

Matematički fakultet  
Univerzitet u Beogradu

# Dizajn prajmera za PCR

## Genetski algoritam i Pareto optimizacije

**Mentor:**

Asistent: Stefan Kapunac

**Profesor:**

Vladimir Filipović

**Studenti:**

Anja Milutinović 235/2021

Lana Matić 143/2021

Smer: informatika

Datum: Septembar 2025.

# SADRŽAJ

<b>1. Uvod.....</b>	<b>2</b>
1.2. Dizajniranje prajmera.....	3
<b>2. Genetski algoritam.....</b>	<b>4</b>
2.1. Jednostavni problem.....	5
2.1.1. Jedinka.....	5
2.1.2. Fitness.....	5
2.1.3. Selekcija.....	7
2.1.4. Ukrštanje.....	7
2.1.5. Mutacija.....	7
2.2. Multiplex problem.....	8
2.2.1. Pareto optimizacija i NSGA-II.....	8
Osnovna ideja NSGA-II:.....	9
Prednosti ovog pristupa:.....	9
2.2.2. Multiplex jedinka.....	9
2.2.3. Fitness.....	9
• Definisanje ciljeva (svi se minimizuju).....	11
2.2.4. Selekcija.....	11
2.2.5. Ukrštanje.....	12
2.2.6. Mutacija.....	12
2.3. Eksperimentalni scenariji.....	12
2.3.1. Jednostavni problem - primer.....	12
2.3.2. Multiplex problem - primer.....	13
3. Diskusija.....	16
<b>4. Zaključak.....</b>	<b>17</b>
5. Literatura.....	18

# 1. Uvod

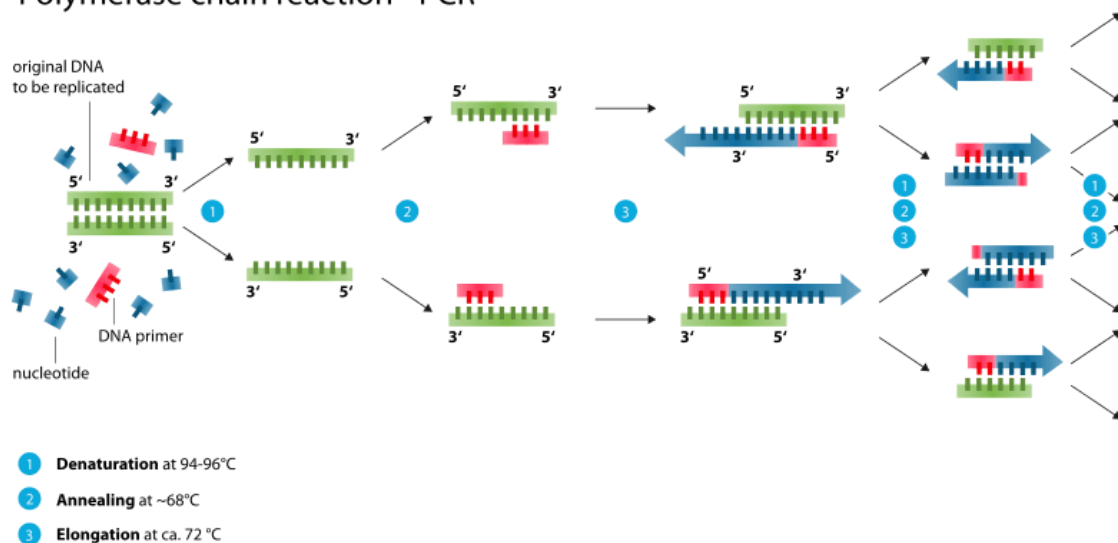
PCR (eng. Polymerase Chain Reaction) je laboratorijska tehnika za umnožavanje specifične DNK sekvence. Kary Mullis je predstavio PCR 1985. godine i kasnije za ovaj doprinos dobio Nobelovu nagradu za hemiju.

Ova tehnika umnožava specifične sekvence DNK ponavljanjem ciklusa:

1. Denaturacija - DNK se zagreva na 95°C kako bi se raskinule vodonične veze između komplementarnih baznih parova dvolančanog molekula;
2. Kaljenje (aniling) - hlađenje denaturisane DNK na temperaturu koja se obično kreće od 55°C do 72°C, omogućavajući prajmerima da se vežu za svoje komplementarne sekvence;
3. Elongacija - temperatura od 75°C do 80°C optimizuje aktivnost Taq polimeraze, termostabilna DNK polimeraza - enzim izolovan iz bakterije *Thermus aquaticus*, i podstiče izduživanje lanca - sinteza DNK

Ovaj ciklus se ponavlja 30–40 puta, što dovodi do eksponencijalnog umnožavanja ciljanog DNK fragmenta i omogućava dobijanje miliona kopija iz samo nekoliko molekula.

## Polymerase chain reaction - PCR



Slika 1. Proces reakcije lančanog umnožavanja - PCR

Zašto je PCR bitan?

1. U biologiji i biotehnologiji PCR omogućava analizu gena, sekvenciranje DNA, kloniranje i gentsko inženjerstvo.
2. U medicini se koristi za dijagnostiku bolesti, detekciju virusa i bakterija, kao i za otkrivanje tumorskih mutacija.
3. U forenzici omogućava identifikaciju osoba iz tragova DNK, dok se u ekologiji koristi za praćenje biodiverziteta putem eDNK.

4. U agronomiji se primenjuje za otkrivanje biljnih bolesti i testiranje genetski modifikovanih organizama (GMO).

## 1.2. Dizajniranje prajmera

Ključni element PCR-a su **prajmeri**, kratke veštačke sekvence DNA koje se vezuju na specifično mesto na DNK u fazi kaljenja i određuju početne tačke sinteze DNK.

U svakom molekulu DNK svaki lanac ima usmernost, jer je polinukleotid napravljen od nukleotida koji se povezuju fosfodiesterским vezama uvek na isti način:

- 5' (pet-prim) kraj – to je kraj lanca gde slobodno stoji 5. ugljenikov atom šećera (deoksiriboze) i na njemu se nalazi slobodna fosfatna grupa ( $-PO_4$ ).
- 3' (tri-prim) kraj – to je kraj lanca gde slobodno stoji 3. ugljenikov atom šećera, i na njemu se nalazi slobodna hidroksilna grupa ( $-OH$ ).

U PCR-u DNA polimeraza uvek produžava u smeru **5'→3'** i može da krene samo sa slobodnog 3'-OH kraja koji obezbeđuje prajmer. Da bi se precizno umnožio tačno određeni segment (**amplikon**), potrebna su **dva prajmera** koja "zagređuju" cilj:

- **Forward (F)** po sekvenci je identičan (referentnoj) niti na mestu vezivanja, ali se vezuje na obrnuti komplement (*reverse-complement*) referentne niti, da bi polimeraza produžavala nadesno - postavlja levu granicu amplikona,
- **Reverse (R)** - po sekvenci je reverse-complement referentne niti na svom mestu vezivanja (vezuje se na referentnu nit), pa polimeraza od njegovog 3' kraja produžava nalevo - postavlja desnu granicu amplikona.

Oni su orijentisani jedan naspram drugog i "gledaju ka unutra", tako da polimeraza sa oba 3' kraja napreduje ka sredini ciljanog regiona.

