

---

*Emilija Milijić*

## *In silico* dizajn leka za HPV-11

---

*Molekularni doking je sve atraktivnija metoda virtuelnog skrininga potencijalnih novih lekova za različite bolesti i stanja. Ideja ovog istraživanja bila je drug repurposing postojećih lekova, sve u cilju pronalaska novog potencijalnog inhibitora transaktivacionog domena (TAD) E2 proteina humanog papilomavirusa tipa 11 (HPV-11), čime bi se onemogućila replikacija virusa. Pripremljena je baza malih molekula Cashew i sprovedena kroz molekularni doking u programu MPI Vina, nakon čega su najuspešniji rezultati sprovedeni kroz ponovni doking u programu Schrödinger Glide. Na osnovu najuspešnijih rezultata molekularnog dokinga, pomoću LSTM Chem, naprednije verzije rekurentne neuronske mreže, generisana su 44 unikatna molekula. Na taj način je formiran finalni set od 59 molekula, koji obuhvata 15 molekula iz baze Cashew koji su se najstabilnije vezali za TAD E2 proteina i 44 generisana molekula. Mali molekuli iz finalnog seta ostvarili su stabilnu vezu sa transaktivacionim domenom E2 proteina, nakon čega su prošli kroz QSAR analizu. Na osnovu dobijenih rezultata, zaključeno je da se potencijalno mogu koristiti u terapiji HPV-11 nakon in vitro eksperimenta u laboratoriji. Cilj ovog istraživanja bio je in silico kvantifikovati vezu malih molekula sa TAD-om E2 proteina HPV-11, dizajnirati nove molekule na osnovu strukture najstabilnije vezanih i ispitati osobine i jednih i drugih.*

---

### Uvod

Papilomavirusi su mali, dvolančani DNK virusi koji mogu inficirati čoveka i druge sisare, odnosno mogu se replikovati u epitelu njihove kože i sluzokože (White *et al.* 2003). Postoji više od 100 tipova humanih papilomavirusa (HPV) koji izazivaju različita medicinska stanja i bolesti, počev od genitalnih kondiloma (HPV-6 i -11) do raka grlića materice (HPV-16, -18 i -31) (White *et al.* 2003). Genom HPV-a se sastoji od oko 8000 baznih parova koji kodiraju za oko 10 poznatih proteina: 8 ranih (E1-E8) i dva kasna (L1 i L2), pri čemu L1 i L2 formiraju virusni kapsid (White *et al.*

---

*Emilija Milijić (2000),  
Niš, učenica 4. razreda  
Medicinske škole  
„Dr Milenko Hadžić”  
u Nišu*

*MENTOR:  
Luka Velimirov,  
student Biološkog  
fakulteta Univerziteta u  
Beogradu*

2003). Pored standardnih DNK replikacionih faktora ćelije domaćina, dva virusna proteina, E1 i E2, direktno su povezani sa replikacijom virusnog genoma (Titolo *et al.* 2000).

Protein E1 ima tri funkcionalna domena: C-terminalnu ATPazu/helikazu, centralni DNK-vezujući domen i N-terminalni funkcionalni protein koji je neophodan za replikaciju *in vivo*, ali neobavezan pri replikaciji *in vitro* (Amin *et al.* 2000). Kao takav, E1 ima ulogu prepoznavanja i vezivanja za virusnu DNK, dok kao helikaza razdvaja DNK ispred replikacione viljuške (D'Abramo i Archambault 2011). Uzimajući u obzir važnost uloge E1 proteina u virusnoj replikaciji, pretpostavlja se da inhibiranje E1 proteina antivirusnim jedinjenjima može imati značaj u razvoju lekova za HPV infekcije. E2 takođe predstavlja atraktivan ciljni protein za antivirusna jedinjenja čiji je mehanizam delovanja sprečavanje replikacije virusne DNK (*ibid.* 2011). E2 protein ima dva funkcionalna domena: N-terminalni transaktivacioni domen (TAD), koji se direktno vezuje za E1 protein, i C-terminalni DNK-vezujući domen (*ibid.*). Transport E1 proteina do virusne DNK je omogućen njegovom interakcijom sa E2 proteinom, koji se sa velikim afinitetom vezuje za odgovarajuća mesta na virusnoj DNK (*ibid.*). Kompleks E1-E2 neophodan je za virusnu replikaciju i sprečavanjem njegovog nastanka se znatno umanjuje šansa da se virusna DNK replicira, samim tim dolazi do smanjenja produktivnosti infekcije.

Jedan od načina za inhibiranje nastanka E1-E2 kompleksa je vezivanje određenog liganda za TAD E2 proteina (D'Abramo i Archambault 2011). Neki od poznatih inhibitora TAD-a E2 proteina, koji se specifično vezuju za aminokiseline Glu-39 i Glu-100, su benzofenon, koji se nalazi u kremama za sunčanje (Gabros *et al.* 2019), repaglinid, koji se koristi u terapiji dijabetesa, zatim, maleimid, kao i klasa malih molekula indandiona (White *et al.* 2003). Mali molekuli koji su postojeći lekovi mogu se primenjivati za različite bolesti i medicinska stanja (Pushpakom *et al.* 2019). Pomenuti mali molekuli interaguju sa TAD-om E2 proteina, i na taj način sprečavaju formiranje E1-E2 kompleksa, nakon čega E1 protein nije u stanju da obavi svoju enzimsku aktivnost (White *et al.* 2003).

