



Sveučilište u Zagrebu



Prirodoslovno-matematički fakultet



Kemijski odsjek

Kemijski seminar I

Izrađen prema:

H. Wang, H. Liu, L. Cai, C. Wang, and Q. Lv, “Using the multi-objective optimization replica exchange Monte Carlo enhanced sampling method for protein-small molecule docking,” *BMC Bioinform*, 18(1), 1–21 (2017).

MONTE CARLO PRISTUP MOLEKULARNOM DOKINGU

Ana Mikelić

Studentica 1. godine Poslijediplomskog sveučilišnog studija KEMIJA,
smjer FIZIKALNA KEMIJA

Zagreb, 2020.

1. Uvod

2. Monte Carlo metode

1. **MC** – jednostavne Monte Carlo metode
2. **REMC** - Monte Carlo metode uz zamjenu replika
3. **MO-REMC** - Monte Carlo metode
uz višekriterijsku optimizaciju zamjene replika
4. **HMO-REMC** – hibridna Monte Carlo metoda
uz višekriterijsku optimizaciju zamjene replika

3. Usporedba odabranih Monte Carlo metoda - RosettaLigand

TScore

IFDelta

Lrmsd

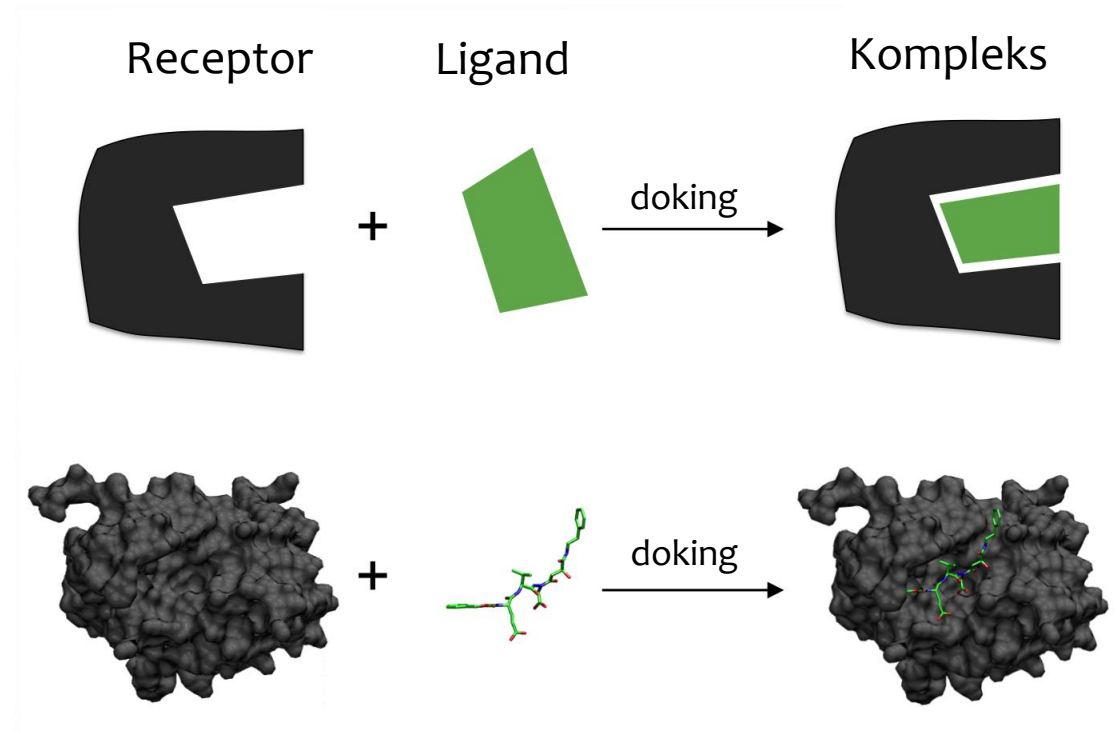
- **Molekularni doking** ⇒ računalne tehnike kojima se predviđaju načini vezanja liganda za receptor, najčešće supstrata u aktivno mjesto enzima

- ✓ 3D-struktura receptora

- ? Energija interakcije s ligandom

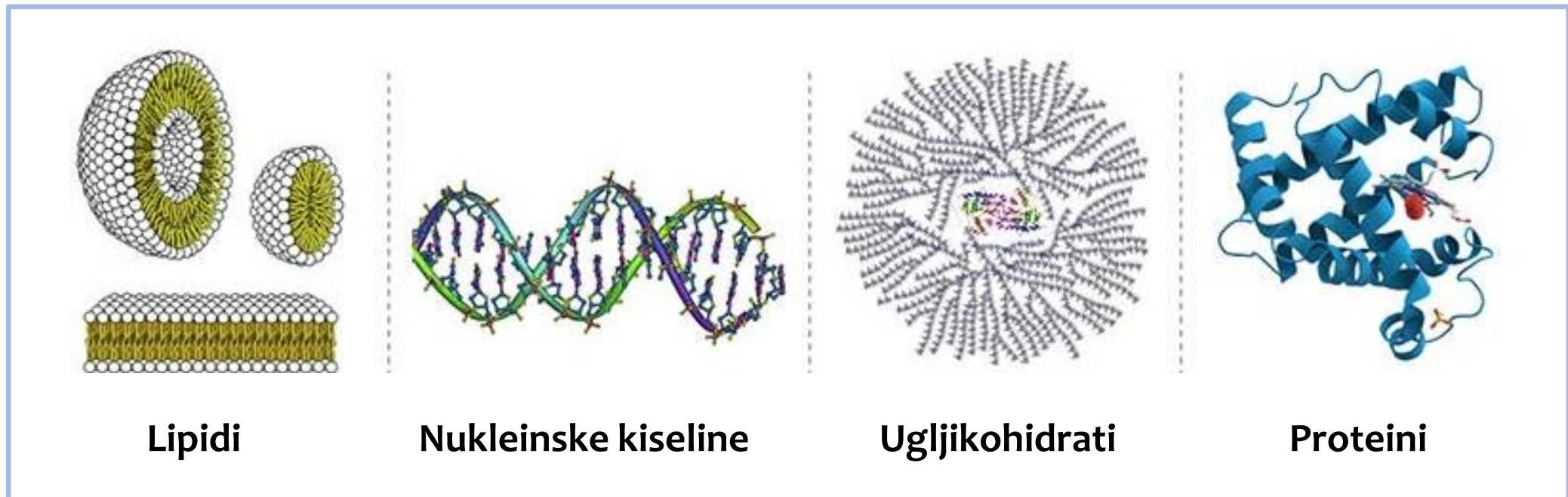
→ Provedi se u 2 koraka:

1. Generiranje svih mogućih konfiguracija
2. Rangiranje na temelju pridružene funkcije bodovanja (engl. *scoring function*) i izdvajanje onih najrelevantnijih



