

# Odkrivanje novih zdravil s superračunalnikom

marec 2024

---

## Avtorji:

- dr. Davor Sluga, asistent
- dr. Ratko Pilipović, docent
- dr. Nejc Ilc, docent

Vsi avtorji so zaposleni na Fakulteti za računalništvo in informatiko Univerze v Ljubljani.

## Prolog

»V redu,« si rečeš, »očitno je to zadnja preizkušnja ...« Modrice na levi in ureznina na desni roki dokazujejo, da je tokratna igra preživetja ena težjih v tvoji karieri poklicnega pustolovca. Utrujen spomin ti govori o orjaških zobnikih in letečih motornih žagah, ki so te ogrožale v prejšnji dvorani. Trenutno pa je vse sumljivo mirno. Si v manjšem atriju in svetloba pozno pomladnega sonca ti dobro dene. Na drugi strani so velika hrastova vrata in zelo ti je všeč umetelno kovana ključavnica – malo manj pa dejstvo, da nimaš še ključa zanjo. Ko preiskuješ prostor, se nenadoma spotakneš ob ovijalko in takoj se celoten prostor strese v temeljih. Prestrašeno gledaš, kako se tla sredi atrija odpirajo v brezno. Na srečo ta potres traja samo toliko, da razkrije ovalen bazen, ki si ga sicer v prijetnejših okoliščinah že nekaj časa želiš na svojem vrtu. Oddahneš si, v bazenu ni piraj ali vodikove kisline. Samo ključi so. Za en bazen ključev torej. Začneš jih preizkušati ...

Povejmo to zgodbo malo drugače. Namesto Indiana Jonesa izberimo znanstvenika z doktoratom iz kemije in popoldansko dejavnostjo v farmaciji.

»V redu,« si rečeš, »očitno imam vse pripravljeno za eksperiment. Potem mi objava v prvi četrtini ne uide ...« Mesece se že nisi dobro naspal, mišice v vratu in ramah te pečejo in potreboval bi nova očala. Ta novi virus nas je očitno vse presenetil. Le kako bi ga onеспособili? Kako bi zavrlji njegovo razmnoževanje ali pa sposobnost, da prodre v človeške celice? Hm, ta beljakovinska bodica, ki je zanj tako značilna, ima nekaj prav zanimivih vezavnih mest. Če nam uspe na ta mesta, ki so kot ključavnica, vezati ustrezno spojino, ki ji pravimo ligand, potem morda imamo rešitev. Imam torej ključavnico, le kakšen ključ bi ji ustrezal? Najbolje, da začnem preizkušati možne spojine. Med vsemi možnimi molekulami v kemijskem vesolju jih je samo  $10^{60}$  takih, za katere ocenjujemo, da imajo farmakološki potencial – da bi bile primerne kot zdravilo za ta presneti kovid. Najbolje, da jih začnem preizkušati ...

## Molekulsko sidranje za odkrivanje novih zdravil

Pri razvoju novih zdravilnih učinkovin raziskovalni laboratoriji uporabljajo robotizirano opremo za visokozmogljivo reševanje (angl. *high-throughput screening*, *HTS*), ki omogoča avtomatizirano eksperimentalno *in vitro* biološko vrednotenje različnih spojin z zmogljivostjo več sto tisoč spojin [1]. Gre za razmeroma drag in zapleten postopek, ki pa ga s hitrim razvojem računalniške znanosti dopolnjujejo in nadomeščajo računalniška orodja za molekulsko modeliranje. Podobno kot poskuse izvajamo *in vivo* ali *in vitro*, lahko govorimo tudi o poskusih *in silico*, kar dobesedno pomeni v *siliciju*, s tem pa v bistvu mislimo na računalniško simulacijo.

Sidranje molekul uvrščamo med metode molekulskega modeliranja, ki je postalo nepogrešljivo orodje odkrivanja novih zdravil ter drugih farmacevtsko-kemijskih raziskav [2]. V kontekstu molekulskega sidranja govorimo tudi o postopkih "navideznega" visokozmogljivega reševanja (angl. *high-throughput virtual screening* - *HTVS*), kjer preiskujemo ustreznost več milijard spojin deset tisočkrat hitreje, kot bi to opravil robotiziran sistem [3]. *In silico* eksperimenti nam tako že v zgodnjih fazah razvoja novih zdravil pomagajo razumeti interakcije med molekulo zdravila in vezavnemu mestu na beljakovini.