*In silico* analizom, odnosno računarskom simulacijom planiranog istraživanja pre rada u laboratoriji, smanjuje se potrošnje hemikalija i olakšava razumevanje i predviđanje mogućih ishoda istraživanja (Mah *et al.* 2011).

Cilj ovog istraživanja je da se u programu Schrödinger Glide (Friesner *et al.* 2006) molekularnim dokingom pojedinačno odredi energija veze benzofenona, repaglinida i klase indandion sa transaktivacionim domenom E2 proteina HPV-11, a potom ispita energija veze svih malih molekula označenih kao lekovi u bazi podataka ChEMBL (Gaulton *et al.* 2012) sa TAD-om E2 proteina, kako bi se postojeći lekovi potencijalno koristili višenamenski. Nakon obrade rezultata, na osnovu SMILES stringova (Bjerrum 2017) malih molekula koji su ostvarili najnižu energiju veze sa TAD-om E2 proteina, generisane su strukture molekula koji predstavljaju nove potencijalne ligande i formiran set malih molekula koju su ostvarili najnižu energiju vezivanja sa TAD-om E2 proteina i novogenerisanih molekula, i ispitane njihove fizičko-hemijske osobine.

## Materijal i metode

Pre samog istraživanja, formirane su dve baze malih molekula koji su poznati lekovi, velika baza Cashew od 4838 malih molekula i mala baza Almond koja je sadržala 33 nasumično izabranih molekula iz baze Cashew. Baza Almond sadrži 33 molekula koji su manuelno nasumično izabrani iz baze Cashew i služi samo kao proba. Mali molekuli iz baze Almond su prošli kroz preliminarni molekularni doking radi testiranja rada i preciznosti programa Schrödinger Glide i kasnijeg upoređivanja rezultata sa rezultatima finalnog dokinga. Urađen je finalni molekularni doking malih molekula iz Cashew baze u programu MPI Vina. Mali molekuli koji su se najuspešnije vezali za TAD E2 proteina, odnosno ostavari li najmanje energije veze, prošli su kroz trening LSTM neuronske mreže na osnovu čega su generisani novi molekuli, strukturno slični uspešno vezanim ligandima. Na kraju je ispitana bioaktivnost najbolje vezanih ligandata i novogenerisanih malih molekula QSAR analizom, odnosno kvantifikovan je odnos strukture ligandata i veze sa receptorom.

**Priprema za molekularni doking.** Oformljena je baza Cashew od 4838 malih molekula čije su trodimenzionalne strukture preuzete iz baze bioaktivnih molekula ChEMBL (Gaulton *et al.* 2012). Molekuli za bazu Cashew su izabrani na osnovu sličnosti strukture sa repaglinidom, benzo-fenonom, maleimidom i klasom indandiona, ili na osnovu oznake *drugs* u bazi ChEMBL. Od 33 nasumično izabrana molekula iz baze Cashew, uključujući i repaglinid, oformljena je nova baza, Almond, koja će proći kroz preliminarni molekularni doking. Određena je vrednost energije veze repaglinida i TAD-a E2 proteina.

Za vizuelizaciju je korišćen program PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Fajlovi preuzeti iz baze ChEMBL su konvertovani iz SDF formata u PDBQT pomoću programa Open Babel GUI. Trodimenzionalna struktura TAD-a E2 proteina HPV-11 preuzeta je iz banke podataka proteina RCSB PDB, gde nosi oznaku 1R6K.

**Preliminarni molekularni doking.** Mali molekuli iz novoformljene baze Almond prošli su kroz proces preparacije za molekularni doking u programu Schrödinger Glide (Friesner *et al.* 2006). TAD E2 proteina je takođe prošao kroz proces preparacije, što uključuje i selekciju aktivnih mesta – aminokiseline Glu-39 i Glu-100. U programu Schrödinger Glide urađeni su preliminarni molekularni doking svih malih molekula iz baze Almond i TAD-a.

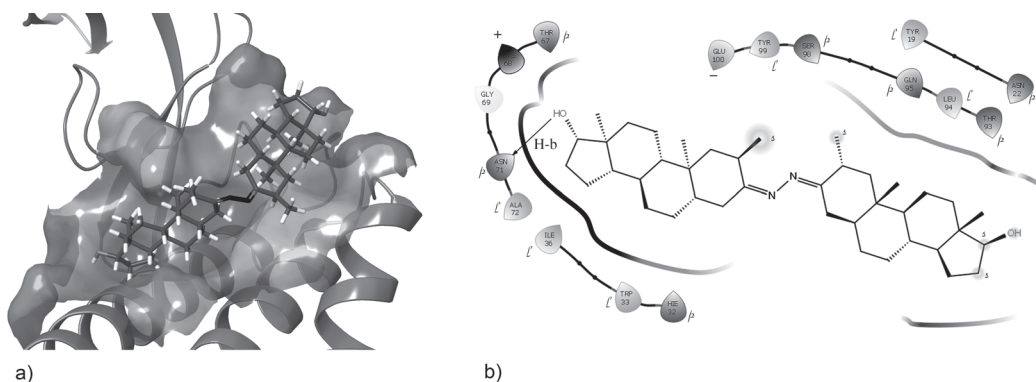
**Finalni molekularni doking.** Mali molekuli iz baze Cashew i TAD E2 proteina prošli su kroz molekularni doking u programu MPI Vina (Trott i Olson 2010) kome je pristupljeno preko Linux klastera. Program MPI Vina je paralelizovana implementacija programa AutoDock Vina i značajno redukuje vreme virtuelnog skrininga koristeći kompjuterske klastere. Nakon dobijenih vrednosti energije vezivanja, 15 nastabilnije vezanih molekula iz baze Cashew je prošlo kroz ponovni, precizniji doking, u programu Schrödinger Glide na standardnom računaru.