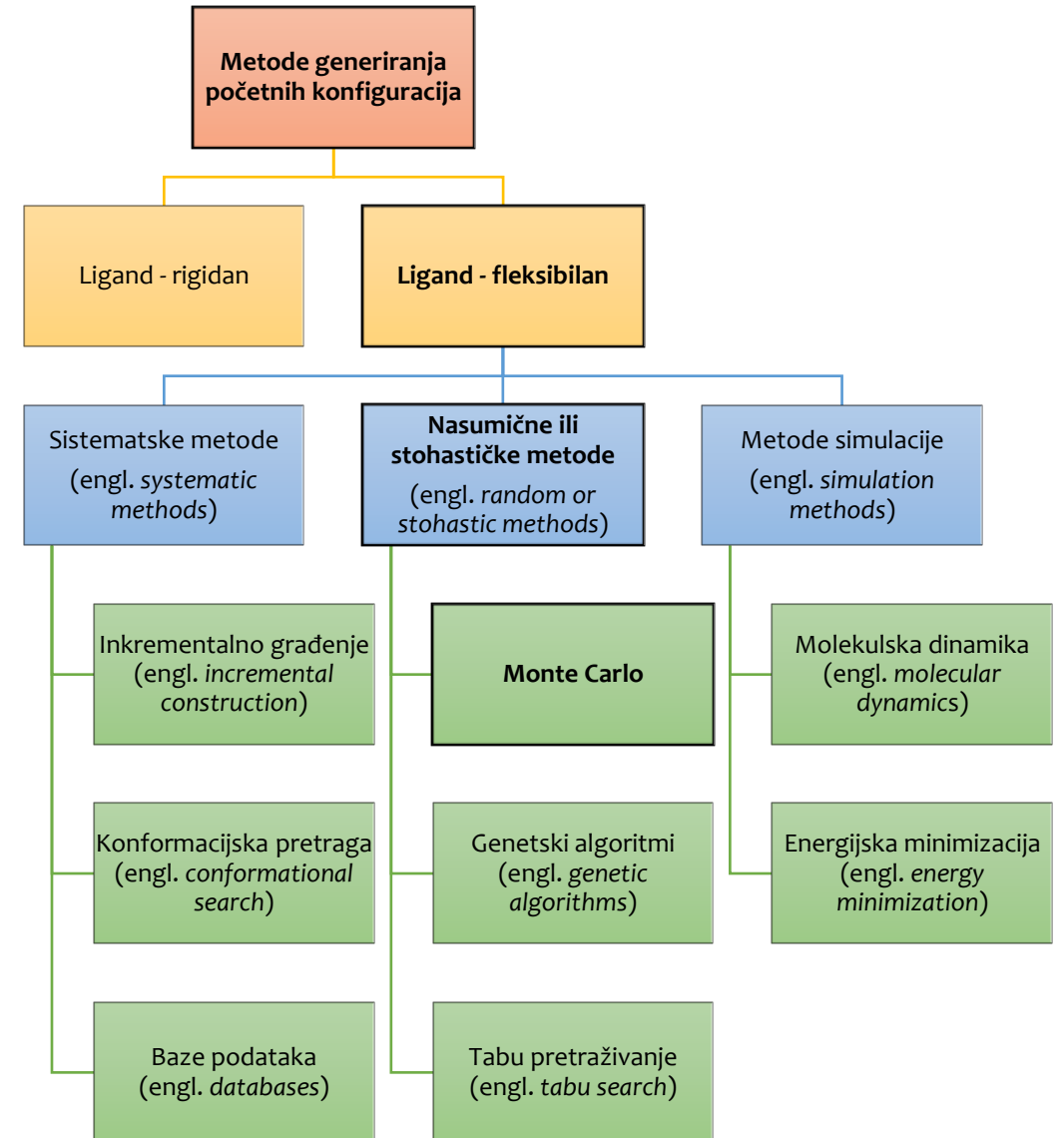
[https://en.wikipedia.org/wiki/Docking_\(molecular\)](https://en.wikipedia.org/wiki/Docking_(molecular))

- Interakcije bioloških makromolekula
 - Fundamentalna uloga
 - Istraživanje lijekova



Prvi korak: Generiranje svih mogućih konfiguracija

- Problem:
 - 6 stupnjeva slobode translacije i rotacije
 - Konformacijski stupnjevi slobode
 - Aproksimacije → različiti algoritmi za doking

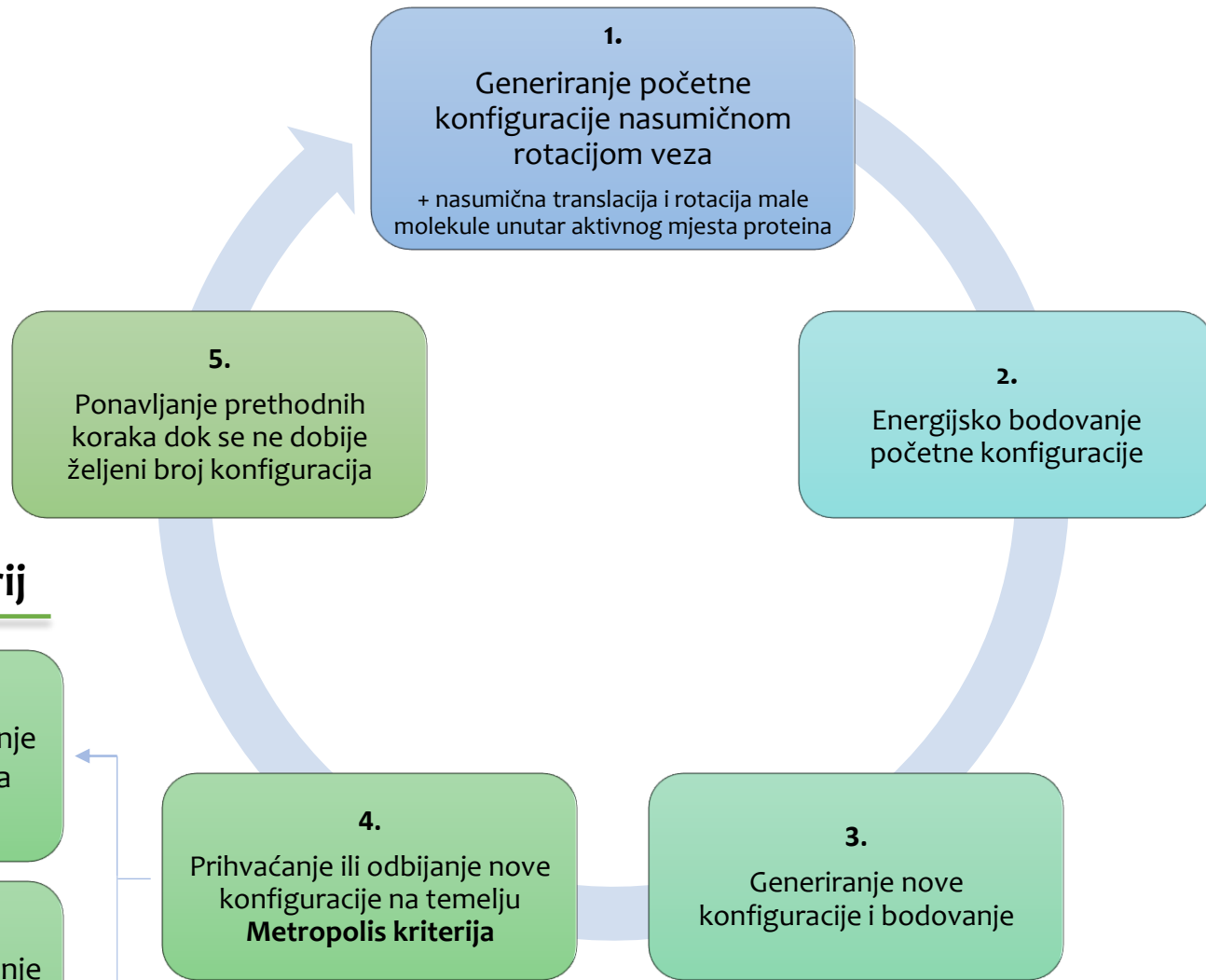
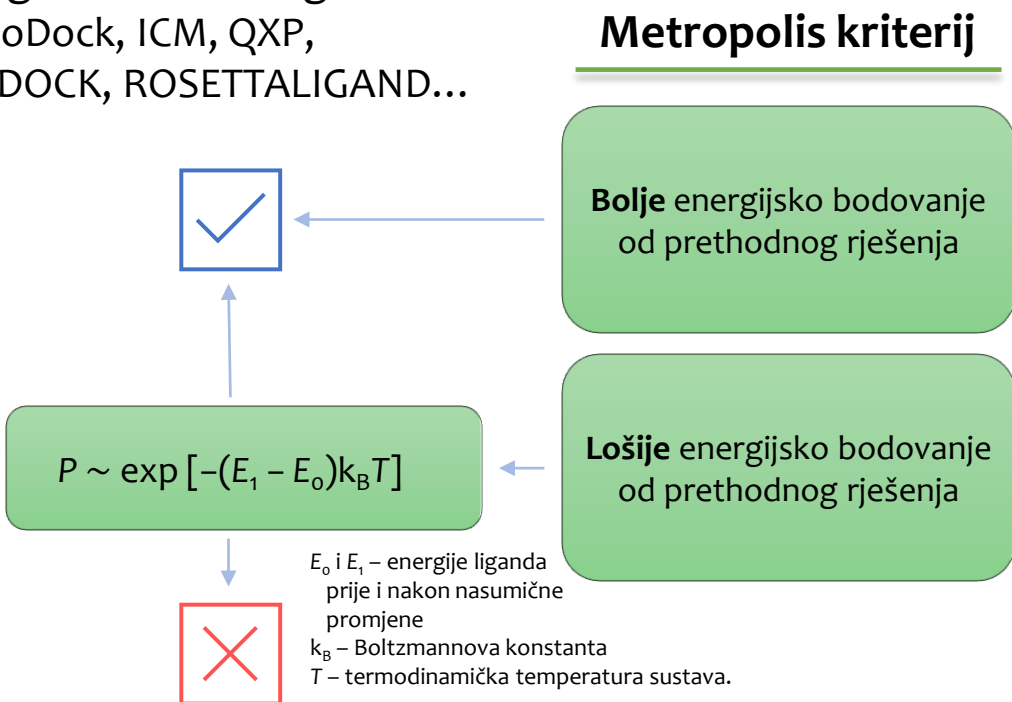


