

Pregledni rad<https://doi.org/10.47960/2232-9080.2026.31.16.25>

Ilustrativan prikaz sinergije teorije grafova i analize peptida kroz kombinatornu optimizaciju

Anton VrdoljakFakultet građevinarstva, arhitekture i geodezije Sveučilišta u Mostaru, doc. dr. sc.,
anton.vrdoljak@fgag.sum.ba

Sažetak: Mnogi suvremeni problemi biomedicine, biofizike i bioinformatike, ali i brojnih drugih područja tehnike i društvenih znanosti, svode se na primjenu matematičkog modeliranja temeljenog na teoriji grafova. Matematičko modeliranje općenito interpretiramo kao čin opisivanja stvarnoga sustava matematičkim formulacijama i jednadžbama, s ciljem stvaranja i implementacije matematičkog modela za naknadne analize te konstruiranje i optimizaciju sustava unutar zadanog okvira. Cilj ovog rada je kroz ilustrativan prikaz sinergije između teorije grafova i analize peptida demonstrirati kako kombinatorna optimizacija omogućuje rješavanje kompleksnih problema biofizike na računski elegantan i efikasan način. Zbog važnosti pri analiziranju i predviđanju svojstava peptida, fokus predloženog modela bit će na konstrukciji (pronalaženju) manjeg, ali još uvijek reprezentativnog, skupa ljestvica aminokiselina. U tu svrhu primijenjena je metoda pretrage promjenjivom okolinom, suvremena metaheuristika u području kombinatorne optimizacije.

Gljučne riječi: teorija grafova, dominantan skup, ljestvice aminokiselina, kombinatorna optimizacija, pretraga promjenjivom okolinom

An illustrative overview of the synergy between graph theory and peptide analysis using combinatorial optimization

Abstract: Many contemporary problems in biomedicine, biophysics, and bioinformatics, as well as in numerous other fields of engineering and social sciences, can be reduced to mathematical modeling based on graph theory. Mathematical modeling is generally interpreted as the act of representing a real-world system through mathematical formulations and equations, aiming to develop and implement a model that facilitates subsequent analysis, design, and system optimization. The aim of this paper is to demonstrate, through an illustrative account of the synergy between graph theory and peptide analysis, how combinatorial optimization enables the resolution of complex biophysical problems in a computationally elegant and efficient manner. Due to their importance in analyzing and predicting peptide properties, the proposed model focuses on constructing a smaller yet representative subset of amino acid scales. To this end, a Variable Neighborhood Search (VNS) approach is employed – a contemporary metaheuristic in the field of combinatorial optimization.

Key words: graph theory, dominating set, amino-acid scales, combinatorial optimization, variable neighborhood search



1. UVOD

Povijest znanosti bilježi brojne primjere moćnih sinergija između naizgled udaljenih disciplina, koje su nerijetko dovodile do korjenitih promjena u ljudskoj misli. Možda najznačajniji takav trenutak predstavlja ujedinjenje geometrije i algebre, čime je stvoren temelj za suvremenu analitičku geometriju. Ta poveznica, odnosno prevođenje prostornih oblika u simbolički algebarski jezik, nije tek matematička zanimljivost, nego ključna paradigma modernog inženjerstva, fizike i računalnih znanosti, jer omogućuje rješavanje kompleksnih problema koji su nerješivi, ili bi ostali teško dohvatljivi, u strogo odvojenim disciplinama. Kada spojimo te dvije discipline, dobivamo mogućnost da vidimo strukturu (geometrija) i izračunamo preciznost (algebra). Drugim riječima, njihov spoj nas uči da svaki problem ima više lica (vizualno i logičko), da rigidna pravila algebre mogu stvoriti beskonačnu kreativnost u geometriji te da razumijevanje jednog duboko obogaćuje razumijevanje drugog. Slične transformativne procese danas opažamo u suvremenoj biomedicini, biofizici i bioinformatici. Tijekom posljednjih nekoliko desetljeća brojna su istraživanja pokazala kako su grafovi izvrstan i nezaobilazan alat za modeliranje sustava koji naglasak stavljaju na veze i odnose među objektima. Prije svega, grafovi pružaju jednostavan i intuitivan prikaz heterogenih i složenih bioloških procesa te biokemijskih reakcija i interakcija. Štoviše, oni olakšavaju modeliranje i razumijevanje složenih molekularnih mehanizama integracijom teorije grafova, strojnoga učenja i algoritama umjetne inteligencije. Unatoč značajnom napretku umjetne inteligencije, precizno modeliranje strukturnih i funkcionalnih svojstava peptida i dalje predstavlja izazov zbog iznimne kompleksnosti njihova prostora pretraživanja. Upravo u tom kontekstu, kombinatorna optimizacija postaje nezaobilazan alat koji omogućuje sustavno unaprjeđenje modela kroz identifikaciju ključnih parametara. Poseban značaj pritom ima konstrukcija reprezentativnih skupova ljestvica aminokiselina (amino karboksilnih kiselina), čime se postiže nužna redukcija dimenzionalnosti bez gubitka ključnih fizikalno-kemijskih i biokemijskih informacija.