**Trening neuronske mreže LSTM i generisanje novih molekula.** Neuronska mreža Long Short-Term Memory Chem (LSTM Chem) je naprednija verzija rekurentne neuronske mreže koja se trenira na bazi reči ili stringova. U našem slučaju LSTM Chem je trenirana na SMILES stringovima 200 hiljada molekula, prema protokolu navedenog u radu Anvite Gupta i saradnika, čime je neuronska mreža prošla kroz opšti trening (Gupta *et al.* 2018). Nakon opšteg treninga, LSTM Chem je fino podešen na 15 molekula iz baze Cashew koji su se najstabilnije vezali za TAD E2 proteina. LSTM Chem je nakon celokupnog treninga generisala 200 molekula. Da bi se proverilo koji su od ovih 200 generisanih molekula validni, a koji od njih su unikatni, korišćen je RDKit u Python-u. Svi validni molekuli imaju osobine koje odgovaraju već poznatim inhibitorima, dok su unikatni svi oni koji nemaju duplikate. Unikatni molekuli (44) koje je generisala LSTM Chem su takođe prošli kroz molekularni doking sa TAD-om E2 proteina.

**QSAR.** Na sajtu Molinspiration izvršena je kvantifikacija odnosa strukture i veze, odnosno QSAR (Quantitative structure–activity relationship) analiza finalnog seta od 59 molekula, odnosno 15 molekula iz baze Cashew koji su se najstabilnije vezali za TAD E2 proteina i 44 generisana unikatna molekula. Za ovaj set od 59 molekula urađena je predikcija fizičko-hemijskih osobina, odnosno QSAR analiza obuhvatala je broj atoma, broj rotabilnih veza, molekulska masa, polarna površina molekula (TPSA), broj odstupanja od pravila Lipinskog i lipofilnost. Pravila Lipinskog određuju kolika je verovatnoća da neko jedinjenje bude lek, odnosno određuju da li neko hemijsko jedinjenje sa biološkom ili farmakološkom aktivnošću ispunjava hemijske i fizičke osobine aktivnog leka. Pravila Lipinskog opisuju osobine molekula bitne za farmakokinetiku u ljudskom organizmu, ali ne predviđa da li je jedinjenje farmakološki aktivno. Pravila Lipinskog su: ne više od 5 donora vodonične veze, ne više od 10 akceptora vodonične veze, molekulska masa ne preko 500 Da, lipofilnost (hidrofobnost) izražena preko logaritma particionog koeficijenta oktanol-voda ( $\log P_{o/v}$ ) da ne premašuje 5 (Lipinski *et al.* 1997). Ukoliko molekul ispunjava barem tri uslova Lipinskog, veća je šansa da ispolji biološku aktivnost, postane aktivan lek i dospe na tržište. Ispitane su osobine koje nalažu i Veberova pravila: da molekul ima 10 ili manje rotabilnih veza i da polarna površina molekula ne prelazi 140 kvadratnih angstrema (Veber *et al.* 2002). Ovom analizom obuhvaćena je i biološka aktivnost, što podrazumeva da je ispitano da li su testirani molekuli inhibitori enzima, ligandi jonskih kanala, inhibitori kinaza, ligandi nuklearnih receptora i inhibitori proteaza.

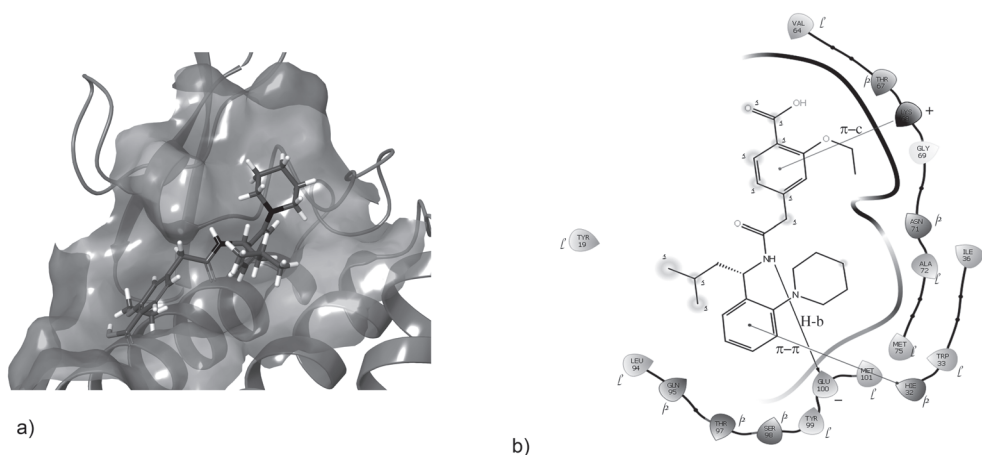
## Rezultati

**Vizuelizacija preliminarne molekularne dokinga.** Molekul koji je ostvario najnižu vrednost energije veze ( $-15.9$  kcal/mol) u programu MPI Vina sa TAD-om E2 protein iz baze Cashew je bolazin. Trodimenzionalni prikaz interakcije bolazina i TAD-a, odnosno formiranje „hidrofobnog džepa” prikazano je na slici 1a, dok je shema te interakcije prikazana na



Slika 1. Interakcija bolazina sa TAD-om. Oznake: -- negativno naelektrisanje, +- pozitivno naelektrisanje, H-b – vodonične veze, *l* – nepolarno, *p* – polarno, *s* – izloženost rastvaraču.