Jednostavne Monte Carlo metode (MC)

- Metropolis MC algoritam
- Markovljev lanac stanja:

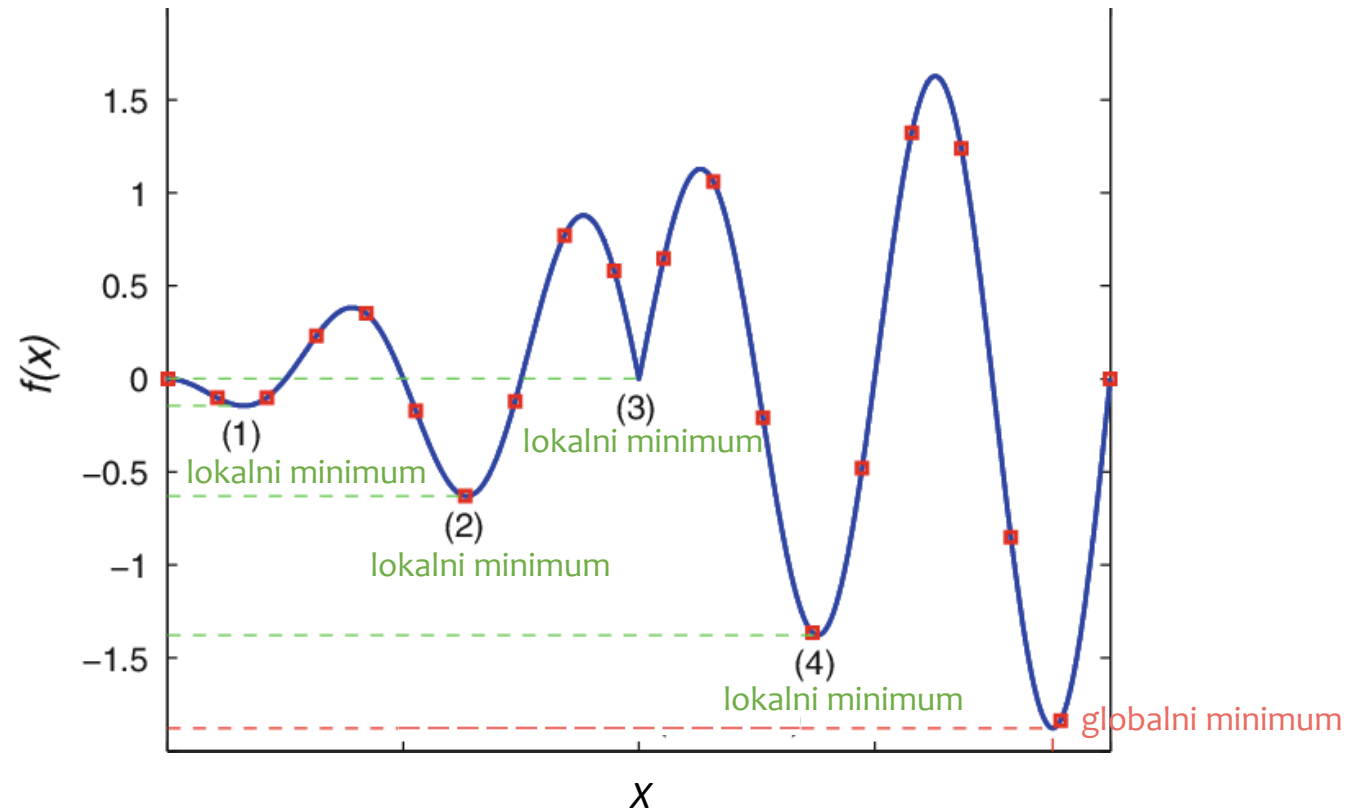
1. Rezultat svakog novog pokušaja ovisi isključivo o prethodnom pokušaju
2. Svaki pokušaj pripada konačnom skupu mogućih rezultata

→ Programi za doking:
AutoDock, ICM, QXP,
MCDOCK, ROSETTALIGAND...



Monte Carlo metode uz zamjenu replika (REMC)

- Simulacije pri nižim temperaturama → problem višestrukih minimuma (engl. *multiple-minima problem*)
→ Simulacije na generaliziranom ansamblu (engl. *generalized ensemble*) ⇒ stanja utežena umjetnim utežnim faktorom ne-Boltzmannove vjerojatnosti



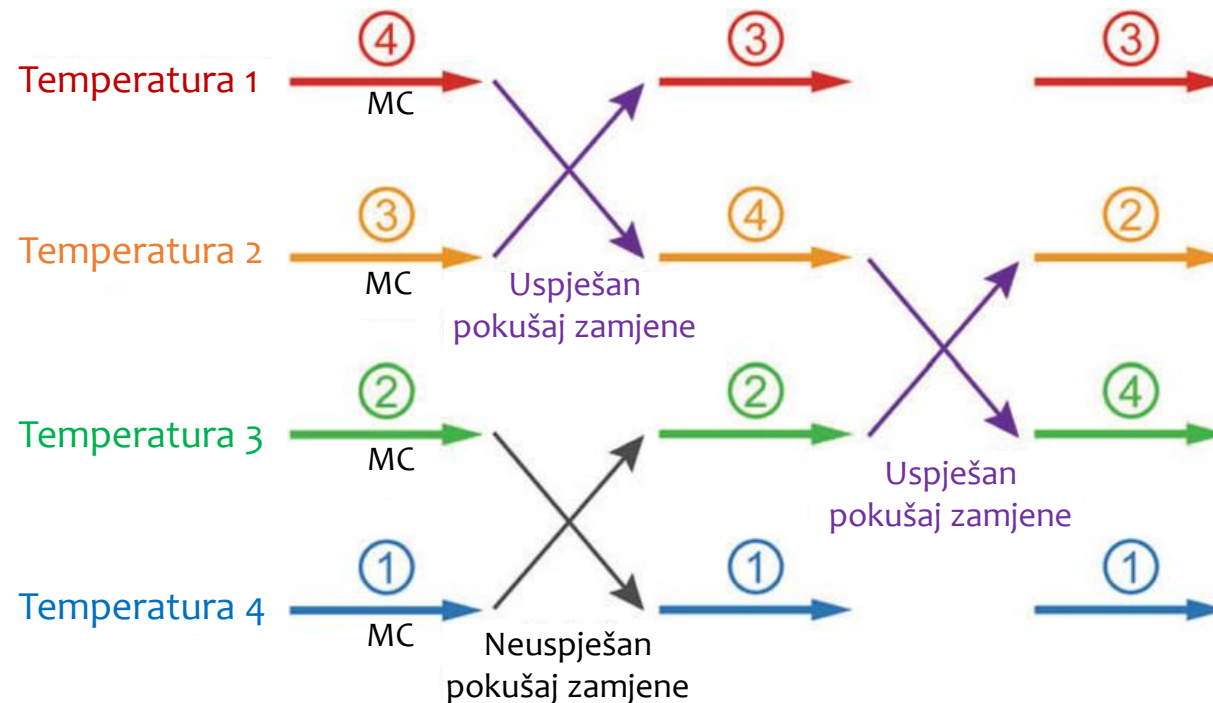
Monte Carlo metode uz zamjenu replika (REMC) (paralelno kaljenje, višestruki Markovljevi lanci)