Dok bi sa jednim prajmerom dobili linearno kopiranje samo jedne komplementarne niti, bez jasne desne (ili leve) granice, a iz toga sledi teško kontrolisana dužina i niža specifičnost, sa oba prajmera dobijamo eksponencijalnu amplifikaciju: od 3. ciklusa nastaje dominantan proizvod tačno definisane dužine (između 3' krajeva F i R), jer oba kraja sada služe kao šabloni za sledeće cikluse.

Kreiranje prajmera za PCR je izazovno jer se istovremeno moraju zadovoljiti brojni kriterijumi:

- prajmer mora imati odgovarajuću dužinu i sadržaj GC baznih parova,
- temperature topljenja ( $T_m$ ) forward i reverse prajmera treba da budu bliske,
- treba izbegavati formiranje sekundarnih struktura (hairpin, dimeri),
- specifičnost mora biti zagarantovana (vezivanje samo za ciljnu sekvencu).

Kod *multiplex PCR*-a kompleksnost dodatno raste jer više prajmerskih parova mora biti međusobno kompatibilno (npr. da ne formiraju cross-dimere).

Ovaj problem spada u NP-teške optimizacione probleme, pa se koriste heuristički pristupi kao što je **genetski algoritam (GA)** u kombinaciji sa **Pareto optimizacijom** za višekriterijumsko rešavanje.

## 2. Genetski algoritam

Genetski algoritam (GA) je heuristička tehnika pretrage i optimizacije inspirisana prirodnom evolucijom. Populacija mogućih rešenja prolazi kroz **selekciju**, **ukrštanje** i **mutaciju**, pri čemu svaka jedinka dobija **fitnes vrednost** koja odražava kvalitet rešenja. Tokom generacija, bolje prilagođene jedinke imaju veću verovatnoću da prenesu svoje „gene“, pa populacija postepeno konvergira ka optimalnim ili blizu-optimalnim rešenjima.

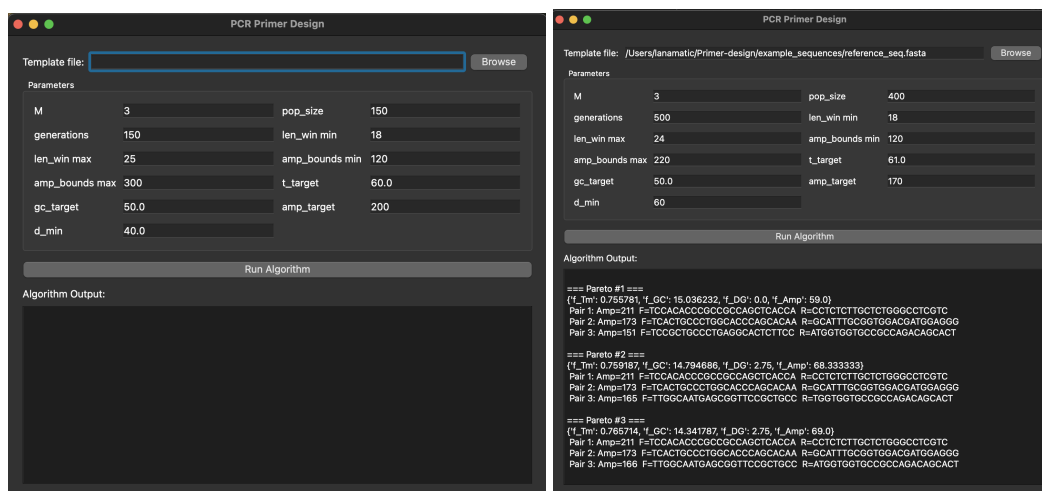
Tipični koraci GA su:

1. **Inicijalizacija** – nasumično stvaranje početne populacije.
2. **Evaluacija** – izračunavanje fitnes vrednosti za svaku jedinku.
3. **Selekcija** – izbor jedinki za reprodukciju na osnovu fitnesa.
4. **Ukrštanje** – kombinovanje genetskog materijala roditelja da bi se dobila nova rešenja.
5. **Mutacija** – unošenje manjih slučajnih promena radi očuvanja diverziteta.
6. **Zamena** – formiranje nove populacije i ponavljanje ciklusa dok se ne dostigne kriterijum zaustavljanja.

U ovom poglavlju prikazujemo kako se prajmeri mogu dizajnirati korišćenjem genetskog algoritma. Na početku razmatramo *jednostavan problem* sa jednom skaliranom fitnes funkcijom, a zatim prelazimo na *multiplex problem* gde se koristi višekriterijumska Pareto optimizacija. Ovaj dvostruki pristup pokazuje kako se isti evolucioni okvir može prilagoditi različitim nivoima složenosti zadatka.

Pored algoritamske implementacije razvijen je i grafički korisnički interfejs koji omogućava jednostavan rad sa DNK sekvencama i parametrima dizajna prajmera. Korisnik može da učitava DNK sekvencu u formatu *.txt* ili *.fasta*, da podesi ključne parametre (ciljna temperatura topljenja, rasponi dužina prajmera i amplicona, minimalni razmak između amplicona i dr.) i da jednim klikom pokrene *multiplex algoritam*. Nakon završetka optimizacije može da pregleda dobijeni Pareto skup optimalnih prajmerskih setova.

Na dijagramu (Slika 1) može se prikazati čitav tok rada – od učitavanja ulaznog fajla, preko parametarskog podešavanja, do evolutivnog računanja i prikaza rezultata.



Slika 1. Grafički korisnički interfejs

## 2.1. Jednostavni problem

Jednostavan problem podrazumeva dizajniranje **jednog para prajmera** za određeni region DNK. Procena kvaliteta kandidata zasniva se na jednoj skaliranoj fitnes funkciji, u kojoj se sabiraju pozitivni kriterijumi i oduzimaju penalizacije.

Iako se ovaj zadatak u principu može rešiti i algoritmima koji rade u polinomijalnom vremenu (jer tražimo samo jedan par prajmera i kriterijumi se mogu proveravati sekvencijalno), mi ga ovde namerno rešavamo genetskim algoritmom. Time pokazujemo kako se isti evolucioni okvir može primeniti i na jednostavne i na znatno složenije slučajeve, čime se uspostavlja prirodan most ka sledećem poglavlju o **multiplex problemu**, gde višekriterijumska Pareto optimizacija postaje neophodna.

Jednostavan problem pruža dobar uvid u način rada genetskog algoritma i postavlja osnovu za složenije višekriterijumske optimizacije u sledećem poglavlju.

### 2.1.1. Jedinka

Pošto su za PCR potrebna dva prajmera kako bi se precizno umnožio tačno određeni segment DNK-a (detaljnije u poglavlju *Dizajniranje prajmera*), jedinka je prajmerski par i opisana je četvorkom:

$$(f_{start}, f_{len}, r_{start}, r_{len})$$

gde su početne pozicije i dužine forward(f) i reverse(r) prajmera.

