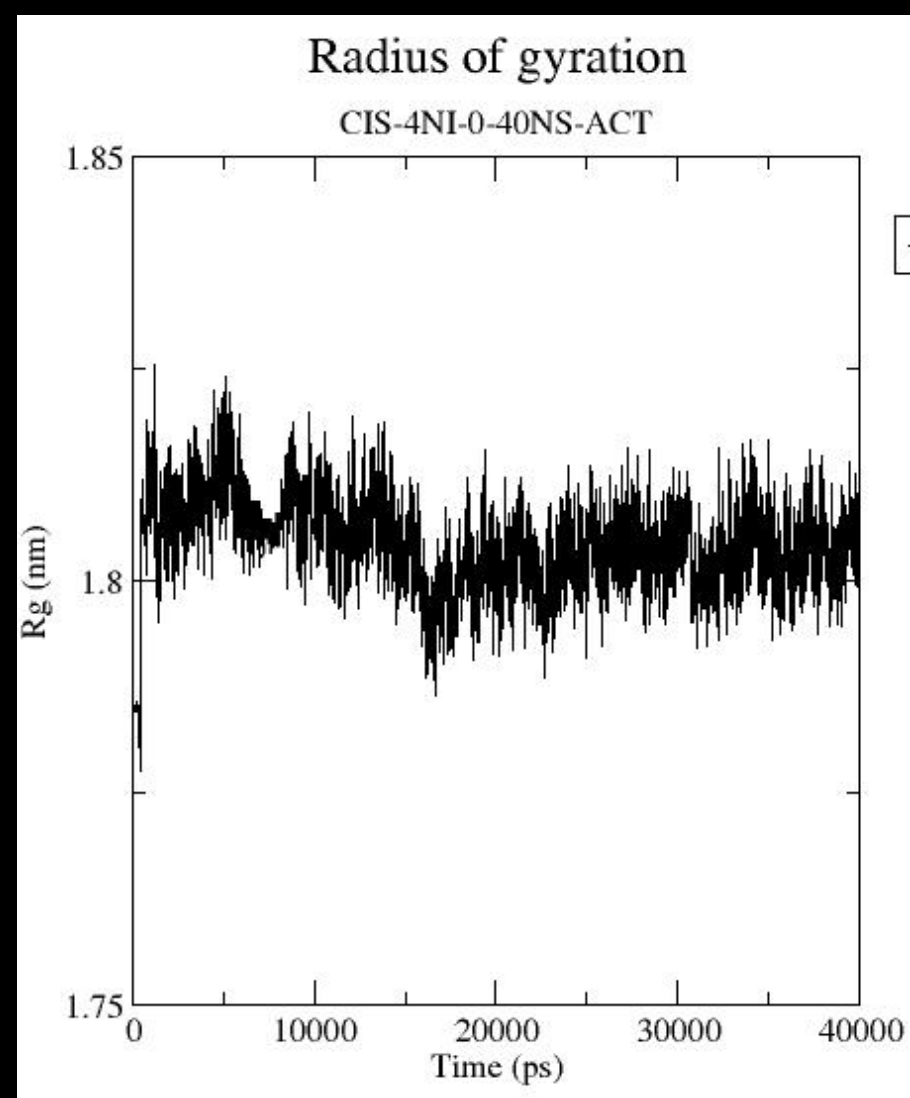