- **M replika** u kanonskim ansamblima

→ Neovisno i simultano simulira se na **različitim temperaturama** određeni broj MC koraka

→ Nakon **n koraka**, susjedne replike mogu **zamijeniti konfiguracije** na temelju određene vjerojatnosti prijelaza

* Ne postoji općenito pravilo koje bi određivalo optimalan broj koraka n između zamjena



→ Zamjena konformacija niže energije iz visokotemperaturnih replika s onima niskotemperaturnih replika

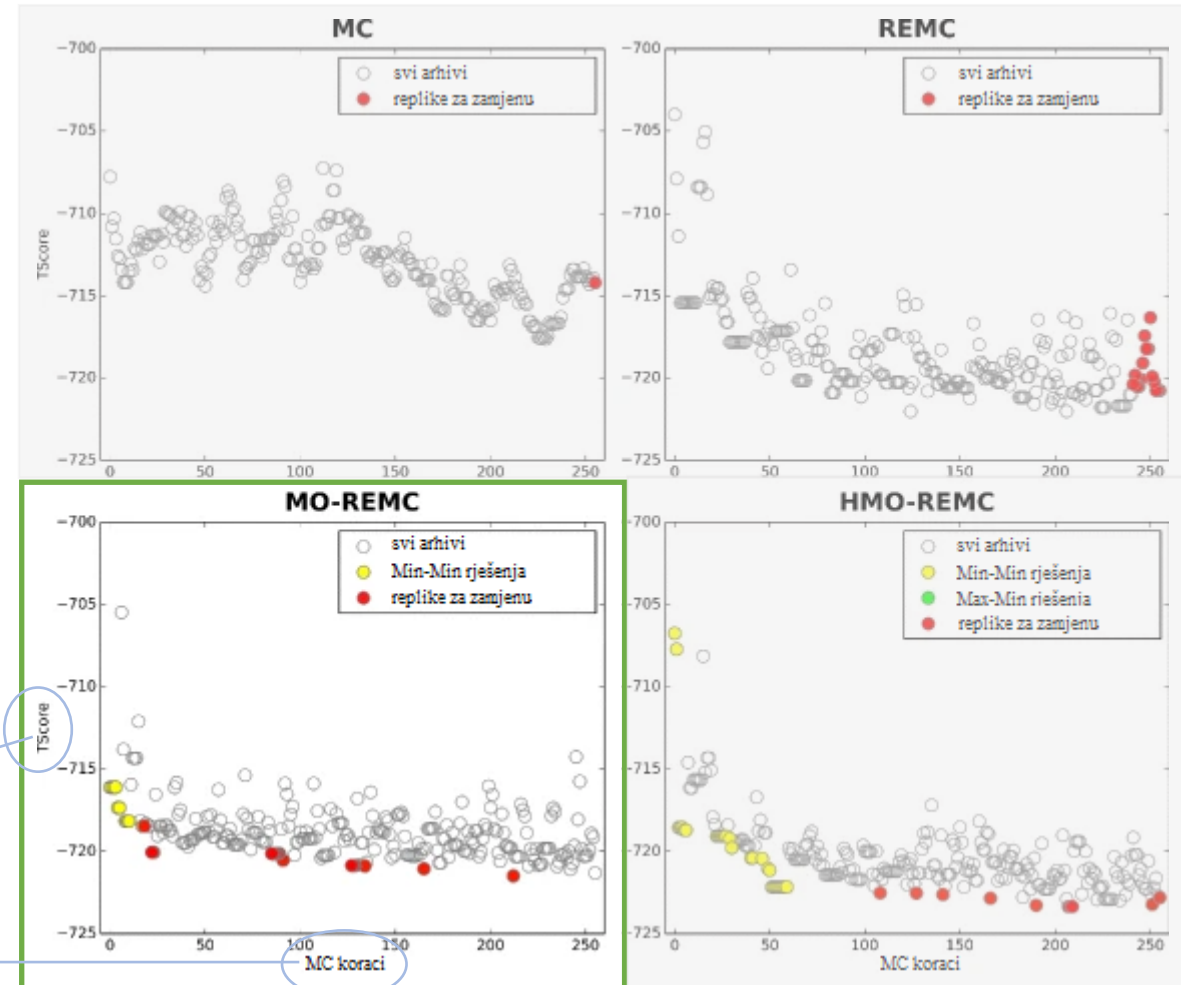
Postoji li bolji način za odabir replika za zamjenu?

Monte Carlo metode uz višekriterijsku optimizaciju zamjene replika (MO-REMC)

- **Replike za zamjenu** = nedominirana rješenja Pareto fronte
 - ⇒ Ograničen broj nedominiranih konformacija pronađenih REMC metodom → set replika sličnih stanjima niže energije ali se istovremeno međusobno maksimalno razlikuju
- * Svaka jedinka u arhivi generirana REMC metodom procjenjuje se pomoću **binarnih kriterija**
 - Ovisnost o prošlosti
 - Bolja konvergencija REMC algoritma
 - povezana s odabirom replika za zamjenu

informacijska mapa dostupnog konformacijskog prostora ovisna o prošlosti

vremenske serije za proces pretrage



H. Wang, H. Liu, L. Cai, C. Wang, and Q. Lv, *BMC Bioinform.* **18** (2017) 1–21

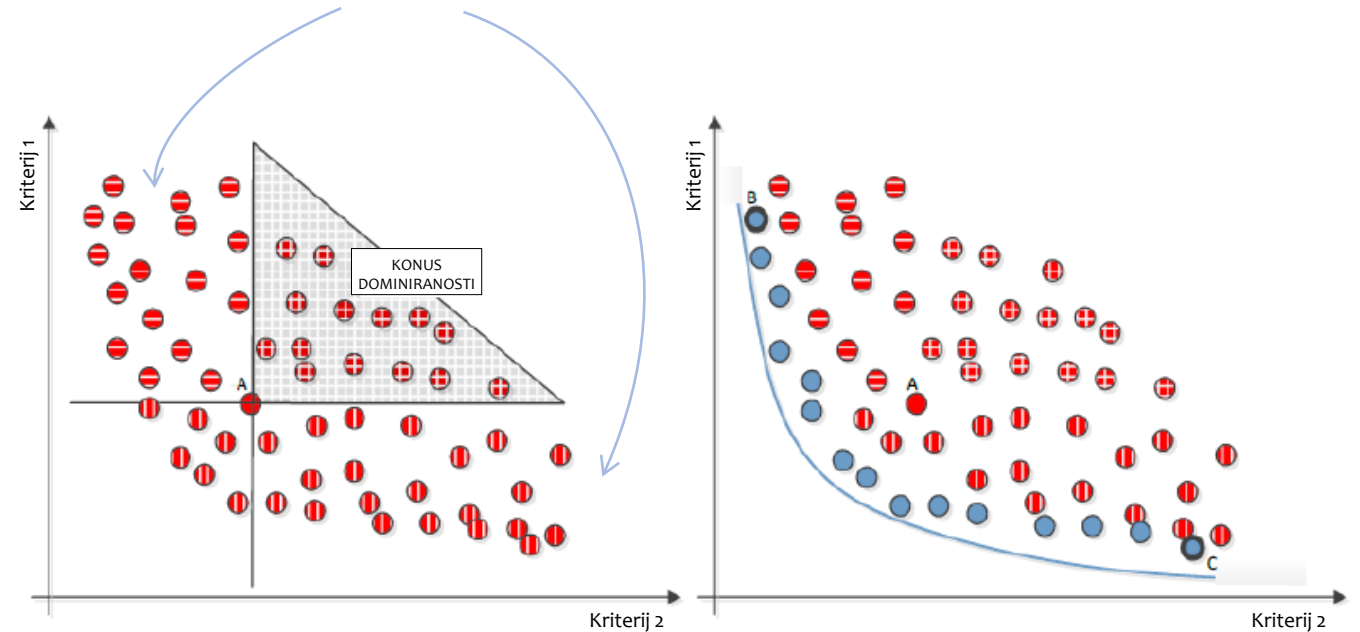
- **Cilj MO-REMC metode uzorkovanja: povećanje brzine konvergencije za dostupan konformacijski prostor.**