### 2.1.2. Fitness

Procena kvaliteta jedinke zasniva se na više bioloških i hemijskih kriterijuma koji su objedinjeni u jednu skaliranu fitnes funkciju. Svaki kriterijum se posebno ocenjujemo, a zatim vrednosti sabiramo uz ponderisane nagrade i kazne. Na taj način jedinka dobija jednu brojčanu vrednost koja predstavlja njen kvalitet i omogućava poređenje kandidata.

Kriterijumi za dobar prajmer su:

- **Dužina prajmera** - broj nukleotida prajmera. Optimalna dužina forward i reverse prajmera je **između 18 i 25 nukleotida**. Kraći prajmeri (<18 nt) obično su nestabilni i manje specifični, dok duži (>25 nt) mogu imati previsoku temperaturu topljenja i sklonost ka nespecifičnom vezivanju;
- **Dužina amplikona** - rastojanje između forward i reverse prajmera, tj. prostor koji "zatvaraju" 3' krajevi prajmera:

$$A = (r_{start} + r_{len}) - f_{start}$$

Za standardni PCR najbolji su amplikoni dužine **120–300 bp**. Amplikoni kraći od 100 bp mogu imati smanjenu specifičnost, dok oni duži od 500 bp otežavaju efikasnu amplifikaciju;

- **GC sadržaj** (gc\_content) - procenat G i C nukleotida u prajmeru. G–C parovi formiraju tri vodonične veze (jače vezivanje), dok A–T imaju samo dve. GC% utiče na

stabilnost prajmera i  $T_m$ . Dobri prajmer treba da ima udeo G i C nukleotida između **40% i 60%**, sa optimumom oko 50%. Nizak GC sadržaj (<40%) dovodi do slabog vezivanja, dok visok GC (>60%) može izazvati preveliku stabilnost i formiranje sekundarnih struktura;

- **Temperatura topljenja** ( $T_m$ ,  $tm\_wallace$ ) - temperatura na kojoj se polovina prajmera odvaja od komplementarne niti. Formula:

$$T_m = 2 \cdot (A + T) + 4 \cdot (C + G) .$$

Za stabilno vezivanje i efikasnu amplifikaciju,  $T_m$  oba prajmera treba da bude u opsegu **58–62 °C**, a razlika između forward i reverse prajmera ne sme biti veća od **2 °C**. Time se obezbeđuje da oba prajmera efikasno rade pri istoj temperaturi kaljenja u PCR-u.

- **GC clamp** ( $gc\_clamp\_score$ ) - prisustvo G ili C nukleotida na 3' kraju prajmera. Preporučuje se da u poslednjih 5 nukleotida na 3' kraju prajmera postoje **1–2 G/C nukleotida**. Ova tzv. GC „štikaljka“ stabilizuje vezivanje i poboljšava inicijaciju elongacije. Više od 3–4 GC u 3' repu, međutim, može stvoriti previše stabilno vezivanje i izazvati nespecifične produžetke.
- **Homopolimeri** ( $has\_homopolymer$ ) - dugi niz istih baza (npr. “AAAAA” ili “GGGG”). Dobri prajmeri ne sadrže duge nizove istih nukleotida. Homopolimeri duži od **3–4 nukleotida** dovode do grešaka polimeraze i smanjuju pouzdanost PCR-a.
- **3' komplementarnost** ( $overlap\_3p\_score$ ) - proverava da li se 3' krajevi prajmera mogu vezati za sebe ili za drugi prajmer (self- i cross-dimeri). Za svaku kombinaciju (F–F, R–R, F–R, R–F) traži se **najduže 3' poklapanje dužine 3–6 nt** sa segmentom obrnuti komplement partnera. Kazna raste s dužinom poklapanja i dodatno se uvećava za svaku G ili C bazu u tom delu, jer GC veze jače stabilizuju pogrešno vezivanje.
- **Hairpin strukture** ( $hairpin\_penalty$ ) - intramolekularno vezivanje prajmera sa samim sobom (npr. 3' kraj komplementaran sredini sekvence). Prajmeri ne smeju da formiraju stabilne ukosnice (hairpin). Ako je 3' kraj prajmera komplementaran nekom delu unutrašnje sekvence, može doći do samovezivanja i neuspešne amplifikacije. Kazna raste s dužinom poklapanja i brojem G/C baza, jer duži i GC-bogatiji hairpin značajno povećava rizik neuspešne amplifikacije.
- **Jedinstvenost k-mera** ( $primer\_uniqueness\_score$ ) - proverava koliko su podsekvence (k-meri) prajmera jedinstvene u šablonu. Prajmer treba da se vezuje samo na ciljnom mestu u genomu. To se proverava analizom k-mera (tipično  $k=10$ ); sekvence koje se javljaju više puta penalizuju se, jer smanjuju specifičnost i mogu izazvati amplifikaciju pogrešnih regiona.

Dobijamo konačnu formulu za fitness:

$$Fitness = w_{T_m} \cdot S_{T_m} + w_{GC} \cdot S_{GC} + w_{Len} \cdot S_{Len} + w_{Clamp} \cdot S_{Clamp} + w_{Amp} \cdot S_{Amp} + w_{Uniq} \cdot S_{Uniq} - (p_{Dimer} \cdot P_{Dimer} + p_{Hair} \cdot P_{Hair} + p_{Run} \cdot P_{Run}),$$

gde su:

- $S_{T_m}$  - skor usklađenosti temperature topljenja forward i reverse prajmera
- $S_{GC}$  - skor učešća GC baznih parova
- $S_{Len}$  - skor dužine prajmera

- $S_{Clamp}$  - skor GC clamp-a na 3' kraju
- $S_{Amp}$  - skor dužine amplicona
- $S_{Uniq}$  — skor jedinstvenosti prajmera

Kazne:

- $P_{Dimer}$  - penal ako postoji komplementarnost 3' krajeva između ili unutar prajmera (self- i cross-dimeri),  $P_{Hair}$  - penal ako prajmer formira hairpin strukturu. To su su realne vrednosti koje rastu proporcionalno dužini 3'-poklapanja (3–6 nt) i broju G/C baza u tom delu.
- $P_{Run}$  - penal ako prajmer sadrži homopolimere  $\geq 4$  nt

Težinski koeficijenti korišćeni u implementaciji su:

- $w_{Tm} = 1.0$ ,  $w_{GC} = 1.0$ ,  $w_{Len} = 0.6$ ,  $w_{Clamp} = 0.5$ ,  $w_{Amp} = 2.0$ ,  $w_{Uniq} = 0.8$ ,  $p_{Dimer} = 3.0$ ,  
 $p_{Hair} = 2.0$ ,  $p_{Run} = 1.5$

### 2.1.3. Selekcija

Za izbor roditeljskih jedinki koristi se **turnirska selekcija**. Algoritam nasumično izabere mali broj jedinki iz trenutne populacije (turnir veličine  $k$ ), a zatim za roditelja uzima onu sa najvećim fitnessom.

### 2.1.4. Ukrštanje

Za stvaranje novih jedinki koristi se **jednopoliziciono ukrštanje**, prilagođeno genima prajmerskog para.

Genetski zapis jedinke sadrži četiri broja: početak i dužinu forward prajmera i početak i dužinu reverse prajmera. Funkcija bira nasumičnu tačku preseka unutar genoma roditelja. Novi kandidat dobija prvu polovinu gena od roditelja p1, a drugu polovinu od roditelja p2. Ako nakon više pokušaja ukrštanja nije generisana validna jedinka, algoritam vraća boljeg od dva roditelja.