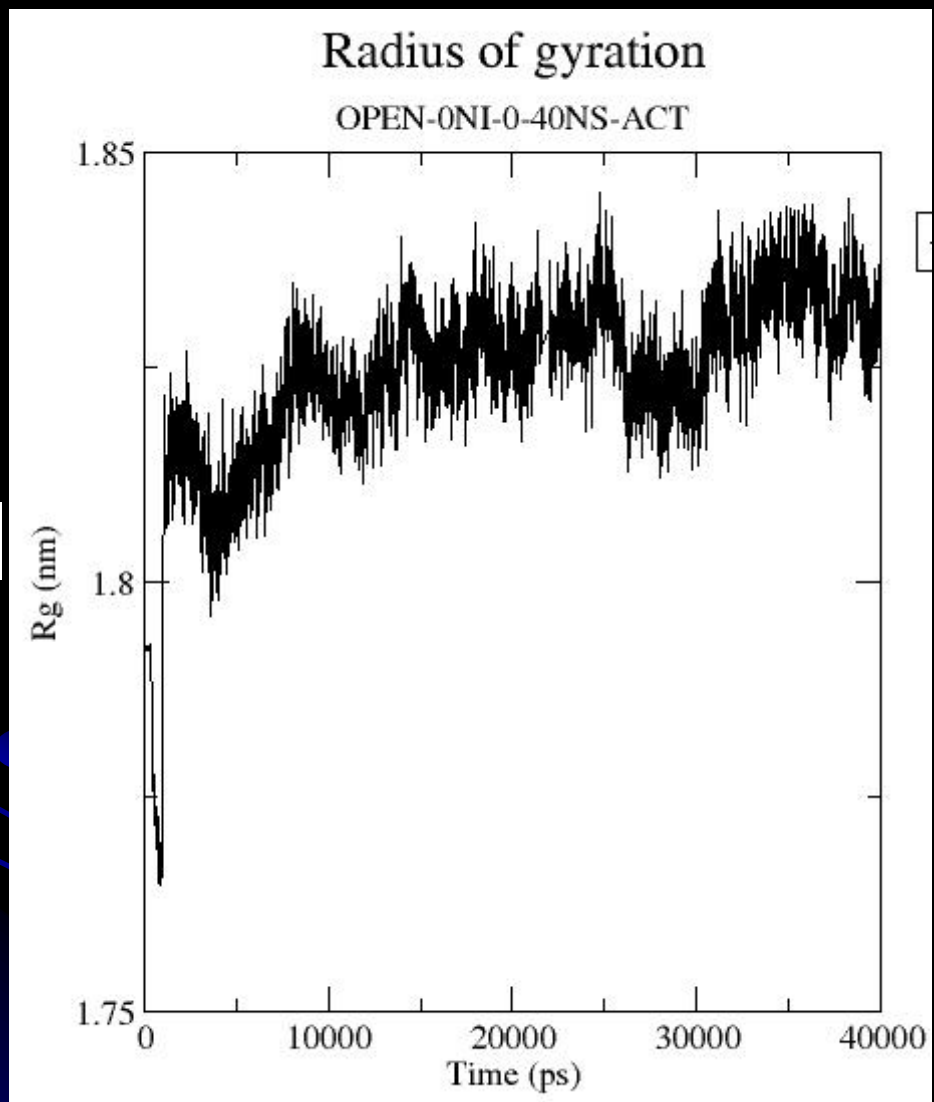