Višekriterijska optimizacija

- **Višekriterijski optimizacijski problem** (engl. *multiobjective optimization problem*) sastoji se od nekoliko kontradiktornih kriterija koji se moraju istovremeno optimizirati
- Postupak: pronaći **Pareto optimalan skup** = skup međusobno **nedominiranih** rješenja

Pr. optimizacijskom problemu s definirana dva kriterija koje treba minimizirati i koji ne sadrži ograničenja

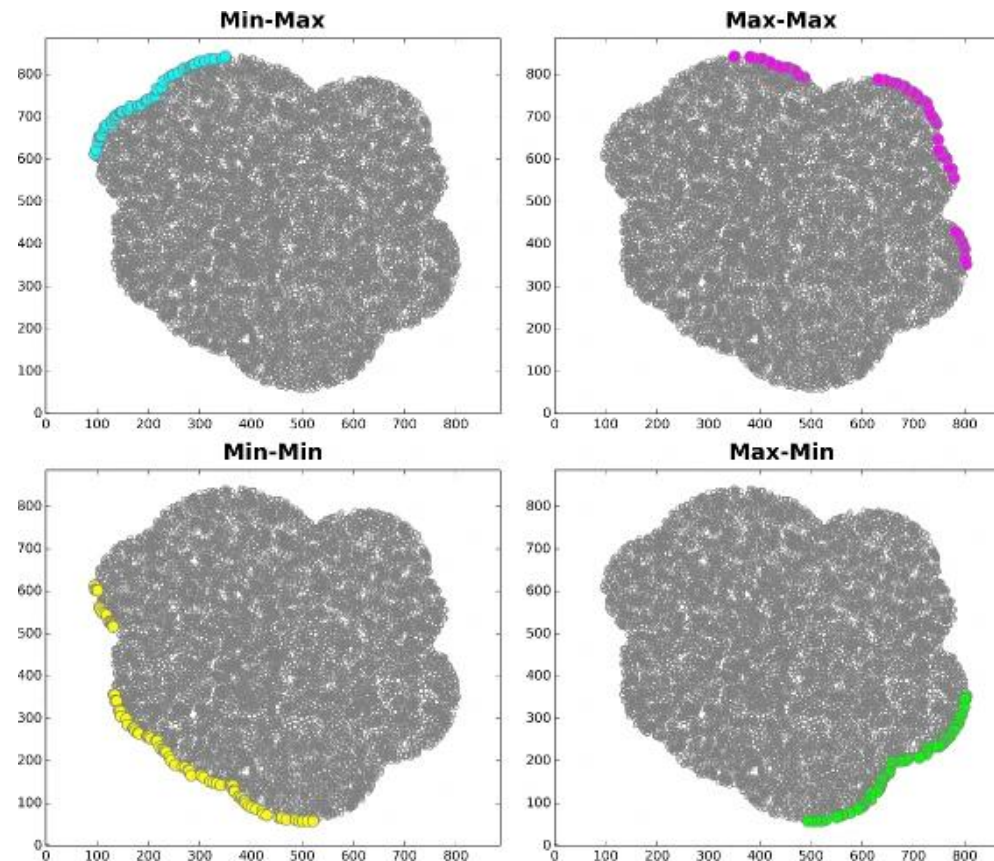
- * U prostoru definiranom kriterijima, leži na površini poznatoj kao **Pareto fronta**
- * Skupovi **nedominiranih rješenja** \Rightarrow rješenja bolja od svih ostali barem po jednom kriteriju



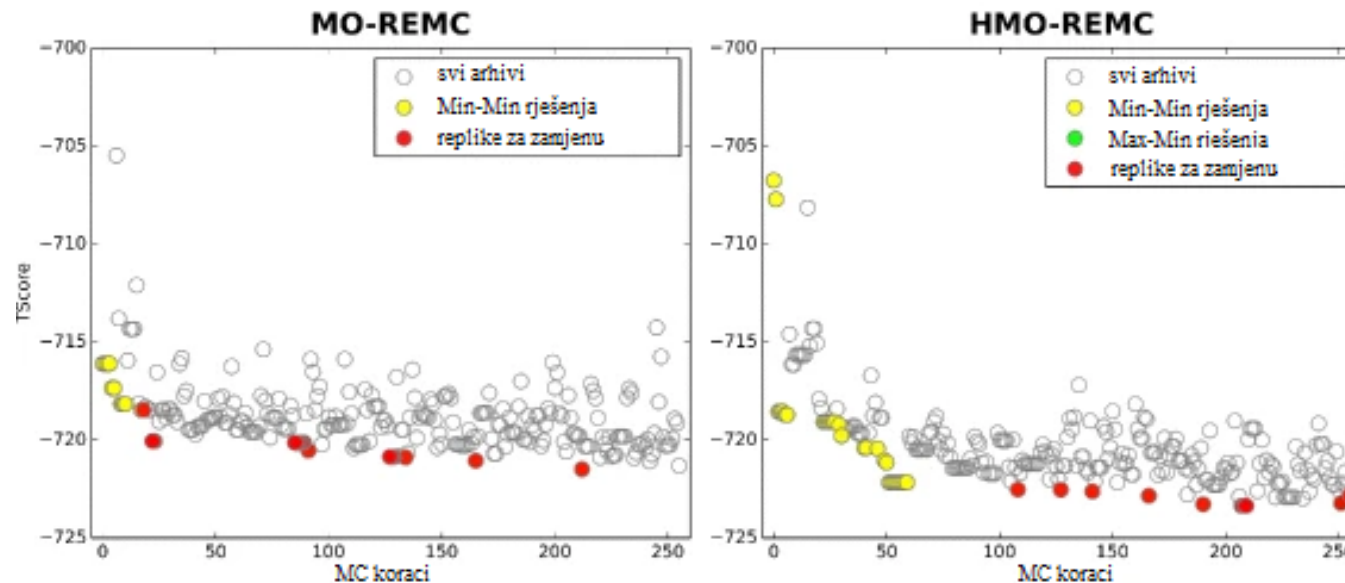
L. Jambrečić, *Višeciljna optimizacija konstrukcije tankera za prijevoz nafte*, Diplomski rad, Fakultet strojarstva i brodogradnje, Sveučilište u Zagrebu, 2017, str. 11.