### 2.1.5. Mutacija

Mutacija uvodi **slučajne promene** u genetski zapis jedinke kako bi se očuvala raznovrsnost populacije i sprečilo prerano konvergiranje algoritma u lokalni optimum.

Za svaku od četiri genske komponente ( $f_{start}$ ,  $f_{len}$ ,  $r_{start}$ ,  $r_{len}$ ) generiše se slučajni broj; ako je on manji od zadate **verovatnoće mutacije**, komponenta se menja. Promene su male i kontrolisane:

- dužine forward i reverse prajmera se povećavaju ili smanjuju za  **$\pm 1$  nt**,
- startne pozicije se pomeraju nasumično do  **$\pm 5$  nukleotida**.



## 2.2. Multiplex problem

Multiplex PCR podrazumeva **istovremenu amplifikaciju više različitih regiona DNK** u jednoj PCR reakciji. Za svaki region potrebno je pronaći poseban par prajmera, a svi parovi moraju biti međusobno kompatibilni. To znači da se pored zahteva koje ispunjava svaki pojedinačni prajmer (dužina, GC sadržaj,  $T_m$ , odsustvo homopolimera, dimera i hairpin struktura) moraju uzeti u obzir i međusobne interakcije svih prajmerskih parova. Zbog ovih dodatnih ograničenja **problem višestruko raste u složenosti**:

- Broj mogućih kombinacija prajmerskih parova raste približno kvadratno sa brojem ciljanih regiona.
- Mora se obezbediti usklađenost temperatura topljenja među svim prajmerima.
- Neophodno je sprečiti cross-dimerizaciju između prajmera iz različitih parova.
- Poželjno je da se dužine amplicona razlikuju kako bi se proizvodi jasno razdvajali na gel elektroforezi ili u real-time PCR-u.

Za razliku od jednostavnog problema, gde se koristi jedna skalirana fitnes funkcija, ovde je teško ili nemoguće sve kriterijume svesti na jedan broj bez gubitka bitnih informacija. **Zato se primenjuje višekriterijumska Pareto optimizacija.**

**Pareto optimizacija** ne traži jedno „najbolje“ rešenje, već generiše **Pareto front** – skup rešenja koja nisu dominirana ni u jednom kriterijumu. Jedno rešenje dominira nad drugim ako je bar u jednom kriterijumu bolje, a ni u jednom lošije. Dobijamo više jednakovrednih kompromisa, a istraživač u laboratoriji može naknadno da izabere rešenje koje najbolje odgovara eksperimentalnim uslovima (npr. prednost manjih amplicona ili maksimalne homogenosti  $T_m$  vrednosti).

Primena genetskog algoritma sa Pareto optimizacijom omogućava:

- Istovremeno optimizovanje više ciljeva, bez potrebe za veštačkim ponderisanjem.
- Očuvanje diverziteta rešenja, što je važno da bi se dobio širok spektar potencijalnih prajmerskih setova.
- Prirodno proširenje istog evolutivnog okvira koji je korišćen za jednostavan problem, čime se pokazuje da metod može da se prilagodi različitim nivoima složenosti zadatka.

### 2.2.1. Pareto optimizacija i NSGA-II

Kod multiplex PCR-a za svako ciljano mesto u genomu potrebno je pronaći poseban par prajmera, a svi parovi moraju biti međusobno kompatibilni. Svako rešenje je zato opisano vektorom ciljeva (usklađenost  $T_m$ , ravnoteža GC sadržaja, minimalna energija dimerizacije, kontrolisana dužina amplicona).

Ovi ciljevi su često **konfliktni**: poboljšanje jednog može pogoršati drugi. Zbog toga nije dovoljno sabrati sve kriterijume u jednu vrednost, već je potrebna **višekriterijumska optimizacija**.

Za rešavanje ovog problema primenjen je **algoritam NSGA-II (Non-dominated Sorting Genetic Algorithm II)**, standard u multi-objective evolutivnim tehnikama.

Osnovna ideja NSGA-II:

- **Dominacija:** Rešenje A *dominira* nad rešenjem B ako je A nije lošije ni po jednom kriterijumu i bolje je bar po jednom.
- **Pareto front:** Skup rešenja koja nisu nadjačana ni u jednom kriterijumu. Prvi Pareto front sadrži rešenja najbolje ravnoteže svih ciljeva.
- **Crowding distance:** Meri gustinu rešenja oko jedinke unutar istog fronta. Veća udaljenost znači veći doprinos raznovrsnosti.

Prednosti ovog pristupa:

- **Višekriterijumska priroda:** Svi ciljevi (Tm, GC, energija dimera, dužina amplicona) optimizuju se istovremeno, bez ručnog ponderisanja.
- **Raznovrsnost rešenja:** Crowding distance osigurava da algoritam pokrije širok spektar kompromisnih rešenja.
- **Praktična fleksibilnost:** Istraživač može naknadno izabrati kompromisni set prajmera koji najbolje odgovara laboratorijskim uslovima.

### 2.2.2. Multiplex jedinka

U multiplex problemu jedinka predstavlja čitav **skup prajmerskih parova**, a ne samo jedan par kao u *jednostavnom problemu*. Svaki par je i dalje opisan četvorkom

$$(f_{start}, f_{len}, r_{start}, r_{len}),$$

ali se sada više takvih parova kombinuje u jedinstveni hromozom.

Osobine multiplex jedinke:

- Svi parovi moraju međusobno da se ne ometaju – ne smeju da formiraju cross-dimere niti da se međusobno preklapaju na nepoželjan način.
- Mora se voditi računa o globalnoj usklađenosti temperatura topljenja i o razlikama dužina amplicona, kako bi se PCR produkti jasno razdvajali pri analizi.
- U Pareto optimizaciji svaki cilj (npr. homogenost Tm ili minimalna energija potencijalnih dimera između različitih parova) posmatra se za celu jedinku kao skup.

### 2.2.3. Fitness

U multiplex okruženju svaki kandidat (jedinka) je skup više prajmerskih parova. Kvalitet kandidata opisujemo **vektorom ciljeva** (multi-objective), bez sabiranja u jedan skalar. U skladu sa NSGA-II, **svi ciljevi se minimizuju** (manje je bolje).

- Temperatura topljenja: **SantaLucia nearest-neighbor Tm.**  
Za razliku od jednostavnog Wallace pravila (linearni zbir  $2 \cdot (A+T) + 4 \cdot (G+C)$ ), u multiplexu koristimo termodinamički model najbližih suseda (nearest-neighbor model SantaLucia, 1998) jer:

1. uzima u obzir **dinukleotidne susede**,

2. računa **entalpiju** ( $\Delta H$ , kcal/mol) - mera totalne enrgije termodinamičkog sistema, i **entropiju** ( $\Delta S$ , cal/(mol·K)) - mera neuređenosti sistema, za cele oligonukleotidne duple spirale,
3. uključuje **konzentracioni termin** ( $C_t/4$  za slučaj ne-samokomplemenata) i **korekciju na soli**,
4. daje **kontinuiranije i realističnije  $T_m$**  vrednosti (bolje poravnate sa laboratorijskim uslovima), što je presudno kada više prajmera mora raditi na istoj temperaturi kaljenja.