Ni(II)



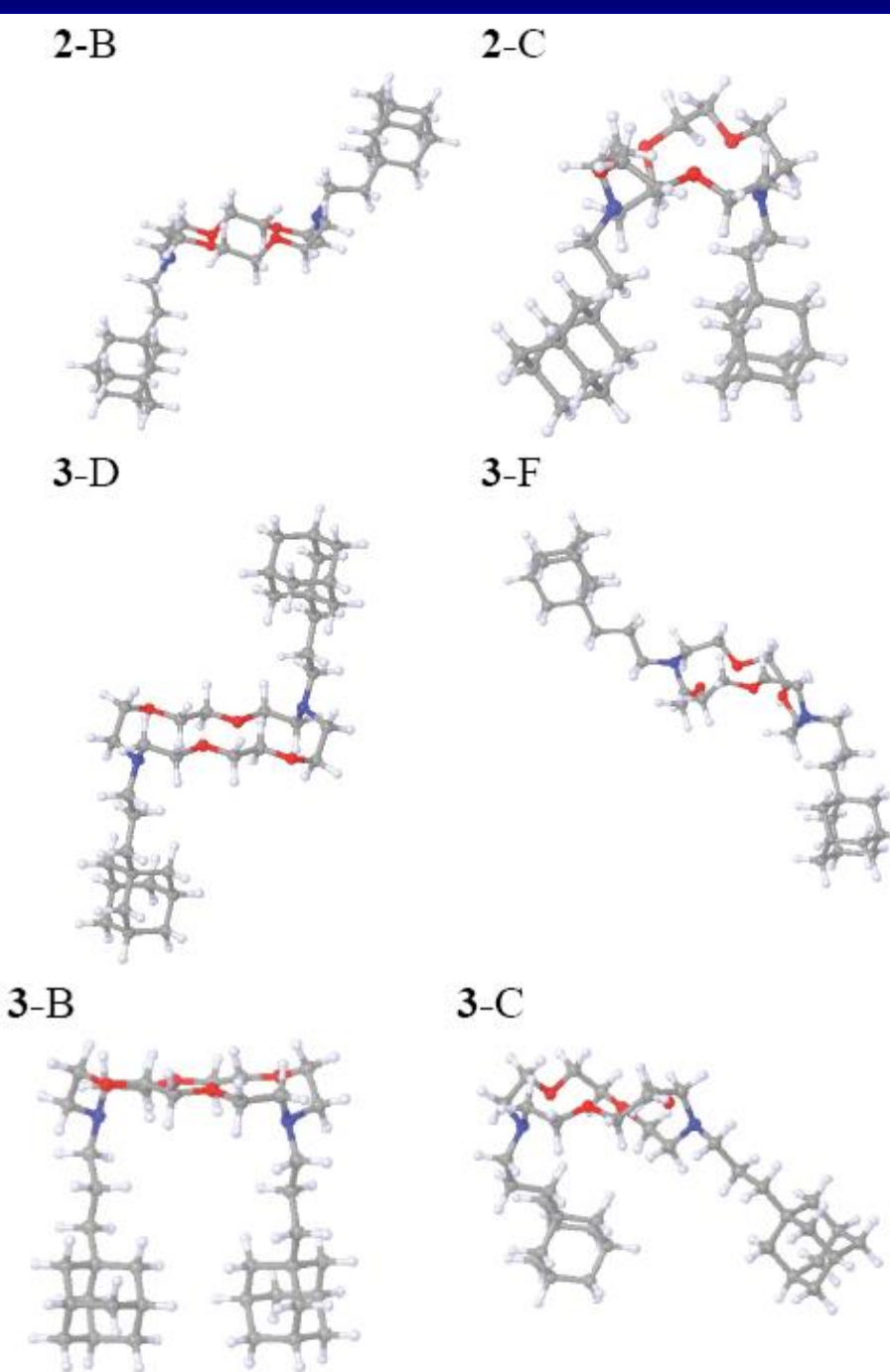




# KLASTER-ANALIZA

- postoje različiti algoritmi koji po različitim kriterijima klasteriraju (grupiraju u klustere) konformacije dobivene MD simulacijom
- Gromsov algoritam klasteriranje radi na temelju RMSD vrijednosti: potrebno je izabrati graničnu RMSD vrijednost (*cutoff*) koja će određivati jesu li dvije konformacije dovoljno slične da budu smještene u isti klaster ili pripadaju različitim klasterima.
- broj klastera, određen pri istoj graničnoj vrijednosti, za dva ili više sustava ukazuje i na fleksibilnost tog sustava (fleksibilniji sustavi imat će veći broj klastera)
- predstavnici pojedinih klastera se uspoređuju u svrhu analize trajektorije