Figure 1. Bolazine-TAD interaction. Tags: -- negative charge, +- positive charge, H-b – hydrogen bonds, *l* – nonpolar, *p* – polar, *s* – solvent exposure.



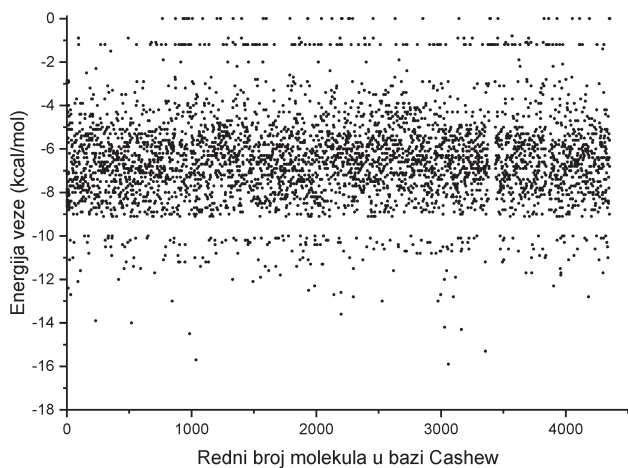
Slika 2. Interakcija repaglinida sa TAD-om. Oznake: -- negativno naelektrisanje, +- pozitivno naelektrisanje, H-b – vodonične veze,  $\pi$ - $\pi$  – pi-pi steking interakcija,  $\pi$ -c – pi-katjon interakcija, *l* – nepolarno, *p* – polarno, *s* – izloženost rastvaraču.

Figure 2. Repaglinide-TAD interaction. Tags: -- negative charge, +- positive charge, H-b – hydrogen bonds,  $\pi$ - $\pi$  – pi-pi stacking interaction,  $\pi$ -c – pi-cation interaction, *l* – nonpolar, *p* – polar, *s* – solvent exposure.

slici 1b. U programu Schrödinger Glide bolazin i TAD ostvarili su energiju veze od  $-3.41$  kcal/mol. Programu su kao aktivna mesta TAD-a zadate aminokiseline Glu-39 i Glu-100 jer je pokazano da se poznati ligandi TAD-a specifično vezuju za ove aminokiseline. Međutim, bolazin je preko hidroksilne grupe u svojoj strukturi ostvario vodonične interakcije sa aminokiselinom Asn-71 TAD-a E2 proteina (slika 1b).

Repaglinid je ligand TAD-a E2 proteina i energija veze repaglinid-TAD iznosila  $-4.3$  kcal/mol u programu Schrödinger Glide. Na slici 2a prikazano je formiranje „hidrofobnog džepa” kao rezultat interakcije repaglinida i TAD-a. Shema interakcije prikazana je na slici 2b, gde se vidi da je repaglinid ostvario pi-pi steking interakcije sa His-32, pi-katjon interakcije sa Lys-68 i očekivane vodonične interakcije sa Glu-100. Očekivana vodonična interakcija sa Glu-39 nije detektovana.

**Energija veze.** Vrednosti energije veze dobijene u programu MPI Vina za sve male molekule iz baze Cashew sa TAD-om E2 proteina prikazane su na dijagramu (slika 3). Najnižih 15 vrednosti navedene su u tabeli 1.

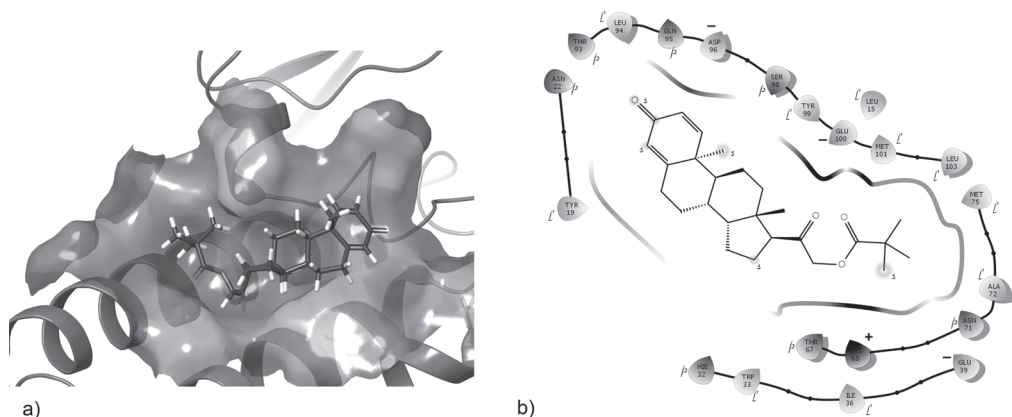


Slika 3. Energije veze 4838 mala molekula iz baze Cashew sa TAD-om E2 proteina (kcal/mol)

Figure 3. Bond energy of 4838 small molecules from Cashew database with TAD of E2 protein (kcal/mol)