Da strnemo: sidranje molekule je torej postopek vezave molekule oz. liganda v vezavno mesto druge molekule, imenovane receptor ali terapevtska tarča. Ligand je tipično majhna molekula z nekaj deset atomi, receptor pa je navadno makromolekula, denimo beljakovina ali nukleinska kislina [4]. Slika 1 prikazuje primer molekule liganda v vezavnem mestu receptorja. Receptor je prikazan kot siva gmota z vdolbinami in izrastki, z rdečo polprosojno plastjo je označeno vezavno mesto in v njem idealna poza liganda, ki se veže na to mesto.

## V čem je sploh problem?

Namen uvodnega domišljjskega sestavka je, da začutimo izjemno računsko zahtevnost procesa molekulskega sidranja, saj moramo preiskati ogromen prostor možnih poz liganda. Poleg tega v raziskavah HTVS preverjamo na milijarde različnih ligandov. Imamo pa srečo, da lahko ta problem enostavno razbijemo na medsebojno neodvisne izračune ustreznosti poz liganda in tako omogočimo izrabo porazdeljenih in vzporednih računalniških arhitektur.

V zadnjem desetletju na področju visokozmogljivega računalništva prihaja do premika od klasičnih računalniških arhitektur, kjer celoten posel obdela centralna procesna enota, proti heterogenim sistemom. Moderne heterogene sisteme navadno sestavljajo večjedrne centralne procesne enote (CPE), grafične procesne enote (GPE) in drugi pospeševalniki, denimo programirljivo polje logičnih vrat (angl. *field programmable gate-array*, *FPGA*). Med najhitrejšimi 500 superračunalniki na svetu jih 170 izkorišča pospeševalnike, med katerimi izrazito prednjačijo GPE [5]. Najhitrejši in hkrati energijsko najučinkovitejši superračunalnik, ameriški Frontier, sestavljajo računska vozlišča z eno CPE in štirimi GPE, pri čemer slednje prispevajo več kot 99 % njegove računske moči. V času pisanja tega prispevka je najzmogljivejši slovenski superračunalnik, HPC Vega (Slika 2), na 198. mestu v svetovnem merilu.

Velik potencial heterogenih računalniških arhitektur je trenutno na področju molekulskega sidranja skorajda neizkoriščen, kar je do nedavnega veljalo tudi za odprtokodno orodje CmDock, ki ga razvija slovenska skupnost znanstvenikov in drugih prostovoljcev [6]. Gre za programsko opremo, ki ima svoje začetke v leta 1998 razvitem programu RiboDock in v osnovi teče samo na enem jedru CPE. Vsekakor to predstavlja veliko omejitev hitrosti izvajanja.

## Naš pristop

V nadaljevanju opisujemo zasnovo vzporednega algoritma za molekulske sidranje, ki lahko teče na GPE, in prve rezultate na slovenski superračunalniški gruči HPC Arnes. K dodajanju podpore za GPE smo pristopili s pomočjo standarda in knjižnic OpenCL, ki ga podpira večina proizvajalcev strojne opreme. Omogoča vzporedno izvajanje programov na različnih napravah, kot so CPE, GPE in FPGA. Razviti algoritem izkorišča naravno vzporednost optimizacije z genetskim algoritmom, ki se uporablja v orodju CmDock.

Genetska optimizacija je stohastičen proces, kjer preiskujemo prostor možnih rešitev s pomočjo operacij, ki jemljejo navdih iz naravne evolucije; to so operacije križanja in mutacije. Slednje se izvajajo nad množicami osebkov, ki so v našem primeru poze ligandov. Najboljši osebki se ohranjajo in preko genetskih operatorjev razvijajo skozi generacije, dokler ne dosežemo točke, ko so osebki že tako dobri, da se ne izboljšujejo več. Začetna populacija osebkov mora biti velika, da lahko ustrezno prečesemo celoten prostor rešitev. Dobra novica je, da lahko osebke obdelujemo vzporedno in neodvisno od ostalih. S tem zagotovimo visoko izkoriščenost GPE, ki zmore vzporedno izvajati na tisoče niti, ki so jim bili dodeljeni posamezni osebki.

Najpomembnejšo vlogo v optimizacijskem postopku igra cenilna funkcija, ki pove kako dober je nek osebek in pravzaprav vodi celoten postopek iskanja optimalne rešitve. V našem scenariju mora oceniti, kako dobro se nek osebek (ligand) prilaga vezavnemu mestu beljakovine. Za prvi prototip smo izdelali nekoliko enostavnejšo cenilno funkcijo kot je privzeta za CmDock, saj smo jo morali popolnoma prilagoditi za izvajanje na GPE.