Među različitim heurističkim pristupima, pretraga s promjenjivom okolinom ističe se svojom sposobnošću sustavne promjene strukture susjedstva tijekom procesa pretraživanja. Ovakav pristup omogućuje učinkovito izbjegavanje lokalnih minimuma (optimuma) i istraživanje širih prostora rješenja, što je od presudne važnosti pri rješavanju nelinearnih problema optimizacije skupa ljestvica aminokiselina. Primjenom ove metode u okviru teorije grafova, otvara se put prema računalno održivijim modelima koji zadržavaju visoku razinu biološke relevantnosti.

Ovaj je rad strukturiran na sljedeći način. Nakon uvodnih razmatranja, u drugom poglavlju detaljno se analizira teorijska podloga sinergije teorije grafova i analize svojstava peptida, s posebnim osvrtom na matematičku reprezentaciju bioloških mreža. Treće poglavlje posvećeno je opisu predloženog modela, u kojem se definira problem optimizacije skupa ljestvica aminokiselina te obrazlaže implementacija metaheuristike pretrage s promjenjivom okolinom. U četvrtom poglavlju iznose se i diskutiraju rezultati numeričkih eksperimenata, s fokusom na računalnu efikasnost i biološku relevantnost dobivenih podskupova. Rad završava petim poglavljem, u kojem se donosi zaključak te daju smjernice za buduća istraživanja u području kombinatorne optimizacije u biofizici.

2. TEORIJSKA PODLOGA (Teorija grafova u modeliranju molekularnih sustava)

Teorija grafova u suvremenoj biofizici ne predstavlja samo pomoćni alat za vizualizaciju, već temeljni formalni jezik koji omogućuje mapiranje kompleksnih bioloških entiteta u egzaktnu matematičku strukturu. Ako molekulu peptida promatramo kao graf, u kojem aminokiseline predstavljaju vrhove, a njihove interakcije bridove, otvaramo prostor za primjenu robusnih algoritamskih rješenja. Ovakva apstrakcija omogućuje nam da se odmaknemo od puke deskripcije kemijskih svojstava i fokusiramo na

Vrdoljak, A.,

Ilustrativan prikaz sinergije teorije grafova i analize peptida kroz kombinatornu optimizaciju

topološku organizaciju sustava. Upravo ta topološka perspektiva, u kombinaciji s numeričkim vrijednostima ljestvica aminokiselina, stvara sinergijski okvir unutar kojeg je moguće precizno predvidjeti ponašanje i funkcionalnost peptida.

Eksperimentalna i teorijska svojstva aminokiselina, kao temeljnih građevnih jedinica peptida i proteina, opsežno su istražena. Svaka istraživačka metoda dodjeljuje numeričku vrijednost pojedinoj aminokiselini, a takvo se pridruživanje definira kao ljestvica aminokiselina. Njihova primjena u bioinformatičari ključna je za interpretaciju i predviđanje ponašanja biopolimera (peptida i proteina). Međutim, broj dostupnih ljestvica iznimno je velik; primjerice, postoji više od stotinu ljestvica povezanih isključivo s hidrofobnošću. Takva hiperprodukcija podataka predstavlja značajno računsko opterećenje za algoritme koji generiraju potpuno nove (korisne) deskriptore peptida kombinacijom više ljestvica. Stoga je od primarnog interesa za biofizičare i bioinformatičare konstruirati manji, ali reprezentativan skup ljestvica te sačuvati što je više moguće informacija.

Kako bi se postigla željena redukcija podataka uz minimalan gubitak informacija, u ovom radu predložimo graf-teorijski pristup modeliranju odnosa među ljestvicama aminokiselina. U takvom modelu, svaka ljestvica predstavlja vrh grafa, dok bridovi definiraju razinu sličnosti ili korelacije među njima. Ovakva formalizacija omogućuje nam da problem selekcije reprezentativnog skupa transformiramo u precizno definiran graf-teorijski problem, poput pronalaženja optimalnih podskupova vrhova koji maksimalno pokrivaju informacijski prostor. Primjenom ove metodologije, složena biofizička svojstva aminokiselina postaju podložna egzaktnoj kombinatornoj analizi, čime se osigurava sustavna selekcija deskriptora bez oslanjanja na heurističke pretpostavke.