Tabela 1. Energije veze TAD-a E2 i malih molekula iz baze Cashew

Naziv molekula	Energija veze (kcal/mol)
Bolazin	-15.9
Roksolonijum	-15.3
Cinoksolon	-14.5
Dezoksikortikosteron-pivalat	-13.9
Triamcinolon-benetonid	-13.6
Epikriptin	-13.0
Deloksolon	-13.0
Lapisterid	-12.8
Bolmantalat	-12.7
Mergokriptin	-12.3
Amcinonid	-12.1
Algeston-acetofenid	-12.1
Halcinonid	-12.0
Vinleurozin	-11.8
Diflomotekan	-11.7



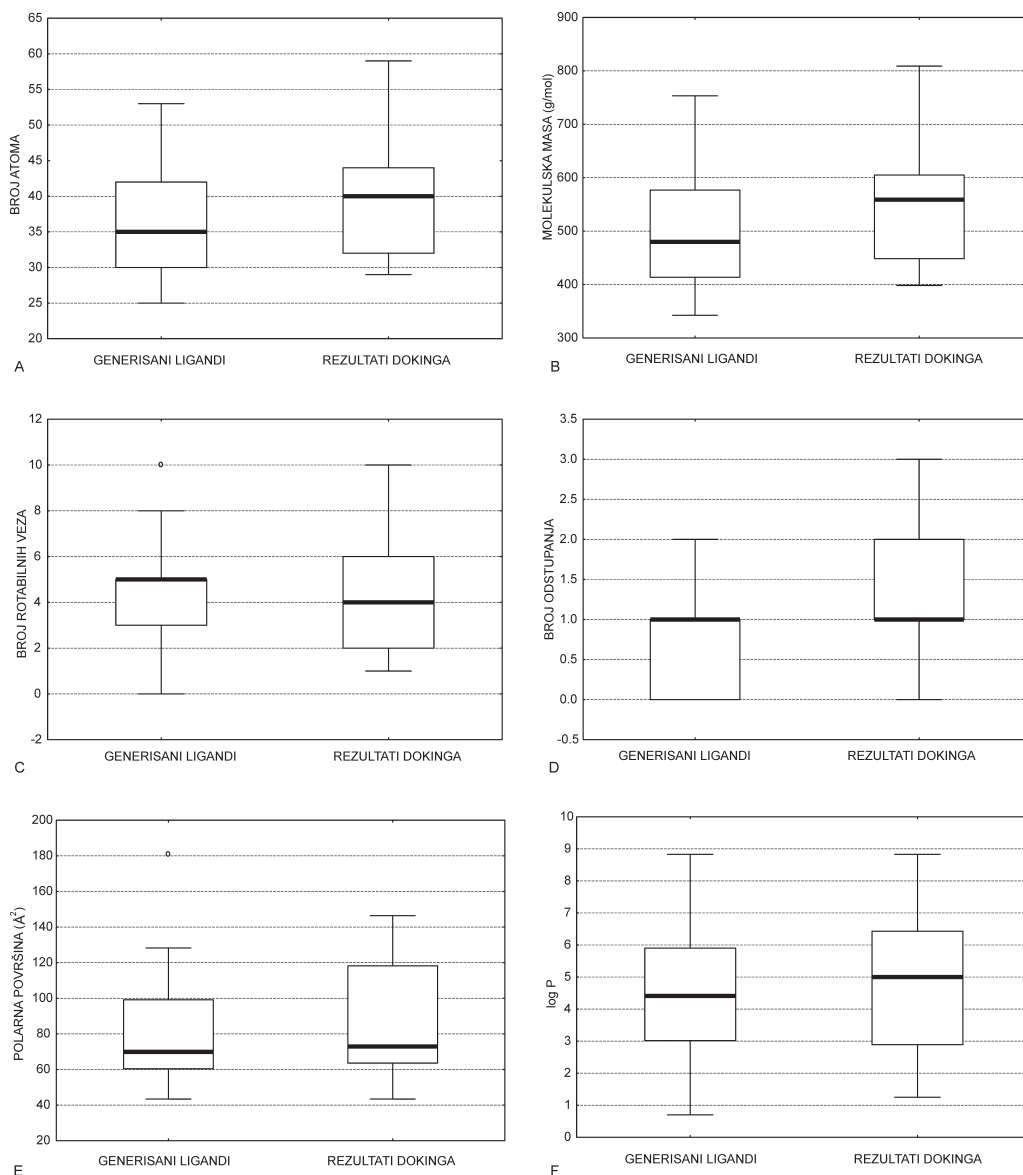
Slika 4. Interakcija Liganda 23 sa TAD-om. Oznake: -- negativno naelektrisanje, + – pozitivno naelektrisanje, *l* – nepolarno, *p* – polarno, *s* – izloženost rastvaraču.

Figure 4. Ligand 23-TAD interaction. Tags: -- negative charge, + – positive charge, H-b – hydrogen bonds, *l* – nonpolar, *p* – polar, *s* – solvent exposure.

**Molekuli generisani neuronskom mrežom LSTM Chem.** Nakon treninga, LSTM Chem je generisala 200 molekula, od kojih su 94 bila validna i 44 unikatna. Novogenerisani Ligand 23, od 44 unikatna molekula, ostvario je energiju veze od  $-3.38$  kcal/mol u programu Schrödinger Glide. Ligand 23 izabran je kao predstavnik seta od 44 unikatna molekula jer je ostvario energiju veze sličnu bolazinu ( $-3.41$  kcal/mol) i sa većim odstupanjem sličnu repaglinidu ( $-4.3$  kcal/mol) što ukazuje na reverzibilnost ove interakcije. Formiranje „hidrofobnog džepa” prikazano je na slici 4a, dok je shema interakcije Liganda 23 i TAD-a prikazana na slici 4b.