# KLASTER-ANALIZA - primjer



**Table 2.** Monte Carlo conformational search of compounds 1–6 and referent compounds 7–8

Crown ethers	NS <sup>(a)</sup> within 10 kJ/mol	NS <sup>(b)</sup> within 20 kJ/mol	<i>N</i> of clusters <sup>(c)</sup>	Conform. represented by each cluster (%) <sup>(d)</sup>	Conf. energy <sup>(e)</sup> (kJ/mol)	Solv. energy <sup>(f)</sup> (kJ/mol)
1	46	513	2	1-A 89 %	482.5	-55.2
				1-B 11 %		
2	178	1517	3	2-A 75 %	477.7	-66.8
				2-B 24 %		
				2-C 1 %		
3	113	878	6	3-A 43 %	479.9	-65.6
				3-B 25 %		
				3-C 15 %		
				3-D 12 %		
				3-E 3 %		
				3-F 2 %		
4	6	160	3	4-A 66 %	495.3	-67.2
				4-B 17 %		
				4-C 17 %		
5	44	652	3	5-A 68 %	461.4	-72.4
				5-B 27 %		
				5-C 5 %		
6	43	581	3	6-A 58 %	457.0	-77.4
				6-B 30 %		
				6-C 12 %		
7	1	20	1	7-A 100 %	266.4	-40.3
				8-A 90 %		
8	94	1214	3	8-B 9 %	314.12	-69.1
				8-C 1 %		

<sup>(a)</sup> Number of structures found within the energy window of 10 kJ/mol.

<sup>(b)</sup> Number of structures found within the energy window of 20 kJ/mol.

<sup>(c)</sup> Number of clusters found by the cluster analysis of the structures found within the energy window of 10 kJ/mol.

<sup>(d)</sup> Percentage of conformations found within the energy window of 10 kJ/mol that are represented by each cluster; cluster structures have the same names as on the Figure S1 in Supplementary Materials.

<sup>(e)</sup> Force field energy of the energetically most favorable conformation found by the MC search.

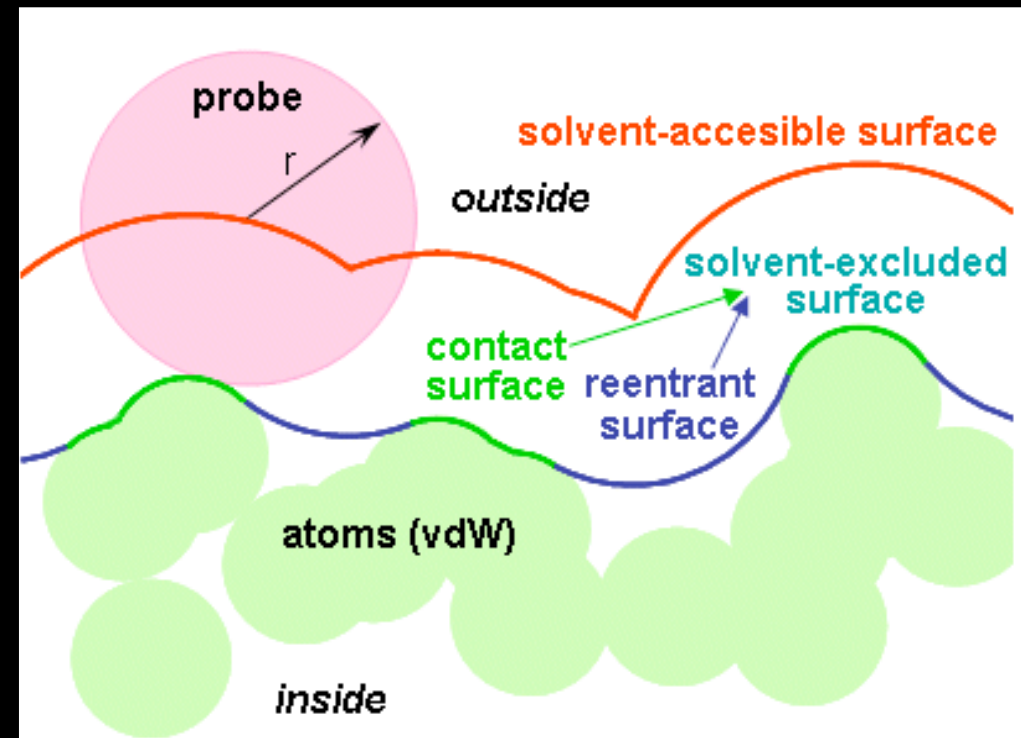
<sup>(f)</sup> Solvation energy in chloroform for the energetically most favorable conformation.