Transformacija problema selekcije reprezentativnog skupa ljestvica u optimizacijski zadatak nalaže uvođenje preciznog graf-teorijskog aparata. U tom kontekstu, sustav modeliramo pomoću sljedećih osnovnih definicija [10]:

Definicija 1. Graf je uređena trojka $G = (V, E, \varphi)$, gdje je V neprazan skup vrhova, E skup bridova koji je disjunktan s V , a φ funkcija koja svakom bridu iz E pridružuje dva, ne nužno različita vrha iz V . Graf često zapisujemo kao uređeni par $G = (V, E)$ ili samo G . Broj vrhova grafa G označavamo s n , a broj bridova s m .

Definicija 2. Graf $G = (V, E)$ koji nema petlji (bridova koji povezuju vrh sa samim sobom) i višestrukih bridova (dva ili više bridova koji imaju isti par krajeva) naziva se jednostavnim grafom. U suprotnom, graf se naziva pseudografom.

Definicija 3. Težinski graf je uređeni par (G, ω) , gdje je $G = (V, E)$ graf, a $\omega : E \rightarrow \mathbb{R}_0^+$ funkcija koju nazivamo težinska funkcija. Graf koji nije težinski naziva se neuteženim grafom.

Dakle, svaki brid $e \in E$ spaja dva vrha $u, v \in V$, tj. kažemo brid e je incidentan vrhovima u i v , a funkcija ω svakom bridu e pridružuje nenegativan broj $\omega(e)$, kojeg nazivamo težinom brida e . Dodatno, kažemo još kako su u i v krajevi brida e , kao i vrhovi u i v su incidentni s bridom e , te kako su u i v susjedni vrhovi, i pišemo $e = uv$.

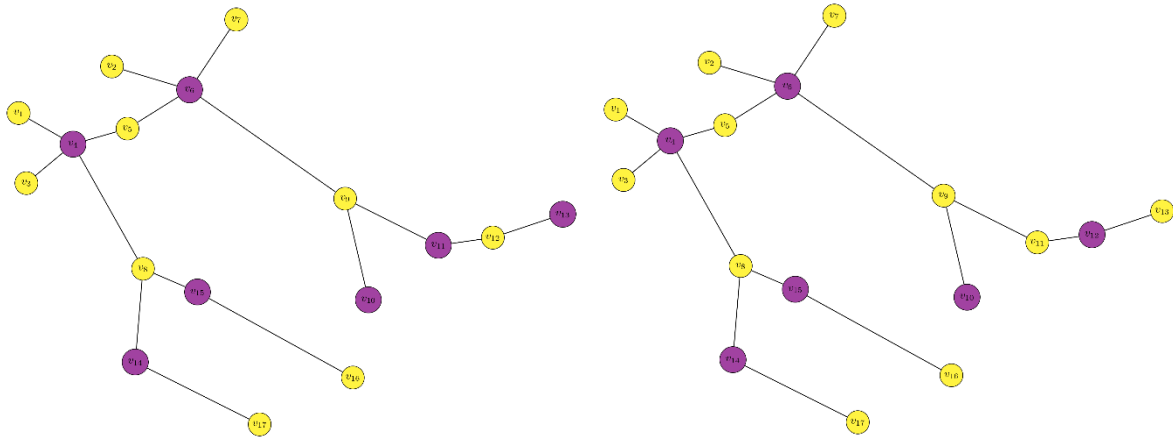
Definicija 4. Pod susjedstvom (okolinom) vrha $v \in V$ jednostavnoga grafa $G = (V, E)$ podrazumijeva se skup

$$N_G(v) = \{u \in V \mid vu \in E\} \quad (1)$$

susjeda vrha v . Stupanj (valencija) vrha $v \in V$ u jednostavnome grafu G je broj susjeda vrha v , $\deg(v) = d_G(v) = |N_G(v)|$ (broj bridova od G incidentnih s vrhom v). Izolirani vrh je vrh sa stupnjem nula. Vrh sa stupnjem jedan naziva se list ili viseći vrh.

Vrdoljak, A.,

Ilustrativan prikaz sinergije teorije grafova i analize peptida kroz kombinatornu optimizaciju

Slika 1. Minimalan (lijevo) i minimuman dominantan skup nekog grafa G [10]

Definicija 5. Neka je $G = (V, E)$ graf. Za $D \subseteq V$ kažemo da je dominantan skup grafa G ako za svaki vrh $v \in V \setminus D$ postoji najmanje jedan $u \in D$ koji je susjedan s v , odnosno za kojega vrijedi $uv \in E$.

Definicija 6. Dominantan skup D grafa G je minimalan ako ne postoji pravi podskup $D' \subset D$ koji je dominantan skup grafa G . Dominantan skup najmanje kardinalnosti (veličine) naziva se minimuman dominantan skup, a njegov kardinalni broj određuje (donji) dominacijski broj $\gamma(G)$ za dani graf G .