**Fizičko-hemijske osobine malih molekula iz finalnog seta.** QSAR analizom predviđene su fizičko-hemijske osobine 59 molekula iz finalnog seta (slika 5) u koje spadaju broj atoma (A) i molekulska masa (B) kojima se opisuje veličina molekula, broj rotabilnih veza (C) koji određuje na koliko načina se molekul može orijentisati prema receptoru, broj odstupanja od pravila Lipinskog (D), polarna površina (E) i lipofilnost (logP, gde je P particioni koeficijent) (F) koja opisuje sposobnost molekula da prođe kroz ćelijsku membranu ili da sa njom interaguje. Osobine generisanih liganada se ne razlikuju bitno od osobina liganada iz tabele 1.

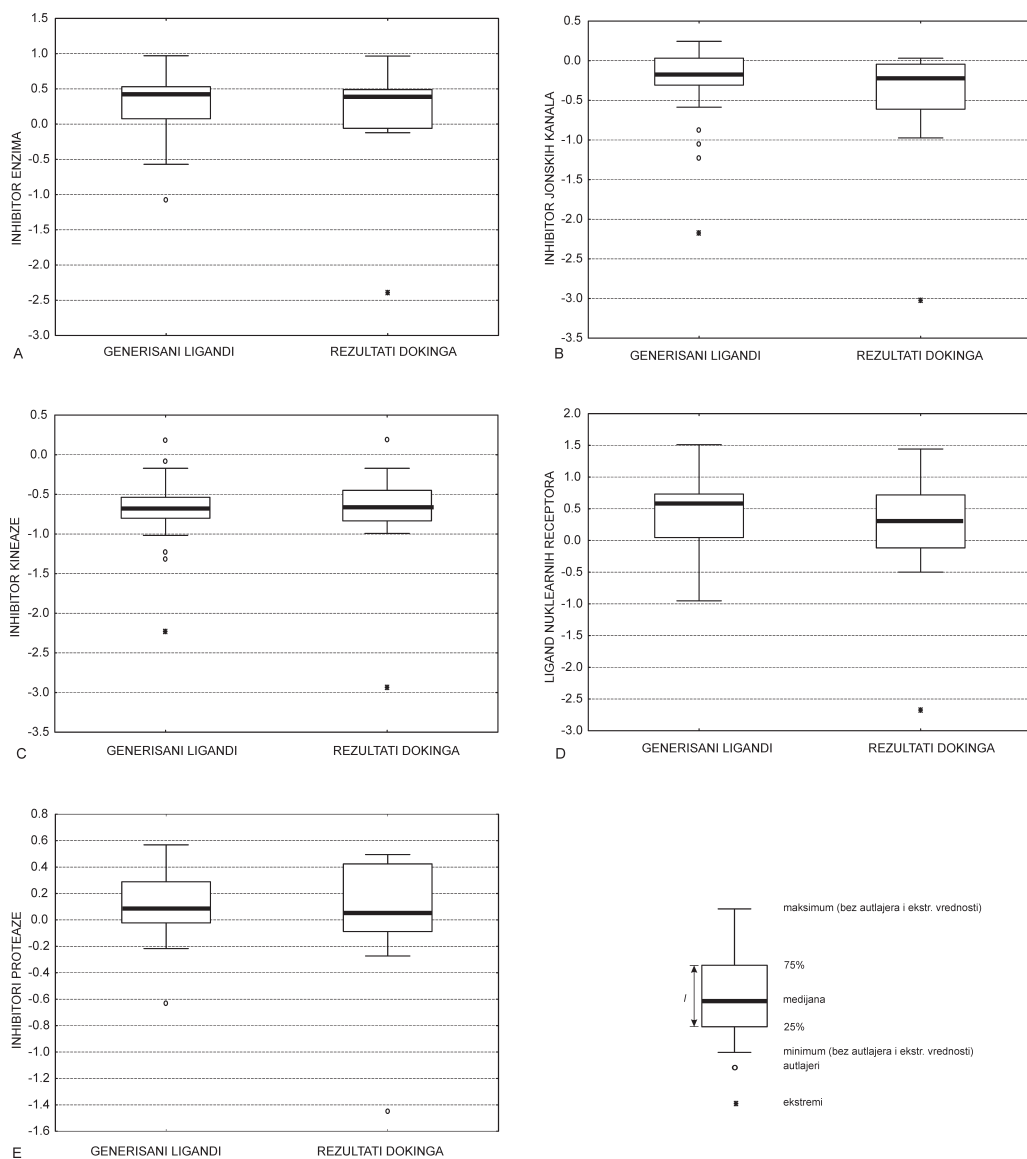
**Biološka aktivnost malih molekula iz finalnog seta.** QSAR analizom je utvrđena i bioktivnost 59 molekula iz finalnog seta (slika 6) gde je opisano da li su molekuli inhibitori enzima (A), ligandi jonskih kanala (B), inhibitori kineaze (C), ligandi nuklearnih receptora (D) i inhibitori proteaze (E). Biološka aktivnost novogenerisanih liganada i liganada iz tabele 1 je vrlo slična, na grafiku D se vidi izvesno odstupanje, odnosno novogenerisani ligandi imaju veću predispoziciju da budu ligandi nuklearnih receptora, medijana je za 0.25 viša od medijane liganada iz tabele 1. Ostale osobine su vrlo slične.