# OTAPALU DOSTUPNA POVRŠINA

- korištenjem probe definiranog promjera lako se može izračunati otapalu dostupna površina makromolekule
- iznos polumjera probe određuje preciznost računanja površine (manja proba znači veću preciznost)

- **SAS** – Solvent Accesable Surface

- te **ASP** – Atomic Solvation Parameter – određuje se na temelju particijskog koeficijenta, predstavlja energiju za prijenos površine atoma ili skupine u vodu iz nekog drugog medija



Based on <http://geometry.molmovdb.org/3v/>  
and <http://www.netsci.org/Science/Compchem/feature14e.html>

# POISSON-BOLTZMANNANOVA JEDNADŽBA ZA RAČUNANJE ELEKTROSTATSKOG POTENCIJALA

$$V_e = \frac{Z_1 Z_2 e^2}{D r}$$

- Coulombov izraz

- problem **ZASIJENJENJA** zbog prisutnosti ostalih specija između  $Z_1$  i  $Z_2$

POISSON-BOLTZMANOVA JEDNADŽBA riješava problem zasjenjenja i omogućava računanje elektrostatskog potencijala u svakoj točki sustava.

$$\nabla D(r) \nabla V_e(r) - D(r) k^2 \sinh[V_e(r)] + \frac{4\pi Z e \rho(r)}{k_B T} = 0$$

$\nabla$  - gradient (vektorska derivacija u odnosu na vektor  $r$ )