Dakle, dominantan skup grafa $G = (V, E)$ je podskup D od V , takav da je svaki vrh iz grafa ili u dominantnom skupu, ili je njegov susjed u dominantnom skupu [3]. Svaki minimuman dominantan skup je minimalni dominantan skup, ali obrat nužno ne vrijedi (vidi Sliku 1.) [13]. Dominacijski broj $\gamma(G)$ za graf G sa Slike 1. je 6. Nadalje, svaki graf ima barem jedan dominantan skup: ako je $D = V$, tj. skup D sadrži sve vrhove grafa G , tada je skup D prema definiciji dominantan skup grafa G , budući je $V \setminus D$ nužno prazan skup. Za grafove bez izoliranih vrhova uvijek postoje dva disjunktna dominantna skupa: ako je $D_m =$ minimalni dominantan skup grafa G , tada je $V \setminus D_m$ dominantan skup. Ako je G graf s n vrhova i m bridova, pri čemu je $0 \leq m \leq \binom{n}{2}$, onda za njegov dominacijski broj vrijedi [1]:

$$\gamma(G) \leq \left\lfloor n + 1 - \sqrt{1 + 2m} \right\rfloor. \quad (2)$$

Iako su prethodne definicije postavile jasan matematički okvir, važno je naglasiti da je pronalaženje minimalnog ili minimumnog dominantnog skupa u općem grafu dokazano NP-potpun problem [4]. To znači da s porastom broja ljestvica aminokiselina (vrhova grafa), vrijeme potrebno za pronalaženje egzaktno optimalnog rješenja raste eksponencijalno, što klasične metode pretrage čini praktički neupotrebljivima. Upravo zbog te računalne složenosti, u rješavanju ovog problema nužno je osloniti se na napredne metaheurističke pristupe. U širem spektru poznatih (meta)heuristika odabrali smo pretragu promjenjivom okolinom (engl. *Variable Neighborhood Search*, VNS) – metodu koja zahvaljujući svojoj sposobnosti sustavne promjene strukture susjedstva (okolina) omogućuje učinkovito pronalaženje

Vrdoljak, A.,

Ilustrativan prikaz sinergije teorije grafova i analize peptida kroz kombinatornu optimizaciju

visokokvalitetnih rješenja u razumnom vremenskom okviru. U idućem poglavlju detaljno se opisuje implementacija ove metaheuristike [7], koja je od svog predstavljanja devedesetih godina dvadesetog stoljeća doživjela brojne primjene i unaprjeđenja [8].

Neka je r_0^2 neki zadani broj (tzv. prag sličnosti). Neka je S skup svih ljestvica aminokiselina. Kažemo kako je $T \subset S$ reprezentativan skup ljestvica ako za svaku ljestvicu $s \in S$ postoji ljestvica $t \in T$ takva da je kvadrat koeficijenta korelacije r^2 između ljestvica s i t veći ili jednak r_0^2 .

Problem pronalaženja minimalnog reprezentativnog skupa ljestvica aminokiselina može se svesti na problem pronalaženja minimalnog dominantnog skupa. Neka je G jednostavan graf sa skupom vrhova $S(G)$ i skupom bridova:

$$E(G) = \{s_1s_2 : s_1, s_2 \in S(G) \text{ i korelacija između } s_1 \text{ i } s_2 \text{ je } \geq r_0^2\}. \quad (3)$$

Svaki dominantan skup u promatranome grafu G odgovara reprezentativnom skupu ljestvica aminokiselina i svaki reprezentativan skup ovih ljestvica odgovara dominantnom skupu. Dakle, minimalan dominantan skup u ovome grafu je minimalan reprezentativan skup ljestvica.

Nažalost, kako je ranije istaknuto, pronalaženje minimalnog dominantnog skupa u općem grafu je dokazano NP-potpun problem.

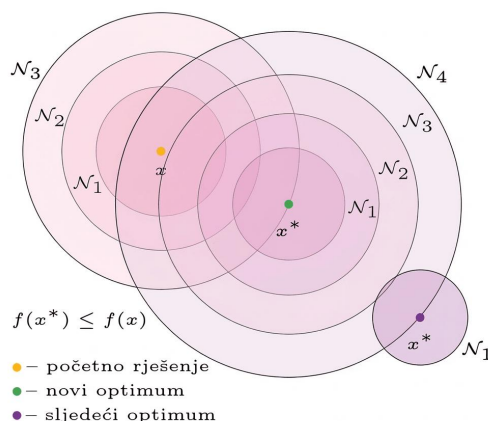
3. METODOLOGIJA I MODEL (Optimizacija skupa ljestvica aminokiselina primjenom VNS)