Slika 5. Fizičko-hemijske osobine generisanih jedinjenja i najuspešnijih rezultata iz baze Cashew dobijenih molekularnim dokingom: A) broj atoma, B) molekulska masa, C) broj rotabilnih veza, D) broj odstupanja od pravila Lipinskog, E) polarna površina molekula i F) lipofilnost (mol/mL), izražena preko logP, gde je P oktanol-voda particioni koeficijent. Grafikoni predstavljaju rezultate za 44 generisana liganda (levo) i 15 liganada iz tabele 1 koji su prošli kroz molekularni doking (desno).

Figure 5. Physico-chemical properties of the generated compounds and the most successful results from the Cashew database obtained by molecular docking: A) number of atoms, B) molecular mass, C) number of rotatable bonds, D) number of deviations from Lipinski's rule, E) polar surface of molecules and F) lipophilicity (mol/mL), expressed in logP, where P is the octanol-water partition coefficient. The graphs represent the results for 44 generated ligands (left) and 15 ligands from Table 1 that underwent molecular docking (right).





Slika 6. Biološka aktivnost generisanih jedinjenja i najuspešnijih rezultata molekularnog dokinga iz baze Cashew: A) inhibitor enzima, B) inhibitor jonskih kanala, C) inhibitor kinaze, D) ligand nuklearnih receptora i E) inhibitor proteaze.. Grafikoni predstavljaju rezultate za 44 generisana liganda (levo) i 15 liganada iz table 1 koji su prošli kroz molekularni doking (desno).

Figure 6. Bioactivity of the generated compounds and the most successful results of molecular docking from the Cashew database: A) enzyme inhibitor, B) ion channel inhibitor, C) kinase inhibitor, D) nuclear receptor ligand and E) protease inhibitor. The graphs represent the results for 44 generated ligands (left) and 15 ligands from Table 1 that underwent molecular docking (right).

## Diskusija

U programu Schrödinger Glide dobijeni su rezultati koji su podudarniji sa rezultatima sličnih istraživanja. Program MPI Vina kao rezultate dao je ekstremno niske vrednosti energije veza, što ukazuje na ireverzibilno vezivanje, i očekivano je da one budu manje precizne s obzirom da je program obrađivao veću bazu podataka, odnosno bazu Cashew. Bolazin je u programu Schrödinger Glide ostvario energiju veze od  $-3.41$  kcal/mol sa TAD-om E2 proteina, dok je u programu MPI Vina ostvario energiju veze od  $-15.9$  kcal/mol (tabela 1). S obzirom na to da vrednost energije veze repaglinid-TAD iznosi  $-4.3$  kcal/mol i s obzirom na to da je Schrödinger Glide obrađivao manje podataka, zaključuje se da je značajno selektivniji od MPI Vina. Potencijalni metodološki problem je upravo ovo nepoklapanje Schrödinger Glide-a i MPI Vina. Pokazano je da 15 malih molekula koji su ostvarili najstabilniju vezu sa TAD-om imaju dovoljno jaku interakciju, ali je to potrebno ispitati *in vitro*.

Ligand 23 pokazuje veliku strukturnu sličnost sa bolazinom i repaglinidom. Međutim, upadljive razlike su nedostatak azotnih funkcionalnih grupa kod Liganda 23 i manja molekulska masa u odnosu na bolazin, dok je repaglinid slične molekulske mase.

QSAR analizom pokazano je da 59 malih molekula iz finalnog seta ima povoljne osobine i da se potencijalno mogu koristiti u terapiji HPV-11, što se zaključuje prevashodno na osnovu pravila Lipinskog. U većini slučajeva nema više od jednog odstupanja od ovog pravila zato što su ligandi generisani neuronskom mrežom „dizajnirani“ da fizičko-hemijski liče na postojeće, poznate ligande TAD-a.

## Zaključak

Postoji više od 100 tipova humanog papiloma virusa koji izazivaju različita medicinska stanja i bolesti, jedan od njih je HPV-11 koji izaziva kondilome. Virusni proteini E1 i E2 HPV-11 direktno su povezani sa replikacijom virusnog genoma. E2 predstavlja posebno atraktivan target-protein za antivirusna jedinjenja čiji je mehanizam sprečavanje replikacije virusne DNK. Poznati inhibitori transaktivacionog domena E2 proteina su repaglinid, benzofenon, maleimid, klasa malih molekula indandiona itd. Dobijene su vrednosti energije veza ovih molekula sa transaktivacionim domenom E2 proteina i repaglinid, koji je ostvario energije veze  $-4.3$  kcal/mol sa TAD-om E2 proteina, uzet je kao jedan od referentnih molekula. Svi molekuli iz baze molekula ChEMBL sa oznakom „lekovi“ prošli su kroz molekularni doking u implementaciji programa MPI Vina. Molekuli koji su ostvarili najniže vrednosti energije veze sa TAD-om E2 prošli su kroz molekularni doking u programu Schrödinger Glide, gde je bolazin ostvario energiju veze od  $-3.41$  kcal/mol što je slično repaglinidu te je zaključeno da se on najuspešnije vezao za receptor iz oformljene baze malih molekula. Na osnovu 15 najuspešnije vezanih molekula (tabela 1),

neuronskom mrežom su generisana 44 molekula koji su fizičko-hemijski slični najuspešnije vezanim molekulima iz oformljene baze. Istakao se novogenerisani Ligand 23 koji je fizičko-hemijski sličan bolazinu i predlaže se njegova sinteza. Važno je ove rezultate potvrditi ili opovrgnuti in vitro, dakle sintetisati Ligand 23 i ispitati biološko dejstvo Liganda 23 i bolazina. Za potpuniji uvid u materijal koji je korišćen u ovom istraživanju i dobijene rezultate pogledati sadržaj na repozitorijumu (Milijić 2019).