$V_e(r)$  – elektrostatski potencijal ovisan o  $r$

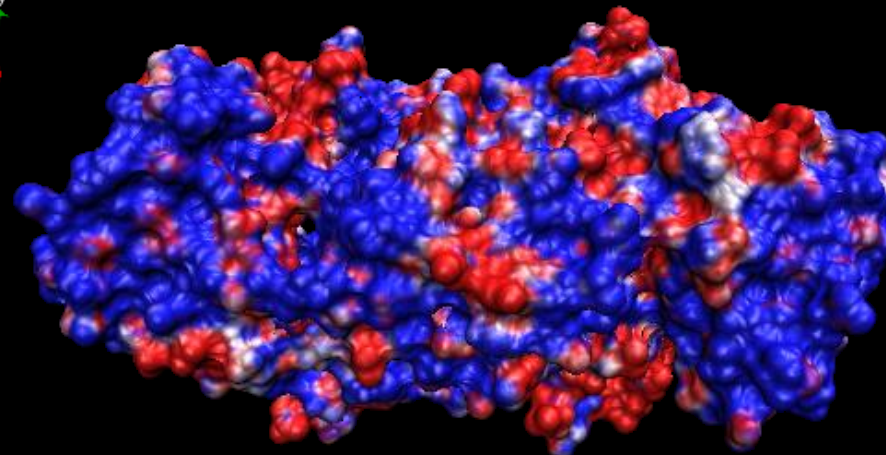
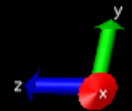
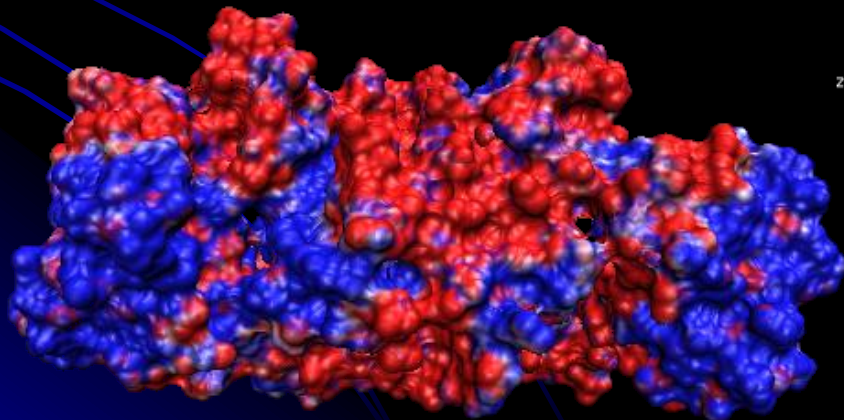
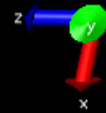
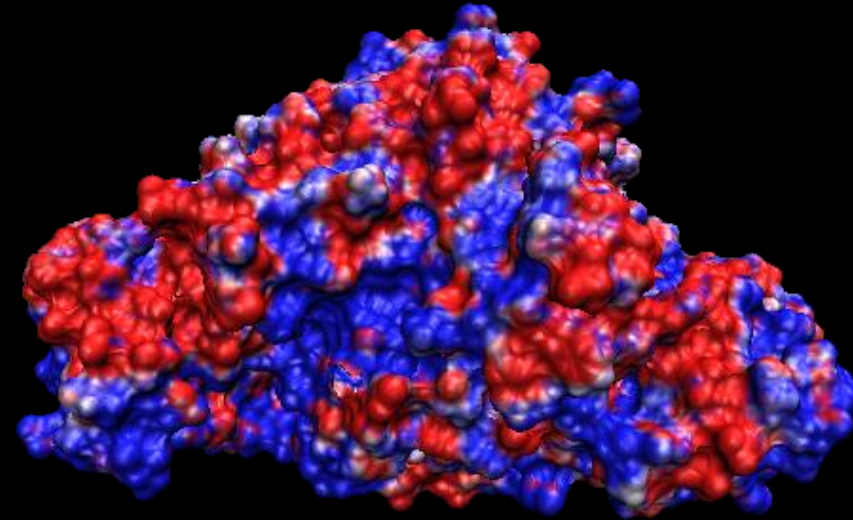
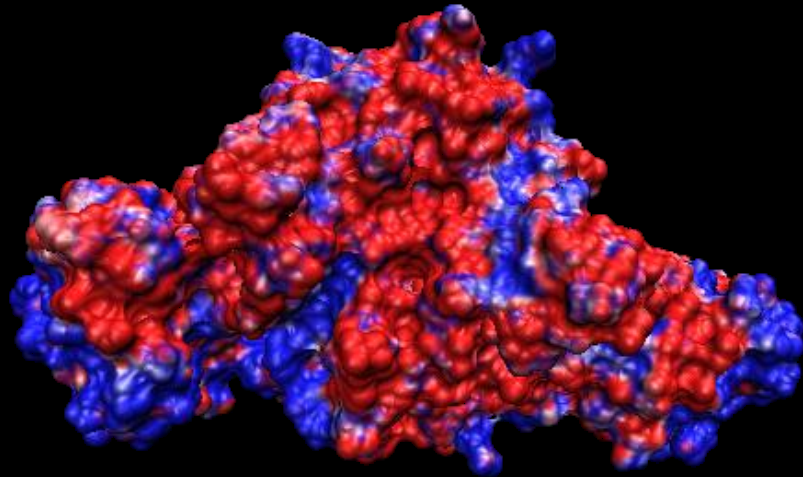
$D(r)$  – dielektrična konstanta ovisna o  $r$