Metodu pretrage promjenjivom okolinom (u daljnjem tekstu VNS metaheuristika ili samo VNS) predložili su Nenad Mladenović i Pierre Hansen 1997. godine kao robustan pristup za rješavanje kompleksnih problema kombinatorne i globalne optimizacije. Za razliku od egzaktnih metoda koje mogu, barem u teoriji, pružiti optimalno rješenje, odnosno prihvatljivo (izvedivo) rješenje koje optimizira (minimizira ili maksimizira) vrijednost zadane funkcije cilja, metoda pretrage promjenjivom okolinom, poput ostalih heurističkih metoda, daje prihvatljivo rješenje bez jamstva optimalnosti. Temeljna filozofija VNS-a počiva na ideji da lokalni minimum (optimum) u odnosu na jednu strukturu susjedstva ne mora nužno biti lokalni minimum (optimum) u odnosu na neku drugu strukturu. Shodno tome, VNS algoritam teži sustavnom istraživanju susjedstva (okolina) trenutnog rješenja, korištenjem skupa unaprijed definiranih struktura susjedstva. Budući različite instance problema imaju različit „ambijent“ i složenost, odabir prikladnih struktura predstavlja izazovan zadatak jer one izravno definiraju prostore rješenja – različite strukture susjedstva mogu dovesti do različitih prostora rješenja. Sustavnom izmjenom okolina tijekom procesa pretraživanja, algoritam uspješno „izranja“ iz lokalnih zamki i usmjerava se prema globalno optimalnom rješenju. U kontekstu konstrukcije reprezentativnog skupa ljestvica aminokiselina, VNS nam omogućuje učinkovito istraživanje prostora podskupova vrhova grafa, težeći pronalaženju onog podskupa koji zadovoljava uvjet dominantnosti uz minimalnu kardinalnost.

Učinkovitost VNS metaheuristike proizlazi iz sljedećih triju ključnih opažanja [2]:

1. Lokalni minimum (optimum) u odnosu na jednu strukturu susjedstva nije nužno lokalni optimum u odnosu na neku drugu strukturu susjedstva;
2. Globalni minimum je lokalni minimum u odnosu na sve moguće strukture susjedstva;
3. Za mnoge probleme, lokalni minimumi u odnosu na jednu ili više struktura susjedstva su relativno blizu jedni drugima.

Prvo opažanje nedvojbeno opravdava uvođenje više okolina. Zahvaljujući njemu VNS metaheuristika može efikasno izlaziti iz lokalnih minimuma. Kada se „zaglavi“ u lokalnome minimumu



Slika 2. Shema lokalne pretrage s više okolina (VNS)

ova metoda mijenja okolinu i istražuje druge dijelove prostora rješenja, odnosno omogućuje „bijeg“ iz lokalnih minimuma, ali i nastavljanje pretrage. Zadnje opažanje, inače empirijska činjenica, je od posebne važnosti jer sugerira da lokalni minimum pruža informaciju o globalnom minimumu (npr. oni mogu dijeliti zajedničke vrhove u grafu ljestvica). Drugim riječima, empirijski je potvrđena neophodnost sprovođenja sustavne pretrage okolina oko lokalnog minimuma (sve dok se ne pronade bolje rješenje). Način na koji se u okviru VNS-a razmatra više susjedstava tijekom lokalne pretrage (s f je označena funkcija cilja), ilustriran je Slikom 2.

Kako bi se ovi principi primijenili, osnovni VNS algoritam (engl. *Basic Variable Neighborhood Search*, BVNS) ciklički ponavlja tri faze: protresanje (engl. *shaking*), lokalnu pretragu (engl. *local search*) i sustavnu promjenu okoline (engl. *neighborhood change*). Pseudokod ovog algoritma [6] prikazan je Slikom 3. Pojednostavljeno, nakon što se definiraju konačan skup okolina N_k , $k = 1, \dots, k_{\max}$, koje će biti korištene u pretrazi, nasumično početno rješenje x , i odredi kriterij zaustavljanja, pristupa se procesu protresanja, odnosno bira se nasumično rješenje x' iz prve okoline $N_1(x)$, a zatim se primjenjuje lokalna pretraga, odnosno postupak lokalnog pretraživanja, u nekoj okolini nasumičnog rješenja x' , koja ne mora nužno biti ista okolina kao ona koja se koristi u procesu protresanja (čak ne mora biti ni iste strukture). Postupkom lokalne pretrage određuje se (pronalaži) neki lokalni minimum x'' . U tom trenutku moguća su tri ishoda:

1. Ako je $x = x''$ (odnosno, ako smo na dnu iste doline) primjenjuje se lokalna pretraga na sljedeću okolinu rješenja x , tj. $N_2(x)$;
2. Ako je $x \neq x''$ i $f(x'') \geq f(x)$, pronađeni lokalni minimum je lošiji od prethodnog, pa se u ovom ishodu postupak također ponavlja za $N_2(x)$;
3. Ako je $x \neq x''$ i $f(x'') < f(x)$, novopronađeni lokalni minimum je bolji od prethodnog, te se pretraga centrira oko x'' i ista počinje ponovo od prve okoline.