## Rezultati

Za potrebe testiranja in validacije smo si pri razvoju pomagali s superračunalniško infrastrukturo. Uporabljali smo predvsem Arnesovo gručo, ki je trenutno druga najzmogljivejša superračunalniška gruča v Sloveniji s preko 90 računskimi vozlišči, ki vsebujejo več kot 4400 procesorskih jeder, 48 GPE Nvidia V100 in 16 GPE Nvidia H100.

Teste smo izvedli na nekaj tipičnih molekularnih tarčah in ligandih različnih kompleksnosti, ki smo jih pridobili iz odprte banke podatkov o beljakovinah RCSB [7]:

- **3PTB: protein tripsin in zaviralec benzamidin (Slika 3)**
- **3CS9: tironska kinaza BCR-ALB1 in zaviralec nilotinib (Slika 4)**
- **600K: regulator apoptoze BCL-2 in zaviralec venetoklaks (Slika 5)**

Primerjali smo čas izvajanja sidranja med zmogljivo CPE AMD EPYC 7702P ter GPE Nvidia V100 za 20 začetnih populacij osebkov. Poleg tega smo merili odstopanje poze ligandov od referenčnih eksperimentalnih poz. S tem smo preverjali, ali vzporedna različica genetskega algoritma na GPE deluje pravilno.

Algoritem na GPE doseže različne pohitritve, odvisno od velikosti in kompleksnosti problema, kar prikazuje Preglednica 1.

	3PTB	3CS9	600K
Pohitritev izvajanja	4,9	18,1	24,2

**Preglednica 1: Efektivna pohitritev molekularnega sidranja na različnih kompleksih z uporabo GPE Nvidia V100 v primerjavi z osnovnim CmDock, ki smo ga poganjali na AMD EPYC 7702P.**

Tukaj je potrebno povedati, da je dodajanje podpore za grafične pospeševalnike dolgotrajen in zahteven proces, saj je potrebno izvorno programsko kodo najprej do potankosti razumeti, jo ustrezno posodobiti in prilagoditi. Dobljeni rezultati so samo začetek, saj je odprtih še veliko možnosti za optimizacijo in izboljšavo cenilnih funkcij. Končni cilj je zmogljivo orodje za molekularno sidranje, ki bo omogočilo prečesavanje velikih baz potencialnih kandidatov za novo prelomno zdravilo v razumnem času in bo znalo izkoriščati moderno strojno opremo.

## Epilog

*Po celodnevem iskanju pravega ključa se izmučen s hrbtom nasloniš na vrata in pogledaš v nebo. Ravno, ko tvojo pozornost pritegne oblak v obliki pudlja, začutiš, da se vrata odpirajo in da izgubljaš ravnotežje. Očitno si s telesom nevede pritisnil na kljuko. Vrata sploh niso bila zaklenjena ...*

## Zahvale

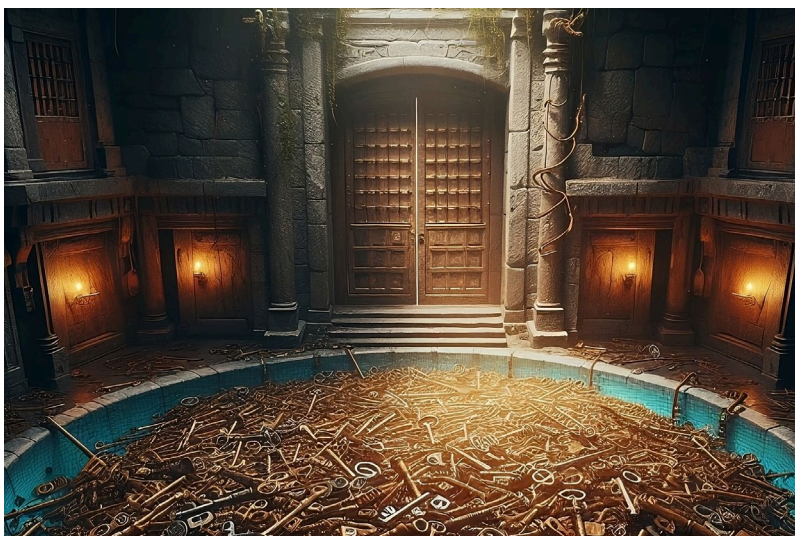
Razvoj programske opreme za vzporedno molekularno sidranje na GPE je potekal v sklopu projekta *Razvoj programskih orodij za računalniško kemijo* v okviru operacije *Študenti UL v delovnem okolju*. Projekt je financiralo Ministrstvo za visoko šolstvo, znanost in inovacije v okviru razpisa *Projektno delo za pridobitev praktičnih izkušenj in znanj študentov v delovnem okolju*.