U upotrebljenoj „lajt“ verziji modela, ukupne  $\Delta H$  i  $\Delta S$  računaju se sabiranjem doprinosa dinukleotida (NN tabela) i inicijalnih termina:

$$\Delta H = \Delta H_{init} + \sum_{i=1}^{n-1} \Delta H_{NN(s_i, s_{i+1})}, \quad \Delta S = \Delta S_{init} + \sum_{i=1}^{n-1} \Delta S_{NN(s_i, s_{i+1})},$$

gde su:

- $\Delta H$  - entalpija (kcal/mol) - ukupna promena entalpije - tj. ukupna toplota koja se oslobađa ili apsorbuje pri formiranju dvostruke spirale DNK.
- $\Delta S$  - entropija (cal/(mol·K)) - ukupna promena entropije - tj. mera promene reda/smerenosti sistema kada se jednolančani DNK upari u dvostruku spiralu.
- $\Delta H_{init}$  i  $\Delta S_{init}$  - inicijalni doprinosi, početni efekat krajeva lanca (5' i 3' kraja) i osnovna nukleacijska energiju potrebna da se započne formiranje dvostruke spirale.
- $\Delta H_{NN(s_i, s_{i+1})}$  i  $\Delta S_{NN(s_i, s_{i+1})}$  - energetski doprinos svakog dinukleotida (svakog para susednih baza  $s_i$  i  $s_{i+1}$ ) u lancu. Na primer, dinukleotid „GC“ ima jači (negativniji) doprinos entalpiji od „AT“, jer G–C parovi formiraju tri vodonične veze.

Zatim se  $T_m$  (u Kelvinima) procenjuje kao:

$$T_m^{(K)} = \frac{1000 \cdot \Delta H}{\Delta S + R \cdot \ln\left(\frac{C_t}{4}\right)} + 16.6 \cdot \log_{10}[Na^+],$$

gde je:

- R gasna konstanta = 1.987 cal/(mol·K)
- $C_t$  - ukupna koncentracija niti (za ne-samokomplemente uzima se  $\frac{C_t}{4}$ )
- poslednji sabirak je aproksimativna korekcija na monovalentnu so

Konačni  $T_m$  je u Celzijusima:

$$T_m^{(°C)} = T_m^{(K)} - 273.15,$$

Bolji je od Wallace-a u multiplexu, zato što preciznije hvata fine razlike u stabilnosti prajmera (sekvencijski kontekst, dužina, snagu-jona, koncentracije), pa omogućava tačniji raspored Tm vrednosti između parova i stabilnije zajedničko kaljenje.

- **Definisanje ciljeva (svi se minimizuju)**

Implementacija vraća četvorku (**f\_Tm, f\_GC, f\_DG, f\_Amp**). Svaki cilj je konstruisan tako da **manja vrednost znači bolje rešenje**.

- **f\_Tm** - usklađenost i homogenost Tm između parova. Računa se prosečno odstupanje srednjeg Tm po paru od ciljane vrednosti (npr. 60 °C) plus mera rasipanja (varijansa) između parova. Manje je bolje: parovi imaju Tm blizu targeta i međusobno su homogeni.
- **f\_GC** – globalna GC ravnoteža preko svih prajmera (F i R). Računa se prosečno odstupanje GC% svake sekvence od cilja (npr. 50%). Manje je bolje: GC je u poželjnom opsegu i ujednačen.
- **f\_DG** – dimeri/hairpin, „ΔG-like“ kazne (intra + inter). Sabrana „energetska“ penalizacija iz:
  - **intra-pair** doprinosa po paru (hairpin/self-dimer procene)
  - **globalnih cross-dimer** interakcija između **svih prajmera u jedinci** (svaki sa svakim, oba smeri)

Manje je bolje: manje i slabije (posebno 3'-GC-bogate) komplementarnosti niži rizik lažne elongacije.

- **f\_Amp** – ciljna dužina i razmak amplicona. Kombinuje:
  - **prosečno odstupanje** svake dužine amplicona od ciljane vrednosti i
  - **penal** ako je **najmanji međusobni razmak** dužina amplicona manji od npr. 40 bp, što otežava razdvajanje na gelu/analizi

Manje je bolje: ampliconi su blizu željene dužine i dovoljno razmaknuti međusobno.

## 2.2.4. Selekcija

Kod višekriterijumske Pareto optimizacije selekcija mora istovremeno da:

- favorizuje rešenja sa boljim Pareto rangom, i
- održava raznovrsnost rešenja unutar istog fronta.

Za to se koristi turnirska selekcija zasnovana na Pareto rang i crowding distance-u. Za dve nasumično izabrane jedinke a i b prvo se poredi **Pareto rang**:

- manji rang znači bolji Pareto front (npr. rang 0 je najbolji front).
- jedinka sa nižim rangom automatski pobeđuje.

Ako su u istom Pareto frontu, poredi se **crowding distance**:

- crowding distance meri prostornu udaljenost od suseda u prostoru ciljeva;

- veća vrednost znači da je rešenje manje „zbijeno“, tj. doprinosi većoj raznovrsnosti Pareto fronta.
- prednost dobija jedinka sa većom crowding distance.

### 2.2.5. Ukrštanje

Kod multiplex jedinki ukrštanje mora da kombinuje celine prajmerskih parova, a ne pojedinačne nukleotide. Koristimo **jednopoloziciono ukrštanje preko liste parova**. Ukrštanje se dozvoljava samo ako oba roditelja koriste istu sta referentnu sekvencu, što sprečava nastanak nelogičnih kombinacija.

Roditelji predstavljaju listu parova. Bira se nasumična tačka preseka i dete dobija prefiks parova iz prvog roditelja i sufiks iz drugog roditelja. Time se čuvaju koherentni parovi (F/R zajedno), umesto da se mešaju geni unutar para. Ako nakon više pokušaja ukrštanja nije generisana validna jedinka, algoritam vraća boljeg od dva roditelja.

### 2.2.6. Mutacija

Kod multiplex jedinki mutacija se primenjuje **na nivou jednog izabranog para prajmera** (lokalna mutacija), čime se pravi mala, kontrolisana promena u skupu bez narušavanja ostatka rešenja.

Nasumično se bira jedan prajmerski par iz jedinice, a zatim se nad tim parom poziva već definisana *mutacija za jednostavan problem*.

U kombinaciji sa ukrštanjem, ovakva mutacija omogućava fino podešavanje pojedinačnih parova u okviru višestrukog PCR seta, što poboljšava ukupnu raznovrsnost Pareto fronta.

## 2.3. Eksperimentalni scenariji

U ovom poglavlju predstavljamo praktične primere rada sistema i upoređujemo rezultate dobijene na jednostavnom problemu i na multiplex problemu.

Cilj je da se pokaže kako razvijeni genetski algoritam i Pareto optimizacija funkcionišu i na realnim DNK sekvencama.