$\rho(r)$  – gustoća naboja

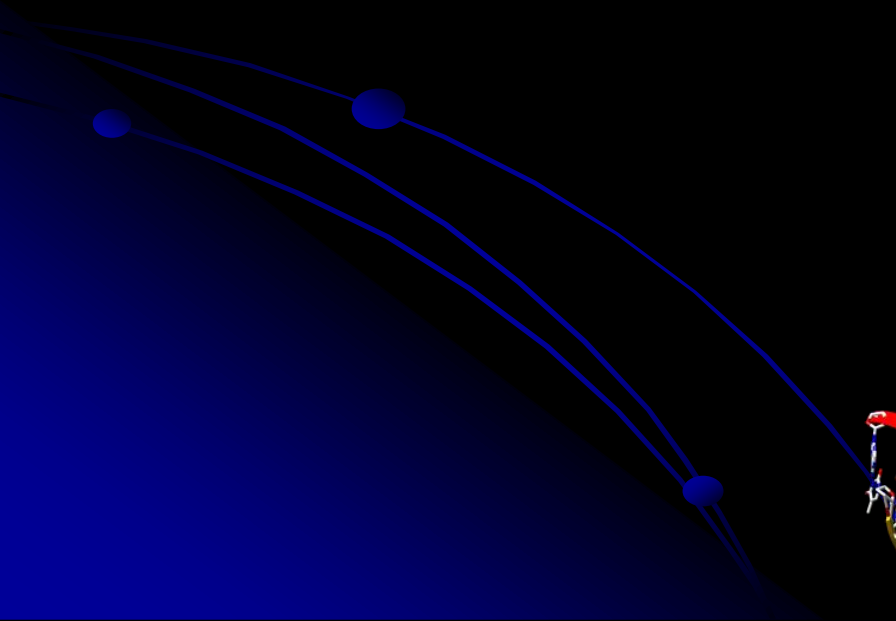
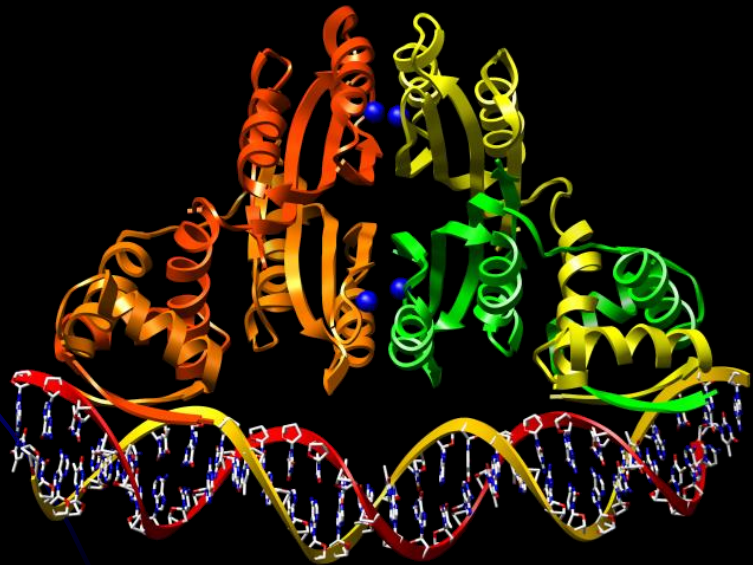
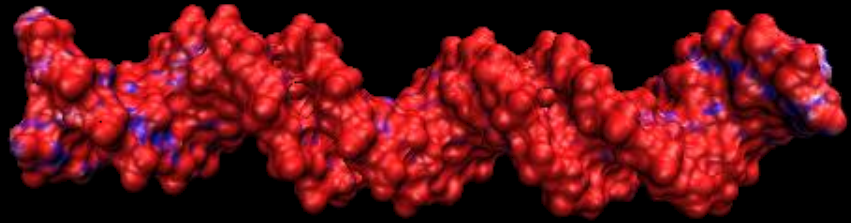
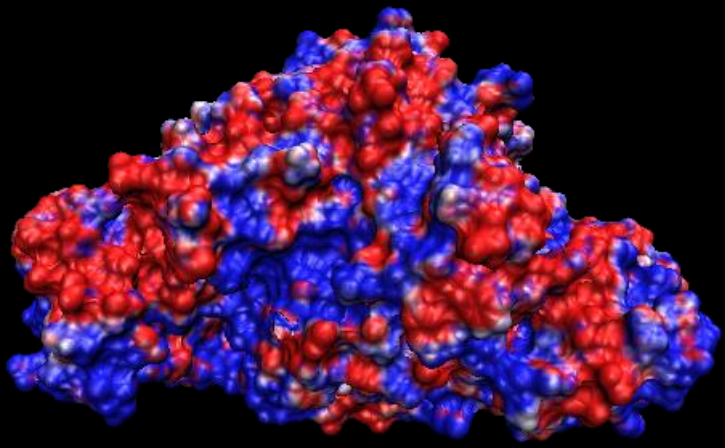
$$K = \frac{8\pi(Ze)^2 I}{D(r) k_B T}$$