Hibridna Monte Carlo metoda uz višekriterijsku optimizaciju zamjene replika (HMO-REMC)

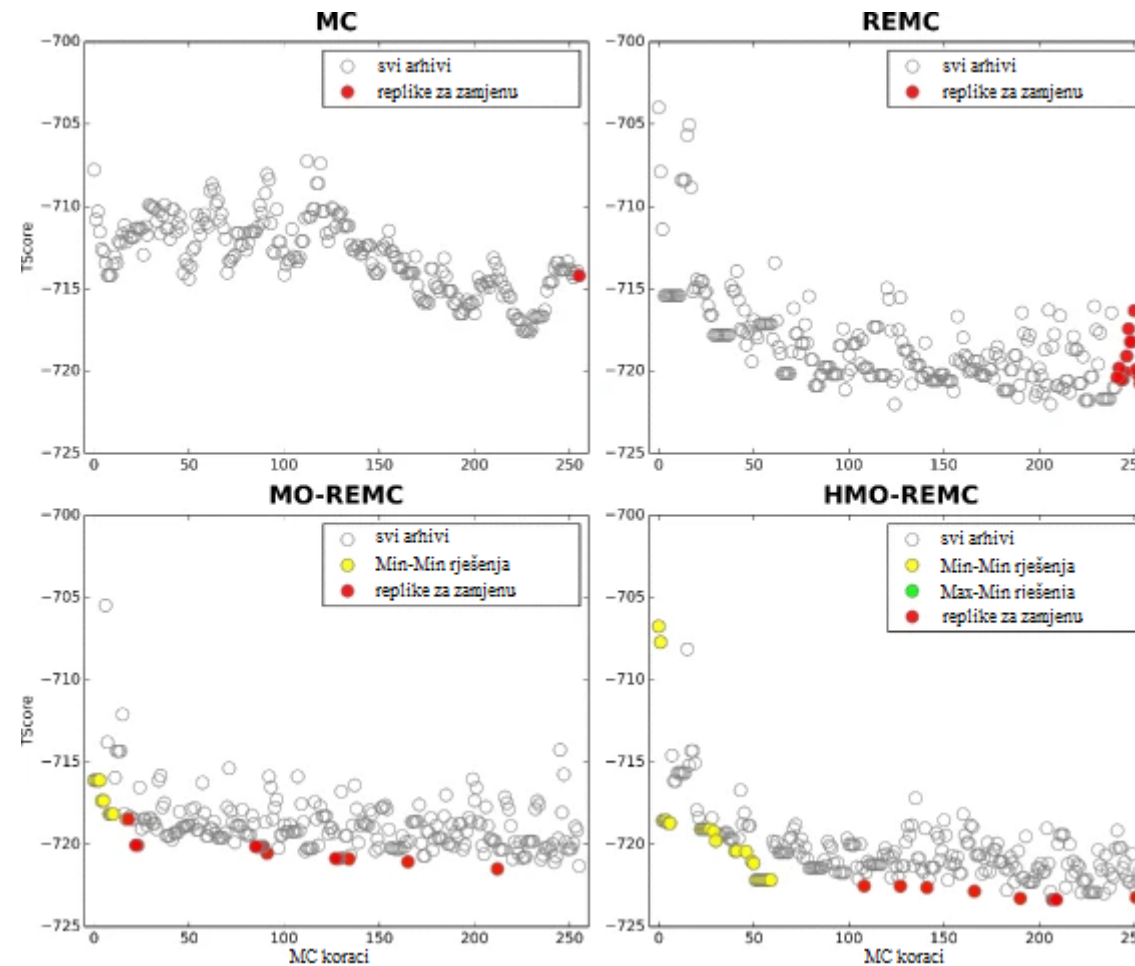
* Optimizacijski problem s dva kriterija → svaki kriterij moguće je minimizirati ili maksimizirati



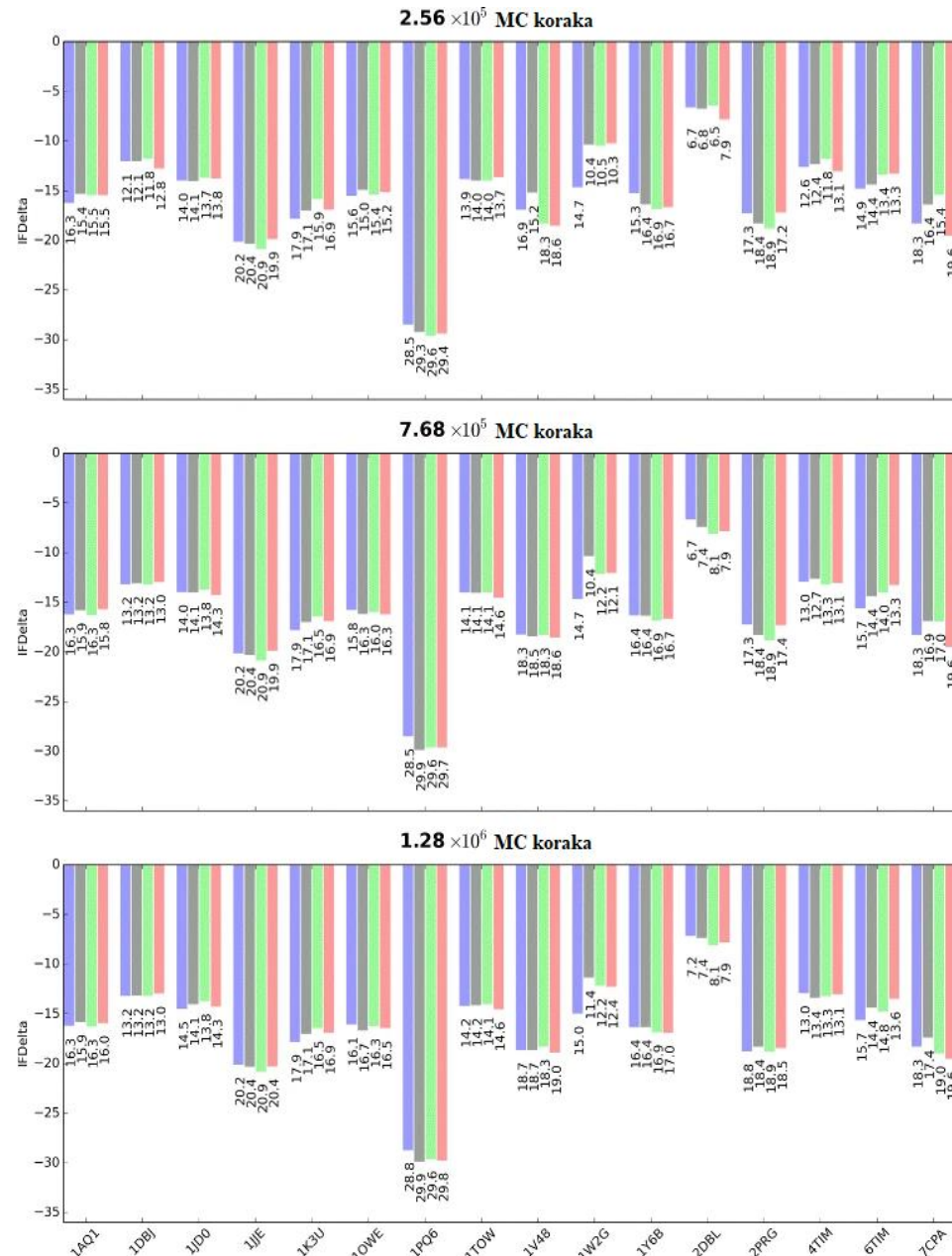
- **Razlika** u odnosu na MO-REMC: postpak odabira elitnih rješenja
→ Odabiru se iz nedominiranih rješenja u arhivima u kojima su i kriterij MC koraka i *Tscore* **minimizirani**, kao i rješenja s **maksimiziranim** kriterijem MC koraka i **minimiziranim** kriterijem *Tscore* vrijednosti.