Ako se dođe do zadnje okoline, odnosno kada se dostigne k_{\max} , bez da se pronade bolje rješenje od (nasumičnog) početnog, iterativni postupak protresanja i lokalne pretrage se ponavlja od početka (postavlja se $k = 1$), dok ne bude zadovoljen kriterij zaustavljanja (npr. maksimalno dozvoljeno vrijeme izvršavanja procedure, maksimalan ukupan broj iteracija, maksimalan broj iteracija između dvije

Vrdoljak, A.,

Ilustrativan prikaz sinergije teorije grafova i analize peptida kroz kombinatornu optimizaciju

Algorithm 1 PSEUDOKOD OSNOVNE PRETRAGE PROMJENJIVOM OKOLINOM**Require:** k_{\max} = redni broj najšire okoline u kojoj se vrši protresanje;Susjedstva = konačan skup okolina \mathcal{N}_k , $k = 1, \dots, k_{\max}$;**Ensure:** Optimalno rješenje;1: *Inicijalizacija:* $x \leftarrow$ NasumičnoPočetnoRješenje();

▷ iz prostora prihvatljivih rješenja

Odrediti kriterijZaustavljanja;

2: **procedure** BVNS(Susjedstva, x , k_{\max} , kriterijZaustavljanja)3: **repeat**4: $k \leftarrow 1$;5: **repeat**6: *protresi:* generiraj točku x' nasumično iz okoline $\mathcal{N}_k(x)$;7: *lokalna pretraga:* primijenite postupak lokalnog pretraživanja počevši od rješenja x' kako biste pronašli rješenje x'' ;8: **if** (x'' je bolje od x) **then**9: $x \leftarrow x''$;▷ centriraj pretragu oko x'' 10: $k \leftarrow 1$;

▷ započni pretragu ponovo od prve okoline

11: **else**12: $k \leftarrow k + 1$;

▷ proširi okolinu

13: **end if**14: **until** $k = k_{\max}$ 15: **until** nije ispunjen kriterijZaustavljanja16: **return** Optimalno rješenje;17: **end procedure**

Slika 3. Pseudokod najrasprostanjenije inačice VNS-a

popravke/poboljšanja globalno najboljeg rješenja, maksimalan broj ponavljanja najboljeg rješenja i sl.). Dakle, može se zaključiti kako algoritam osnovne pretrage promjenjivom okolinom ne prati unaprijed zadanu „putanju” nego istražuje različite okoline trenutno najboljeg rješenja i prelazi u novo samo u slučaju kada naiđe na bolje rješenje od trenutnog. Pomjeranjem do rješenja koje se nalazi relativno „daleko” od trenutnog rješenja, postiže se sustavno pretraživanje prostora prihvatljivih rješenja. Zadržavanjem u trenutno najboljem rješenju, u slučaju kada lokalno pretraživanje nakon protresanja nije odvelo do boljeg rješenja, smanjuje se mogućnost nepotrebnog „lutanja” u prostoru prihvatljivih rješenja. Nasumičnim generiranjem početnog rješenja u nekoj okolini \mathcal{N}_k , osigurava se pretraživanje različitih područja prilikom sljedećeg razmatranja te okoline.

Na koncu ovog poglavlja, dat ćemo kratak osvrt na interesantan matematički okvir za model redukcije na dominantan skup u grafu, te dodatno za pronalaženje minimalnog dominantnog skupa s pomoću metode pretrage promjenjivom okolinom. Predmetni model za početni neprazan skup vrhova grafa uzima skup svih ljestvica aminokiselina. Nadalje, za proizvoljan nenegativan realan broj $r_0^2 \in [0,1]$, tzv. prag sličnosti, početni skup bridova grafa je skup svih onih bridova kojima su oba kraja iz skupa svih ljestvica, uz uvjet da je kvadrat koeficijenta korelacije r^2 između tih krajeva (vrhova), odnosno ljestvica, veći ili jednak r_0^2 . Zadaću pronalaženja minimalnog dominantnog skupa ljestvica ispravno rješavamo VNS-om, uz bitno jamstvo, tj. da za svaku „izbrisanu” ljestvicu postoji još jedna visoko korelirana ljestvica sačuvana u smanjenom skupu. Za potrebe ovog istraživanja i testiranje predloženog modela, korišten je skup podataka preuzet iz otvorene AAindex baze, koja donosi goleme količine numeričkih indeksa koji predstavljaju različita fizikalno-kemijska i biokemijska svojstva aminokiselina, i koja je dostupna pri mrežnoj stranici [11]: <https://www.genome.jp/aaindex/>.