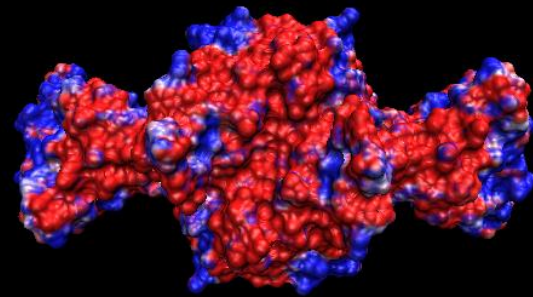
$I$  – ionska jakost,  $I = \sum_{i=1}^N \frac{Z_i^2 C_i}{2}$

# PRIMJER RAČUNANJA ELEKTROSTATSKOG POTENCIJALA POISSON-BOLTZMANNOVOM JEDNADŽBOM

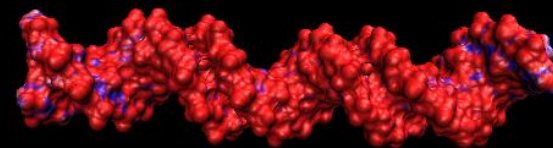
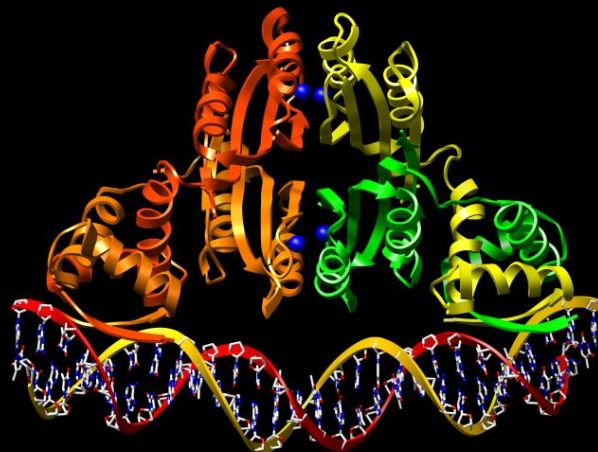
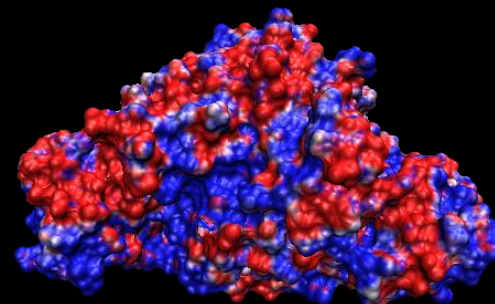
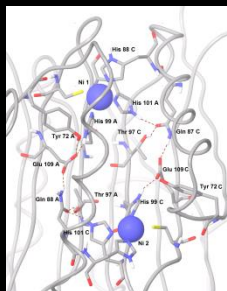








Ni(II)



# OSTALI ALATI ZA ANALIZU TRAJEKTORIJA

- postoji jako velik broj opcija koje su još na raspolaganju, poput analize vodikovih veza, analize udjela sekundarnih struktura i sl.
- popis svih opcija raspoloživih u Gromacs paketu može se naći na poveznici:

<http://manual.gromacs.org/programs/byname.html>

## ZAKLJUČAK

Analiza trajektorije je poput igre detekcije: eksperimentalni podatci, literaturni podatci, iskustvo i sl. nam predstavljaju tragove, a upornost, marljivost, temeljitost i biokemijska intuicija omogućavaju nam dolazak do rezultata. Vizualizacija predstavlja nezamjenjivi dio analize koji se mora potkrijepiti kvantitativnim analizama trajektorije.