Za sodelovanje pri raziskavah in razvoju se zahvaljujemo doc. dr. Marku Jukiću in doc. dr. Črtomirju Podlipniku. Zelo hvaležni smo tudi naslednjim študentom UL FRI in FKKT za njihov neprecenljiv prispevek k izvedbi projekta: Ivani Podlipnik, Špeli Žunec, Tinetu Erentu, Mihi Kovaču, Jakobu Maležiču, Jaki Prelogu in Andreju Sušniku.

## LITERATURA / BIBLIOGRAPHY:

- [1] V. Blay, B. Tolani, S. P. Ho, and M. R. Arkin, "High-Throughput Screening: today's biochemical and cell-based approaches," *Drug Discovery Today*, vol. 25, no. 10, pp. 1807–1821, 2020, doi: 10.1016/j.drudis.2020.07.024.
- [2] L. Pinzi and G. Rastelli, "Molecular Docking: Shifting Paradigms in Drug Discovery," *International Journal of Molecular Sciences*, vol. 20, no. 18, p. 4331, 2019, doi: 10.3390/ijms20184331.
- [3] N. Nikitina, M. Manzyuk, Č. Podlipnik, and M. Jukić, "Volunteer Computing Project SiDock@home for Virtual Drug Screening Against SARS-CoV-2," in *Computer Science Protecting Human Society Against Epidemics*, 2021, vol. 616 IFIP, pp. 23–34. doi: 10.1007/978-3-030-86582-5\_3.
- [4] F. Stanzione, I. Giangreco, and J. C. Cole, "Use of molecular docking computational tools in drug discovery," in *Progress in Medicinal Chemistry*, vol. 60, D. R. Witty and B. Cox, Eds. Elsevier, 2021, pp. 273–343. doi: 10.1016/bs.pmch.2021.01.004.
- [5] TOP500.org, <https://www.top500.org/lists/top500/>
- [6] CmDock repozitorij, <https://gitlab.com/Jukic/cmdock/>
- [7] S. K. Burley, C. Bhikadiya, et al., "RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning," in *Nucleic Acids Research*, 2023, vol. 51, pp. D488–D508. doi: 10.1093/nar/gkac1077.

**SLIKE / FIGURES:**

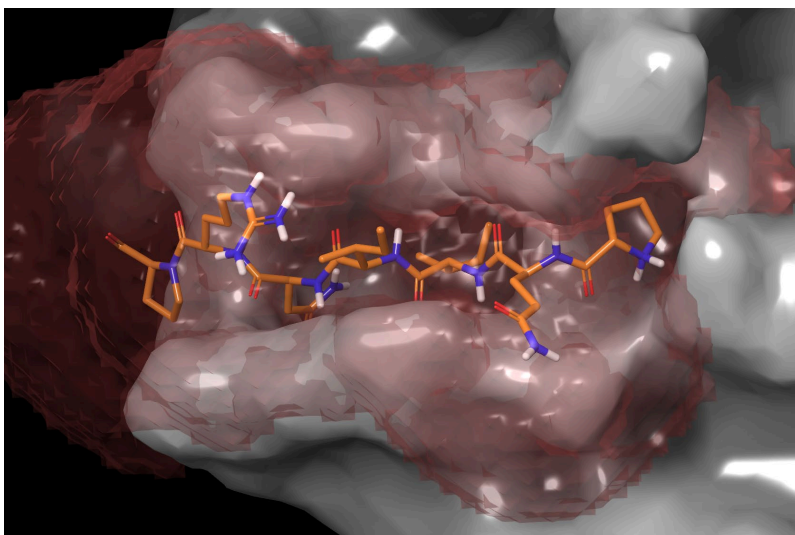


SI: Napovedna grafika

EN: Cover image

Avtorstvo / Source: Bing Copilot, DALL-E 3

Datoteka: fig0-cover-image.jpeg



**Slika 1**

SI: Ligand v vezavnem mestu beljakovine. Vezavno mesto je prikazano kot rdeča polprosojna plast.

EN: Ligand in the binding site of a protein. The binding site is shown as a red translucent layer.