### 2.3.1. Jednostavni problem - veštački primer

Kao ulaz u algoritam korišćena je veštački proširena DNK sekvencu:

ATGCGTACGTTGACCTGATCGATCGGATCCGATGCTAGCTAGCTAGGCTTACGATCGATCG  
(ponovljeno 6 puta)

Ova sekvencu je poslužila kao šablon na kome se traže odgovarajući prajmeri. Parametri genetskog algoritma su podešeni tako da se generiše populacija od 120 mogućih parova prajmera, a proces evolucije se odvija tokom 120 generacija. Na samom početku svaki kandidat (individualni par prajmera) nasumično se formira tako da forward i reverse prajmer zadovoljavaju osnovne uslove dužine (između 18 i 25 nukleotida) i daju amplicon u opsegu od 120 do 300 baznih parova.

Nakon što je algoritam izvršio 120 generacija evolucije, izdvojeni su najbolji prajmeri. Najbolje rešenje imalo je sledeće karakteristike:

- **Forward prajmer:** CTAGCTAGCTAGGCTTACG, startna pozicija 34, dužina 19 bp
- **Reverse prajmer:** ATCCGATCGATCAGGTCAA, startna pozicija 314, dužina 19 bp
- **Amplicon:** dužina 299 bp
- **Fitness vrednost:** 5.508

Ovaj rezultat pokazuje da algoritam uspeva da pronađe par prajmera koji se nalaze na udaljenosti od približno 300 baznih parova, što se uklapa u zadate granice amplicona. Dužina oba prajmera (19 bp) nalazi se u optimalnom rasponu, a njihov nukleotidni sastav je takav da obezbeđuje stabilno vezivanje bez prevelike dominacije GC parova.

### 2.3.2. Multiplex problem - veštački primer

U ovom primeru analizirani su rezultati pokretanja NSGA-II optimizatora nad šablonom koji predstavlja ponovljenu DNK sekvencu:

ATGCGTACGTTGACCTGATCGATCGGATCCGATGCTAGCTAGCTAGGCTTACGATCGATCG \* 8

Sa parametrima:

- pop\_size = 150 (veličina populacije)
- generations = 150 (broj generacija evolucije)
- len\_win = (18, 25) (dozvoljena dužina prajmera u nukleotidima)
- amp\_bounds = (120, 300) (dozvoljena dužina amplicona u baznim parovima)
- t\_target = 60.0 °C (ciljna temperatura topljenja)  
gc\_target = 50 % (ciljni GC sadržaj)
- amp\_target = 200 bp (ciljna prosečna dužina amplicona)
- d\_min = 40 bp (minimalni dozvoljeni razmak između amplicona)

Kao izlaz dobijen je Pareto front rešenja:

### Pareto #1 — ključne vrednosti i proračun

f\_Tm = 3.883075

f\_GC = 5.694444

f\_DG = 0.0

f\_Amp = 65.333333

Ampliconi: 156 bp, 147 bp, 206 bp

Parovi (skraćeno):

- Amp = 156 — F = GCGTACGTTGACCTGATCGATCGGA, R = AGCATCGGATCCGATCGATCAGGTC
- Amp = 147 — F = TGACCTGATCGATCGGATCCGATGC, R = GCATCGGATCCGATCGATCAGGTCA
- Amp = 206 — F = ACCTGATCGATCGGATCCGATGCT, R = GCATCGGATCCGATCGATCAGGTCA

#### Proračun f\_Amp:

Apsolutna odstupanja od cilja (200 bp):

- $|156 - 200| = 44$
  - $|147 - 200| = 53$
  - $|206 - 200| = 6$
- Zbir =  $44 + 53 + 6 = 103$ . Srednja devijacija =  $103 / 3 = 34.3333333333$ .

Međusobni razmaci amplicona:

- $|156 - 147| = 9$
  - $|156 - 206| = 50$
  - $|147 - 206| = 59$
- Minimalni gap =  $9 < d_{\min} (40)$ , pa spacing kazna =  $40 - 9 = 31$ .

Ukupno f\_Amp =  $34.3333333333 + 31 = 65.3333333333$

**Tumačenje:** f\_DG = 0.0 znači da po proveru nema značajnih 3'-komplementarnosti ili hairpina — to je dobra vest, jer ukazuje na nisku verovatnoću prajmer-dimer formiranja. Međutim, glavni nedostatak ovog rešenja je ekstremno mali razmak (9 bp) između prva dva amplicona (156 i 147 bp). U praksi to znači da su proizvodi preblizu da bi se pouzdano razdvojili na standardnoj agaroznoj gelelektroforezi. Dodatno, u ovom rešenju se primećuje da su neki reverse prajmeri identični za više parova (GCATCGGATCCGATCGATCAGGTCA), što može dovesti do međusobne konkurencije i neočekivanih produkata u multiplex postavci.

**Zaključak za Pareto #1:** Termodinamički bezbedan (nema dimer rizika), ali neprihvatljiv zbog razmaka i ponavljanja istih sekvenci. Zahteva restrukturiranje pozicija prajmera.

#### Pareto #2 — ključne vrednosti i proračun

f\_Tm = 3.959025

f\_GC = 5.694444

f\_DG = 7.5

f\_Amp = 46.0

Ampliconi: 152 bp, 173 bp, 206 bp

Parovi (skraćeno):

- Amp = 152 — F = GCGTACGTTGACCTGATCGATCGGA, R = TCGGATCCGATCGATCAGGTCAACG
- Amp = 173 — F = GGCTTACGATCGATCGATGCGTACG, R = GCATCGGATCCGATCGATCAGGTCA
- Amp = 206 — F = ACCTGATCGATCGGATCCGATGCT, R = GCATCGGATCCGATCGATCAGGTCA

#### Proračun f\_Amp:

Apsolutna odstupanja:

- $|152 - 200| = 48$
  - $|173 - 200| = 27$
  - $|206 - 200| = 6$
- Zbir =  $48 + 27 + 6 = 81$ . Srednja devijacija =  $81 / 3 = 27.0$ .

Međusobni razmaci:

- $|152 - 173| = 21$
  - $|152 - 206| = 54$
  - $|173 - 206| = 33$
- Minimalni gap = 21. Spacing kazna =  $40 - 21 = 19$ .

Ukupno f\_Amp =  $27.0 + 19 = 46.0$ .

**Tumačenje:** Ovo rešenje je najbolje od tri po pitanju usklađenosti sa ciljem amplicona (najmanja srednja devijacija: 27 bp). Međutim, f\_DG = 7.5 ukazuje na prisutnost značajnijih termodinamičkih problema (mogući prajmer-dimer ili hairpin signali). U laboratorijskoj primeni to može dovesti do pojave prajmer-dimer traka i smanjene efikasnosti amplifikacije. Minimalni gap = 21 bp je bolji nego u Pareto #1, ali i dalje ispod zahtevanih 40 bp, što znači da i ovde postoji rizik interferencije među ampliconima.

**Zaključak za Pareto #2:** Najbolji izbor ako prioritet predstavlja blisko pogodjenje ciljnih dimenzija amplicona, ali zahteva korekciju prajmera zbog termodinamičkog rizika (f\_DG).