Vrdoljak, A.,

Ilustrativan prikaz sinergije teorije grafova i analize peptida kroz kombinatornu optimizaciju

4. ANALIZA RAČUNALNE EFIKASNOSTI I BIOLOŠKE RELEVANTNOSTI REZULTATA

Eksperimentalna evaluacija predloženog modela provedena je na skupu od 507 ljestvica aminokiselina preuzetih iz otvorene AAindex baze podataka [11], [5]. Za potrebe diskretizacije i konstrukcije neutuženog i jednostavnog grafa odabrana su dva praga sličnosti $r_0^2 = 0,81$ i $r_0^2 = 0,64$. Optimizacija je provedena korištenjem javno dostupnog softverskog rješenja (programa) SCALEZIP (izbornik aminokiselinskih ljestvica), pri čemu je izvršeno novih 50 neovisnih pokretanja (replikacija) algoritma kako bi se ispitala robusnost rješenja za oba odabrana praga sličnosti. SCALEZIP program (besplatan i izrađen na engleskom jeziku) je u prepakiranome obliku (ZIP datoteka) dostupan za preuzimanje pri sljedećoj URL adresi [12]: <https://tinyurl.com/4yvjd3we>. Njegova implementacija je detaljno opisana u pripadajućoj doktorskoj disertaciji [10].

Novi rezultati (osnovna statistika podataka prikazana je Tablicom 1.) također pokazuju iznimnu stabilnost modela: u svih 50 eksperimenata za $r^2 \geq 0,81$ algoritam (s odabranom opcijom k-means) je također konvergirao prema dominantnom skupu kardinalnosti 263. Iako je veličina rezultirajućeg podskupa bila konstantna, uočen je varijabilitet u samom sastavu ljestvica unutar skupa. Ova pojava još jednom sugerira postojanje višestrukih, međusobno ekvivalentnih rješenja koja podjednako učinkovito minimiziraju redundanciju uz zadržavanje informacijske pokrivenosti izvornog grafa. Redukcija od približno 48% (s 507 na 263 ljestvice) pri strožem pragu ukazuje na značajan stupanj preklapanja u postojećim fizikalno–kemijskim i biokemijskim indeksima. Ponovo, još su zanimljiviji rezultati koje smo dobili u svakom od 50 eksperimenata za $r^2 \geq 0,64$ (s istom odabranom opcijom). U ovom slučaju, naši su novi izračuni također završili skupovima različite kardinalnosti (prosjek je bio 150,19, a standardna devijacija je bila približno 0,511). Još jednom, u oba slučaja, VNS metaheuristika se pokazala kao korisna metoda jer je smanjila broj ljestvica za najmanje 10 u svakom slučaju, a ponekad čak i za više od 20 ljestvica.

Slično, uspoređujući rezultate s onima objavljenim u pripadajućem radu [9], važno je napomenuti da, iako repetitivnost rezultata pri $r^2 \geq 0,81$ sugerira moguću optimalnost broja 263, VNS kao heuristička metoda ne pruža formalni dokaz optimalnosti. Postoji mogućnost da prostor rješenja obiluje lokalnim minimumima iste vrijednosti u kojima algoritam završava, dok bi globalni minimum teorijski mogao biti niži.

Tablica 1. Osnovna statistika rezultata istraživanja.

	$r^2 \geq 0,81$	$r^2 \geq 0,64$
Broj eksperimenata:	50	50
Veličina početnog skupa:	507	507
Minimalna veličina optimalnog skupa:	263	149
Maksimalna veličina optimalnog skupa:	263	151
Aritmetička sredina (mean):	263	150, 19
SD:	0	0,511
Srednje vrijeme izvršenja algoritma:	8.5 min.	9.4 min.

5. ZAKLJUČNA RAZMATRANJA I SMJERNICE ZA BUDUĆA ISTRAŽIVANJA

Budući da je predmetni problem NP-potpun, formalni dokaz optimalnosti moguće je dobiti isključivo iscrpnim pretraživanjem prostora (brute-force), što je u ovom slučaju računski neisplativo. Stoga se primjenjuje metaheuristički algoritam koji nudi zadovoljavajuće rješenje u razumnom vremenu, iako ne jamči apsolutnu optimalnost. Shodno tome, u ovom je radu kroz ilustrativan prikaz sinergije između teorije grafova i analize peptida demonstrirana učinkovitost metaheuristike pretrage promjenjivom okolinom (VNS) u rješavanju problema redukcije dimenzionalnosti ljestvica aminokiselina. Transformacijom skupa od 507 ljestvica u jednostavan neuteženi graf i definiranjem problema kao potrage za minimumom dominantnim skupom, postignuta je značajna kompresija podataka bez gubitka ključnih informacijskih veza.