- Replike odabrane u jednoj iteraciji pomoću MC, REMC, MO-REMC i HMO-REMC metode

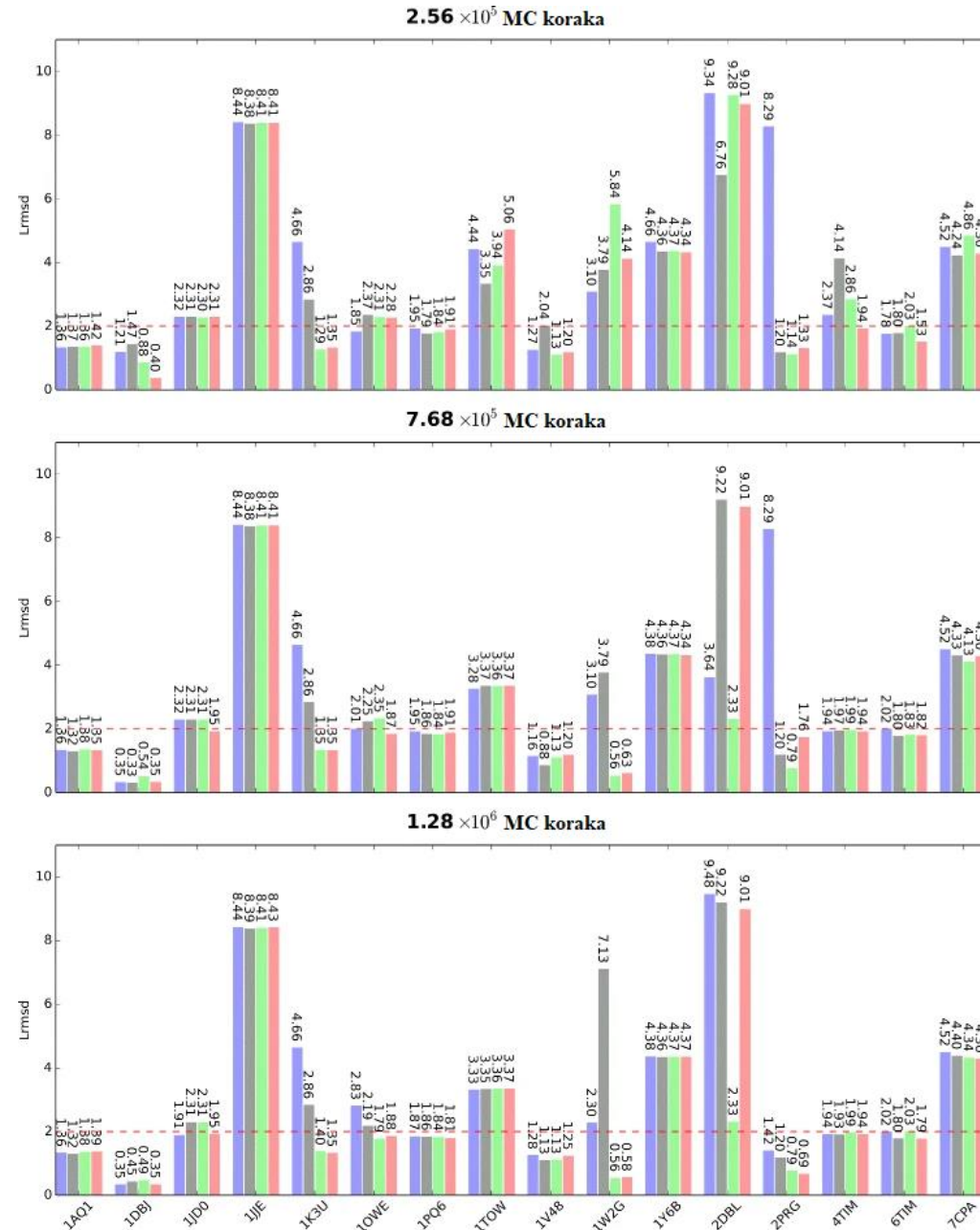


a) **Kvaliteta globalne izvedbe**
(engl. *global performace*)
dokinga u RosettaLigand
računalnom programu →
vezna energija sučelja
proteina (engl. *binding
energy interface delta*,
IFDelta).

**Oznake:**

- Metode:
 - Ljubičasto – MC
 - Sivo – REMC
 - Zeleno – MO-REMC
 - Crveno – HMO-REMC
- 1AQ1, 1DBJ, 1JDO, 1IJE, 1K3U, 1OWE, 1PQ6, 1TOW, 1V48, 1W2G, 1Y6B, 2DBL, 2PRG, 4TIM, 6TIM, 7CPA - receptori
- $2,26 \times 10^5$, $7,68 \times 10^5$ i $1,28 \times 10^6$ – MC koraci (= numR × numC × broj trajektorija) – duljine trajanja simulacije

b) Sveukupna izvedba
(engl. *overall performance*)
dokinga → razlika u RMSD
liganda (engl. *ligand of
RMSD, Lrmsd*).

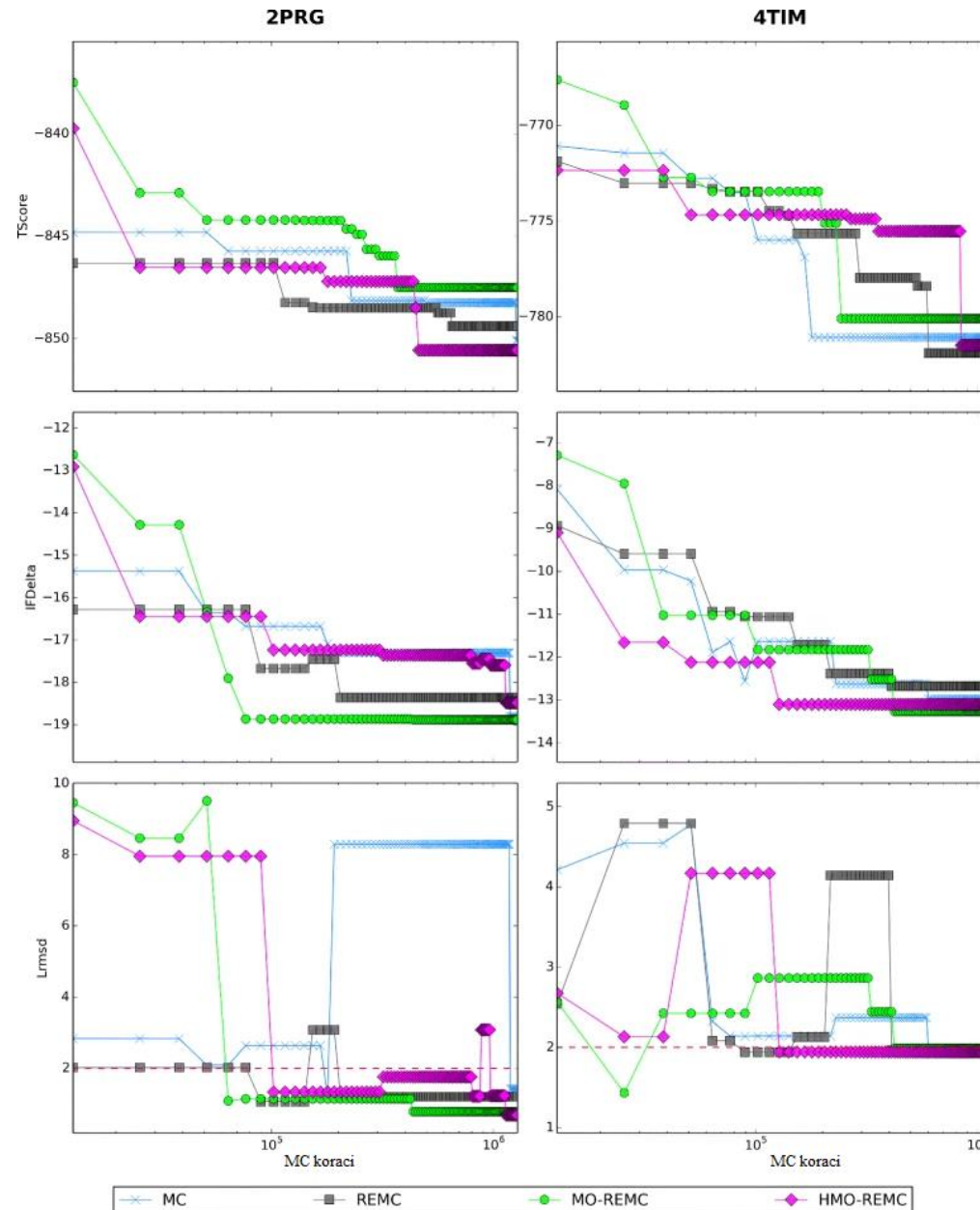


Oznake:

- Metode:
 - Ljubičasto – MC
 - Sivo – REMC
 - Zeleno – MO-REMC
 - Crveno – HMO-REMC
- 1AQ1, 1DBJ, 1JDO, 1IJE, 1K3U, 1OWE, 1PQ6, 1TOW, 1V48, 1W2G, 1Y6B, 2DBL, 2PRG, 4TIM, 6TIM, 7CPA - receptori
- $2,26 \times 10^5$, $7,68 \times 10^5$ i $1,28 \times 10^6$ – MC koraci (= numR × numC × broj trajektorija) – duljine trajanja simulacije

c) **Konvergencija** ukupne vrijednosti energijske funkcije (engl. *energy function total score, TScore*) i *IFDelta* vrijednosti.

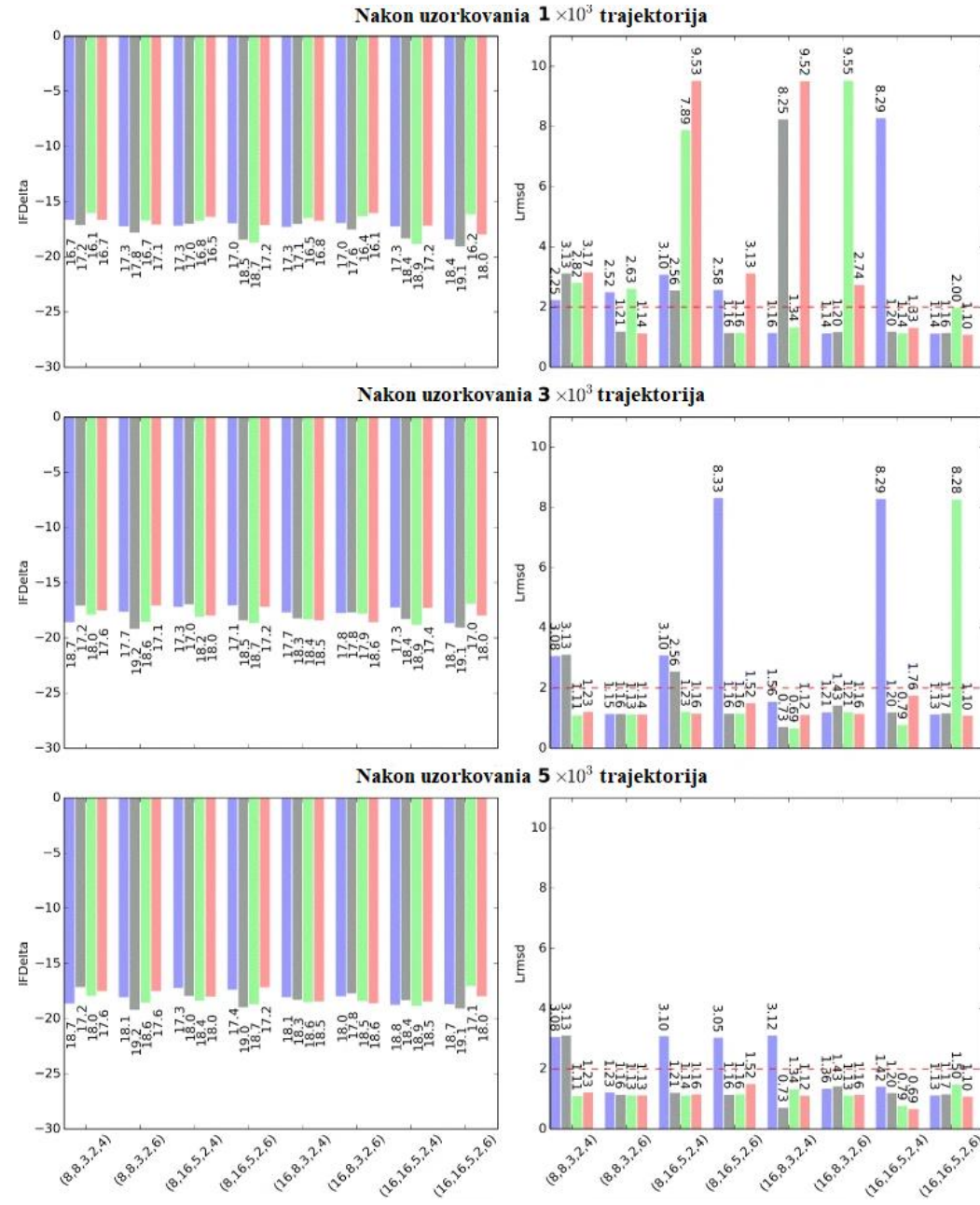
* *Lrmsd* → provjera utjecaja metode na rezultate.



Oznake:

- 2PRG i 4TIM – odabrani receptori
- $MC\ koraci = numR \times numC \times broj\ trajektorija$

d) Valjanost razvijenih metoda (MO-REMC i HMO-REMC) → usporedba više simulacija uz različite početne parametre.



Oznake:

- Metode:
 - Ljubičasto – MC
 - Sivo – REMC
 - Zeleno – MO-REMC
 - Crveno – HMO-REMC
- Apscisa – početni parametri; npr. (8, 8, 3, 2, 4) := (numR, numC, repackNth, minT, maxT)
- 2PRG - odabrani receptor
- 1×10^3 , 3×10^3 i 5×10^3 – broj trajektorija

- Specifično vezanje liganda za receptor nužno je za pravilan rad metabolizma i provođenje signala u tijelu, a razumijevanje spomenute interakcije ključno je i za razvoj lijekova
 - Molekularni doking → generiranje svih mogućih konfiguracija → nasumične metode (Monte Carlo algoritmi)
 - Brzina konvergencije → elitna rješenja (važna komponenta evolucijskih višekriterijskih optimizacijskih algoritama) → MO- i HMO-REMC → informacije Pareto fronte
- Primjena unaprijeđenih metoda može rezultirati boljim uzorkovanjem struktura prilikom molekularnog dokinga male molekule u aktivno mjesto proteina
- MO-REMC i HMO-REMC metode brže konvergiraju prema rješenju u usporedbi s MC i REMC metodama uzorkovanja
- Daljnjim unaprijeđenjem razvijenih metoda, primjerice nasumičnim odabirom replika iz arhiva ili uključivanjem Hamiltonijanske zamjene replika i sl., moglo bi se ostvariti još bolje uzorkovanje struktura

- C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, 2nd ed. (Wiley, Chichester, West Sussex, England, 2004).
- H. Wang, H. Liu, L. Cai, C. Wang, and Q. Lv, *BMC Bioinform.* **18** (2017) 1–21.
- A. R. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed. (Pearson Prentice Hall, Harlow, 2001).
- D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, *Nat. Rev. Drug Discov.* **3** (2004) 935–949.
- X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, *Curr. Comput. Aided-Drug Des.* **7** (2012) 146–157.
- S. Y. Huang, and X. Zou, *Int. J. Mol. Sci.* **11** (2010) 3016–3034.
- V. Salmaso, and S. Moro, *Front. Pharmacol.* **9** (2018) 1–16.
- Y. Okamoto, *J. Mol. Graph. Model.* **22** (2004) 425–439.
- L. Zheng, A. A. Alhossary, C. K. Kwoh, and Y. Mu, *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.* **1–3** (2018) 550–566.