### Pareto #3 — ključne vrednosti i proračun

f\_Tm = 4.081193

f\_GC = 5.027778

f\_DG = 0.0

f\_Amp = 60.0

Ampliconi: 160 bp, 147 bp, 206 bp

Parovi (skraćeno):

- Amp = 160 - F = GCGTACGTTGACCTGATCGATCGGA, R = AGCTAGCATCGGATCCGATCGATCA



- Amp = 147 - F = TGACCTGATCGATCGGATCCGATGC, R = GCATCGGATCCGATCGATCAGGTCA
- Amp = 206 - F = ACCTGATCGATCGGATCCGATGCT, R = GCATCGGATCCGATCGATCAGGTCA

#### Proračun f\_Amp:

Apsolutna odstupanja:

- $|160 - 200| = 40$
  - $|147 - 200| = 53$
  - $|206 - 200| = 6$
- Zbir =  $40 + 53 + 6 = 99$ . Srednja devijacija =  $99 / 3 = 33.0$ .

Međusobni razmaci:

- $|160 - 147| = 13$
  - $|160 - 206| = 46$
  - $|147 - 206| = 59$
- Minimalni gap = 13. Spacing kazna =  $40 - 13 = 27$ .

Ukupno f\_Amp =  $33.0 + 27 = 60.0$ .

**Tumačenje:** Pareto #3 ima f\_DG = 0.0 i najbolji f\_GC ( $\approx 5.03$ ) od svih tri, što znači da su GC% prajmera najbliže ciljnih 50 % i nema uočenih 3'-komplementarnosti u trenutnoj proceni. To ga čini bezbednijim u pogledu prajmer-dimer rizika. Ipak, minimalni gap 13 bp i prosečna devijacija amplicona 33 bp dovode do ukupno lošijeg f\_Amp nego kod Pareto #2. f\_Tm = 4.08 ukazuje da Tm metrički nije idealno usklađen i da postoji prostor za fino podešavanje.

**Zaključak za Pareto #3:** Najbezbednije rešenje iz perspektive termodinamike i GC balansa, ali neprihvatljivo bez korekcije razmaka između amplicona.

Pareto	f_Tm	f_GC	f_DG	f_Amp	Najmanj i razmak (bp)	Glavna napomena
#1	3.883	5.694	0.0	65.33	9	Termodinamički čist (nema dimera), ali <b>ekstremno mali razmak</b> 156↔147 (9 bp) – nepouzdan razdvoj na gelu; ponavljanje R prajmera između parova.
#2	3.959	5.694	7.5	46.00	21	Najbliže cilju dužine (najmanja srednja devijacija), <b>ali</b> prisutni dimer/hairpin signali (f_DG=7.5) i razmak < 40 bp.
#3	4.081	5.028	0.0	60.00	13	Ponovo bez dimera (f_DG=0), dobar GC balans, ali <b>najmanji razmak 13 bp</b> → i dalje ispod kriterijuma d_min=40 bp.

Tabela 1. Poređenje sva tri Pareto rešenja

### 2.3.3. Multiplex problem - realni primer (Homo sapiens ACTB, mRNA)

Kao biološki relevantan test korišćen je FASTA fajl

**Homo sapiens actin beta (ACTB), mRNA** – GenBank oznaka *NM\_001101.5*.

#### Biološka pozadina:

Aktin je osnovni protein ćelijskog skeleta (citoskeleta). Gradi dugačka, tanka vlakna (mikrofilamente) koja:

- daju ćeliji oblik i mehaničku čvrstinu,
- omogućavaju kretanje ćelije i intracelularni transport,
- učestvuju u deobi ćelije.

Beta-aktin ( $\beta$ -aktin) je jedna od nekoliko izoformi aktina i prisutan je u gotovo svim tipovima ćelija kod čoveka. Gen **ACTB** sadrži DNK instrukciju za sintezu  $\beta$ -aktina, pa je čest **referentni gen** u kvantitativnim PCR eksperimentima.

Eksperimentalna postavka:

- **Ulazni šablon:** sekvenca ACTB mRNA iz FASTA fajla.
- **Broj ciljanih regiona (M):** 3
- **Parametri genetskog algoritma (NSGA-II):**
  - **pop\_size** = 400 (veličina populacije)
  - **generations** = 500 (broj generacija evolucije)
  - **len\_win** = (18, 24) (dozvoljena dužina prajmera u nukleotidima)
  - **amp\_bounds** = (120, 220) (dozvoljena dužina amplicona u baznim parovima)
  - **t\_target** = 61.0 °C (ciljna temperatura topljenja)
  - **gc\_target** = 50 % (ciljni GC sadržaj)
  - **amp\_target** = 170 bp (ciljna prosečna dužina amplicona)
  - **d\_min** = 60 bp (minimalni dozvoljeni razmak između amplicona)

Ove vrednosti su podešene tako da odgovaraju **standardnim laboratorijskim uslovima** za dizajniranje prajmera na ljudskim cDNK sekvencama.

Kao izlaz dobijena je Pareto fronta rešenja:

## Pareto #1 — ključne vrednosti i proračun

f\_Tm = 0.755781  
f\_GC = 15.036232  
f\_DG = 0.0  
f\_Amp = 59.0

Ampliconi: 211 bp, 173 bp, 151 bp

Parovi (skraćeno):

- Amp = 211 bp - F: TCCACACCCGCCGCCAGCTCACCA, R: CCTCTCTTGCTCTGGGCCTCGTC
- Amp = 173 bp - F: TCACTGCCCTGGCACCCAGCACAA, R: GCATTTGCGGTGGACGATGGAGGG
- Amp = 151 bp - F: TCCGCTGCCCTGAGGCACTCTTCC, R: ATGGTGGTGCCGCCAGACAGCACT

**Proračun f\_Amp:**

- Apsolutna odstupanja:
  - |211 - 170| = 41
  - |173 - 170| = 3
  - |151 - 170| = 19

Zbir = 63 → srednja devijacija =  $63 / 3 = 21.0$

- Međusobni razmaci:

$$|211 - 173| = 38$$

$$|211 - 151| = 60$$

$$|173 - 151| = 22$$

$$\text{Minimalni gap} = 22 < d_{\text{min}} (60). \text{ Spacing kazna} = 60 - 22 = 38$$

$$\text{Ukupno } f_{\text{Amp}} = 21.0 + 38 = 59.0$$

**Tumačenje:**  $f_{\text{DG}} = 0.0$  znači da u ovom setu nema detektovanih prajmer-dimera ni hairpina, što je odlična vest. Glavni minus je preblizak razmak: između 173 bp i 151 bp ima samo 22 bp, što može otežati razdvajanje na gelu ili multiplex analizu. Amplicon 211 bp je značajno duži od cilja 170 bp i diže deo kazne.

**Zaključak za Pareto #1:** Termodinamički vrlo siguran, ali za multiplex je preporuka skratiti 211 bp ( $\approx 190$  bp) ili povećati razmak između 151 bp i 173 bp.