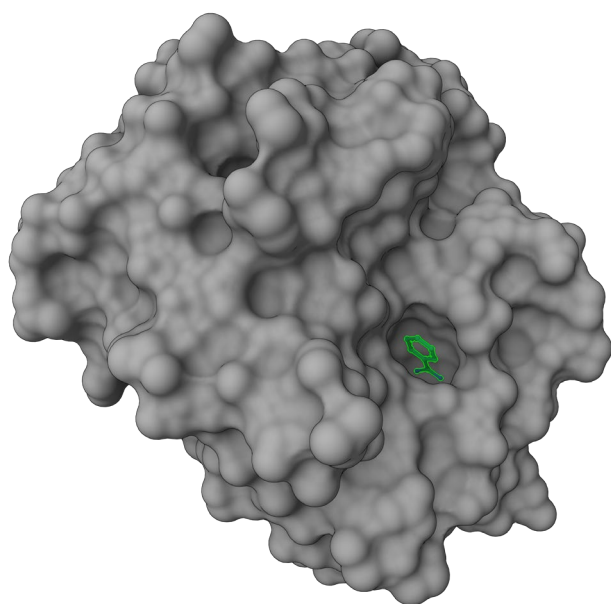
Avtorstvo / Source: Nejc Ilc

Datoteka: fig1-binding-site.png



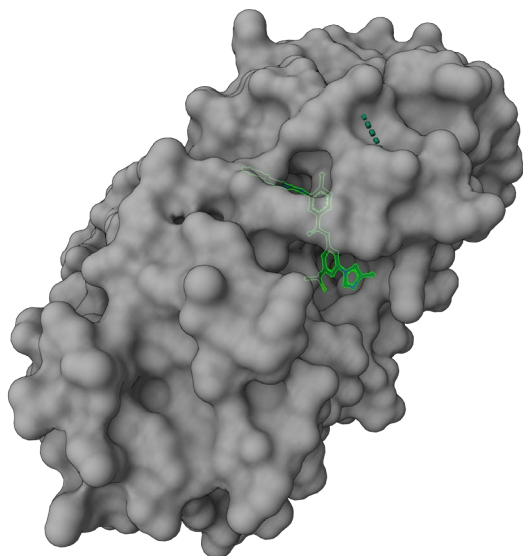
Slika 2

SI: Najzmogljivejši slovenski superračunalnik HPC Vega.  
 EN: HPC Vega, Slovenia's most powerful supercomputer.  
 Avtorstvo / Source: Nejc Ilc  
 Datoteka: fig2-HPC-Vega.jpg



Slika 3

SI: Vizualizacija kompleksa 3PTB. Ligand v referenčni pozi je označen z zeleno barvo.  
 EN: Visualisation of the 3PTB complex. The ligand in the reference pose is highlighted in green.  
 Avtorstvo / Source: RCSB PDB Mol\* Viewer  
 Datoteka: fig3-3PTB.png



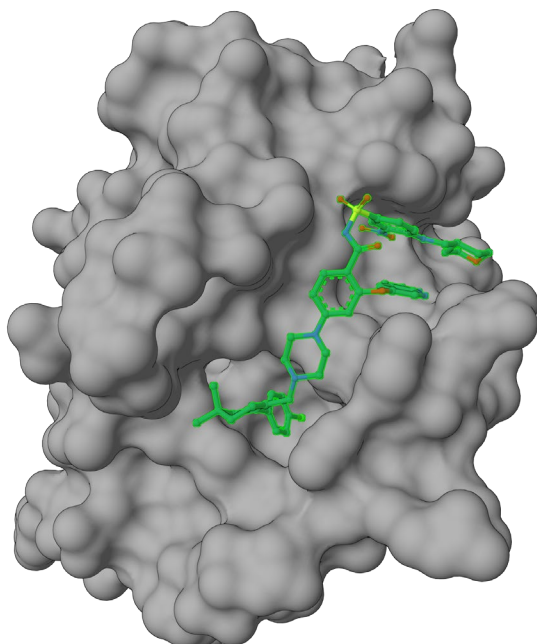
**Slika 4**

SI: Vizualizacija kompleksa 3CS9. Ligand v referenčni pozi je označen z zeleno barvo.

EN: Visualisation of the 3CS9 complex. The ligand in the reference pose is highlighted in green.

Avtorstvo / Source: RCSB PDB Mol\* Viewer

Datoteka: fig4-3CS9.png



**Slika 5**

SI: Vizualizacija kompleksa 600K. Ligand v referenčni pozi je označen z zeleno barvo.

EN: Visualisation of the 600K complex. The ligand in the reference pose is highlighted in green.

Avtorstvo / Source: RCSB PDB Mol\* Viewer

Datoteka: fig5-600K.png