Rezultati dobiveni pri pragu sličnosti $r^2 \geq 0,81$ rezultirali su stabilnim podskupom od 263 ljestvice, čime je potvrđena iznimna robusnost predloženog modela. Posebno je značajno što ovi nalazi potvrđuju rezultate prethodnog istraživanja, dodatno osnažujući hipotezu o visokom stupnju redundancije u postojećim fizikalno–kemijskim i biokemijskim indeksima. Ovakva redukcija izravno olakšava rad istraživačima u području biofizike i bioinformatike, smanjujući računsko opterećenje algoritama za predviđanje svojstava peptida. Preciznije, predloženi model omogućuje bržu preselekciju ključnih deskriptora kod treniranja modela strojnog učenja, sprječava 'prenaučenost' (engl. *overfitting*) te olakšava interpretaciju bioloških rezultata jer se fokusira na informacijski najbogatije indekse. Zaključno, primjena VNS-a pokazala se kao elegantan i efikasan način za navigaciju kroz složene prostore biokemijskih podataka, otvarajući put prema sustavnijem i preciznijem modeliranju molekularnih mehanizama.

Znanstveni doprinos ovog rada očituje se u inovativnoj transformaciji biokemijskog problema u čisto graf-teorijski okvir, čime je omogućena primjena napredne metaheuristike VNS na problemu minimalnog dominantnog skupa ljestvica aminokiselina. Za razliku od prethodnih istraživanja koja se uglavnom oslanjaju na standardne statističke metode poput analize glavnih komponenti (PCA), predloženi model zadržava fizikalni integritet izvornih ljestvica, omogućujući preciznu eliminaciju redundancije uz očuvanje potpune biološke interpretabilnosti rezultata.

Također, bit će zanimljivo istražiti mogu li se slični pristupi generalizirane redukcije primijeniti i na druge kontekste, posebice na kompleksne mreže u stvarnom svijetu u kojima je računalno intenzivno modeliranje primarni izazov. Tehnike korištene u ovom radu mogle bi donijeti značajne prednosti u optimizaciji takvih sustava, otvarajući nove mogućnosti za učinkovitije upravljanje informacijskim tokovima u različitim znanstvenim i tehničkim disciplinama.

LITERATURA

1. Chartrand, G., Eroh, L., Harary, F. & Zhang, P.: How large can the domination numbers of a graph be?, *Australasian Journal of Combinatorics*, 2000, 21, str. 23–35, <https://ajc.maths.uq.edu.au/>
2. Hansen, P., Mladenović, N.: Variable neighborhood search. In: Glover, F. & Kochenberger, G. A. (eds) *Handbook of Metaheuristics*. Springer, New York, 2003, <https://doi.org/10.1007/b101874>
3. Haynes, T. W., Hedetniemi, S. T. & Slater, P. J.: *Fundamentals of Domination in Graphs*, Marcel Dekker, New York, 1998.
4. Karp, R. M.: Reducibility among Combinatorial Problems. In: Miller, R. E., Thatcher, J. W. & Bohlinger, J. D. (eds) *Complexity of Computer Computations. The IBM Research Symposia Series*. Springer, Boston, Massachusetts, 1972, str. 85–103, https://doi.org/10.1007/978-1-4684-2001-2_9

Vrdoljak, A.,

Ilustrativan prikaz sinergije teorije grafova i analize peptida kroz kombinatornu optimizaciju

5. Kawashima, S., Kanehisa, M.: AAindex: amino acid index database, *Nucleic Acids Res.*, 2000, 28, str. 374, <https://doi.org/10.1093/nar/28.1.374>
6. Martí, R., Pardalos, P. M. & Resende, M. G. C.: *Handbook of Heuristics*, Springer, Cham, 2018.
7. Mladenović, N., Hansen, P.: Variable neighborhood search, *Comput. Oper. Res.*, 1997, 24, str. 1097–1100, [https://doi.org/10.1016/S0305-0548\(97\)00031-2](https://doi.org/10.1016/S0305-0548(97)00031-2)
8. Sedlar, J.: Gornje i donje ograde binomnih invarijanti grafova, disertacija, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, 2009.
9. Vrdoljak, A., Vukičević D.: Selector of Amino-Acid Scales Set, *Math. Med. Biol.*, 2024, 41 (3), str. 157–168, <https://doi.org/10.1093/imammb/dqae007>
10. Vrdoljak, A.: Automati grafova, disertacija, Univerzitet u Sarajevu, Prirodno-matematički fakultet, 2025.
11. AAindex: Amino acid index database. Dostupno na: <https://www.genome.jp/aaindex/>, (Pristupljeno 2026-04-15)
12. PhD Project Corner: ScaleZip Software Solution. Dostupno na: <https://tinyurl.com/4yvdj3we>, (Pristupljeno 2026-04-15)
13. Weisstein, E. W.: Minimal Dominating Set. From MathWorld—A Wolfram Web Resource. <https://mathworld.wolfram.com/MinimalDominatingSet.html>, (Pristupljeno 2026-03-02)