**Zahvalnost.** Ivanu Razumeniću, Microsoft Srbija, za pristup kompjuterskom klasteru i Ognjenu Milićeviću, asistentu na Medicinskom fakultetu Univerziteta u Beogradu, za konsultacije tokom realizacije projekta.

---

## Literatura

- Amin A. A., Titolo S., Pelletier A., Fink D., Cordingley M. G., Archambault J. 2000. Identification of domains of the HPV11 E1 protein required for DNA replication in vitro. *Virology*, 272 (1): 137.
- Bjerrum E. J. 2017. SMILES enumeration as data augmentation for neural network modeling of molecules. arXiv preprint arXiv:1703.07076
- D'Abramo C. M., Archambault J. 2011. Small molecule inhibitors of human papillomavirus protein-protein interactions. *The open virology journal*, 5: 80.
- Friesner R. A., Murphy R. B., Repasky M. P., Frye L. L., Greenwood J. R., et al. 2006. *Journal of medicinal chemistry*, 49 (21): 6177.
- Gabros S., Nessel T. A., Zito P. M. 2019. Sunscreens and Photoprotection. U *StatPearls* [Internet]. StatPearls Publishing
- Gaulton A., Bellis L. J., Bento A. P., Chambers J., Davies M., et al. 2012. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 40 (D1): 100.
- Gupta A., Müller A. T., Huisman B. J. H., Fuchs J. A., Schneider P., Schneider G. 2018. Generative recurrent networks for de novo drug design. *Molecular informatics*, 37 (1-2): 1700111.
- Lipinski C. A., Lombardo F, Dominy B. W., Feeney P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23 (1-3): 3.
- Mah J. T. L., Low E. S. H., Lee E. 2011. *In silico* SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery. *Drug discovery today*, 16 (17-18): 800.
- Milijić E. 2019. Petnica BMD2019: HPV drug screening – Emilija Milijić. [https://github.com/lxka/PetnicaBMD2019/tree/HPVDrugScreening?fbclid=IwAR2tQZNqsDV1ssh7pFLmSc39T0oN5cuU8jwqf\\_0n8oifkHtJ2wlDEuoXeev](https://github.com/lxka/PetnicaBMD2019/tree/HPVDrugScreening?fbclid=IwAR2tQZNqsDV1ssh7pFLmSc39T0oN5cuU8jwqf_0n8oifkHtJ2wlDEuoXeev)
- Pushpakom S., Iorio F., Eyers P. A., Escott K. J., Hopper S., et al. 2019. Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, 18 (1): 41.
- Titolo S., Pelletier A., Pulichino A-M., Brault K., et al. 2000. Identification of domains of the human papillomavirus type 11 E1 helicase involved in

oligomerization and binding to the viral origin. *Journal of virology*, **74** (16): 7349.

Trott O., Olson A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, **31** (2): 455.

Veber D. F., Johnson S. R., Cheng H-Y., Smith B. R. Keith W Ward, Kopple K. D. 2002. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, **45** (12): 2615.

White P. W., Titolo S., Brault K., Thauvette T., Pelletier A., *et al.* 2003. Inhibition of human papillomavirus DNA replication by small molecule antagonists of the E1-E2 protein interaction. *Journal of Biological Chemistry*, **278** (29): 26765.

---

*Emilija Milijić*

## *In silico* Drug Design for HPV-11

Molecular docking is an increasingly attractive method for virtual screening of potential new drugs for various diseases and conditions. The idea of this research was to repurpose existing drugs, aiming to find a potential transactivation domain (TAD) inhibitor of the E2 protein of the human papillomavirus type 11 (HPV-11). A database Cashew of 4838 molecules was formed. Molecular docking of the Cashew database was conducted using the MPI Vina program. After that, the best-bound molecules went through re-docking in the Schrödinger Glide program. Based on the most successful results of molecular docking, a recurrent neural network LSTM Chem generated 44 unique molecules. In this way, a final set of 59 molecules was formed, comprising 15 molecules from the Cashew database that are most stably bound to TAD and 44 generated molecules. Small molecules from the final set established a stable bond with the TAD of the E2 protein, after which they underwent QSAR analysis. Based on the obtained results, the conclusion is that they could potentially be used in the treatment of HPV-11 after *in vitro* experimental confirmations. This research aimed to quantify the binding of small molecules to the TAD E2 of the HPV-11 protein *in silico*, to design new molecules based on the structure of the most stably bound molecules, and to examine the properties of both. The most stably bound molecule from the Cashew database was bolazine with a bond energy of  $-3.41$  kcal/mol. Repaglinide, the known inhibitor of TAD, achieved bond energy of  $-4.3$  kcal/mol in Schrödinger Glide. The similarity between these values indicates that bolazine could be a potential inhibitor of HPV-11 replication. Physico-chemical properties and bioactivity of best-bound small molecules and the generated ones were examined and a significant resemblance was found between them (Figure 5 and Figure 6).