## Pareto #2 — ključne vrednosti i proračun

$$f_{\text{Tm}} = 0.759187$$

$$f_{\text{GC}} = 14.794686$$

$$f_{\text{DG}} = 2.75$$

$$f_{\text{Amp}} = 68.333333$$

Ampliconi: 211 bp, 173 bp, 165 bp

Parovi (skraćeno):

- Amp = 211 - F = `TCCACACCCGCCGCCAGCTCACCA`, R = `CCTCTCTTGCTCTGGGCCTCGTC`
- Amp = 173 - F = `TCACTGCCCTGGCACCCAGCACAA`, R = `GCATTTGCGGTGGACGATGGAGGG`
- Amp = 165 - F = `TTGGCAATGAGCGGTTCCGCTGCC`, R = `TGGTGGTGCCGCCAGACAGCACT`

### Proračun $f_{\text{Amp}}$ :

- Apsolutna odstupanja:

$$|211 - 170| = 41$$

$$|173 - 170| = 3$$

$$|165 - 170| = 5$$

$$\text{Zbir} = 49 \rightarrow \text{srednja devijacija} = 49 / 3 = 16.333$$

- Međusobni razmaci:

$$|211 - 173| = 38$$

$$|211 - 165| = 46$$

$$|173 - 165| = 8$$

$$\text{Minimalni gap} = 8 < d_{\min} (60). \text{ Spacing kazna} = 60 - 8 = 52$$

$$\text{Ukupno } f_{\text{Amp}} = 16.333 + 52 \approx 68.33$$

**Tumačenje:** Malo bolja centriranost oko 170 bp nego kod **#1** (niža srednja devijacija), ali lošiji minimalni razmak (8 bp).  $f_{\text{DG}} = 2.75$  ukazuje na blagu dimer/hairpin interakciju

**Zaključak za Pareto #2:** Dobre dužine, ali preblizak razmak i prisutni dimer signali čine ovo slabijim izborom od #1.

### Pareto #3 — ključne vrednosti i proračun

$$f_{\text{Tm}} = 0.765714$$

$$f_{\text{GC}} = 14.341787$$

$$f_{\text{DG}} = 2.75$$

$$f_{\text{Amp}} = 69.0$$

Ampliconi: 211 bp, 173 bp, 166 bp

Parovi (skraćeno):

- Amp = 211 - F = `TCCACACCCGCCGCCAGCTCACCA`, R = `CCTCTCTTGCTCTGGGCCTCGTC`
- Amp = 173 - F = `TCACTGCCCTGGCACCCAGCACAA`, R = `GCATTTGCGGTGGACGATGGAGGG`
- Amp = 166 - F = `TTGGCAATGAGCGGTTCCGCTGCC`, R = `ATGGTGGTGCCGCCAGACAGCACT`

#### Proračun $f_{\text{Amp}}$ :

- Apsolutna odstupanja:  
 $|211 - 170| = 41$   
 $|173 - 170| = 3$   
 $|166 - 170| = 4$   
 $\text{Zbir} = 48 \rightarrow \text{srednja devijacija} = 48 / 3 = 16.0$
- Međusobni razmaci:  
 $|211 - 173| = 38$   
 $|211 - 166| = 45$   
 $|173 - 166| = 7$   
 $\text{Minimalni gap} = 7 < d_{\min} (60). \text{ Spacing kazna} = 60 - 7 = 53$

$$\text{Ukupno } f_{\text{Amp}} = 16.0 + 53 = 69.0$$

**Tumačenje:** Slično **#2**, ali najmanji razmak (7 bp) → najveća spacing kazna.  $f_{DG} = 2.75$  znači isti slab dimer signal.

**Zaključak za Pareto #3:** Termodinamički prihvatljiv, ali najmanje pogodan za multiplex zbog minimalnog razmaka i dimer indikacije.

Pareto	f_Tm	f_GC	f_DG	f_Amp	Najmanji razmak (bp)	Glavna napomena
<b>#1</b>	0.76	15.04	0.0	59.0	22	Termodinamički vrlo siguran (nema dimera), ali najveći amplicon (211 bp) i preblizak razmak između 173 bp i 151 bp.
<b>#2</b>	0.76	14.79	2.75	68.3	8	Dobro centriran oko ciljne dužine (170 bp), ali prisutni dimer/hairpin signali i ekstremno mali razmak.
<b>#3</b>	0.77	14.34	2.75	69.0	7	Sličan #2, ali još manji razmak i iste blage dimer indikacije – najmanje pogodan za multiplex.

Tabela 2. Poređenje sva tri Pareto rešenja

### 3. Diskusija

Dobijeni rezultati pokazuju jasne razlike između jednostavnog genetskog algoritma (GA) i višeciljnog evolutivnog algoritma NSGA-II primenjenog za multiplex PCR dizajn.

#### **Jednostavni problem – jedan par prajmera**

U slučaju običnog GA, cilj je bio da se pronađe jedan par prajmera koji zadovoljava osnovne kriterijume: optimalnu dužinu, poželjnu temperaturu topljenja ( $T_m$ ), balans GC sadržaja i amplicon u definisanom rasponu. Rezultati pokazuju da algoritam uspešno pronalazi stabilan par prajmera bez izraženih problema sa dimerima i sekundarnim strukturama. Primer dobijenih prajmera generisao je amplicon dužine 299 bp, što je unutar definisanih granica, i može se smatrati zadovoljavajućim za standardni PCR eksperiment. Međutim, osnovni GA nije imao potrebu da rešava dodatne konflikte između više parova prajmera, pa je problem bio znatno jednostavniji.

#### **Multiplex problem – više parova prajmera**

NSGA-II algoritam je korišćen za složeniji zadatak, istovremeni dizajn tri nezavisna para prajmera za multiplex PCR. Ovaj pristup uvodi dodatne kriterijume, uključujući međusobnu razliku u veličinama amplikona (kako bi se oni jasno razdvojili na gel elektroforezi) i izbegavanje međusobnih interferencija između prajmera. Analiza Pareto fronta pokazuje da svaki od tri prikazana rešenja predstavlja kompromis između različitih ciljeva.

*U veštačkom primeru* Pareto #1 je bio termodinamički najstabilniji (nema dimera), ali je imao premale razmake između amplikona, dok su #2 i #3 bolje pogađali ciljne dužine, ali su trpeli od većeg rizika dimera ili nedovoljnih razmaka.

*U realnom primeru sa humanim ACTB genom* viđena je slična slika: Pareto #1 je pokazao vrlo sigurnu termodinamiku, ali i nedovoljan razmak između pojedinih amplikona; Pareto #2 je bio bliži ciljnim dimenzijama, ali je imao umeren rizik od dimerizacije; Pareto #3 je pokazao najmanju razdvojenost amplicona i blage dimer indikacije, čineći ga najmanje pogodnim za multiplex PCR.

#### **Uporedna analiza**

U poređenju sa običnim GA pristupom, NSGA-II je pokazao veću fleksibilnost i realističniju primenu, jer u stvarnim eksperimentalnim uslovima multiplex PCR gotovo nikada ne može da zadovolji sve kriterijume u potpunosti. Međutim, cena toga je znatno složenija evaluacija i interpretacija rezultata.

## 4. Zaključak

Jednostavan genetski algoritam pokazao se efikasnim za generisanje pojedinačnih prajmera u okviru standardnog PCR eksperimenta. Dobijeni par prajmera bio je stabilan, sa odgovarajućim karakteristikama i bez značajnih problema vezanih za sekundarne strukture. Ovaj pristup je brz, jednostavan i daje jedno konkretno rešenje koje se lako koristi u praksi.

Nasuprot tome, NSGA-II algoritam je omogućio višeciljno optimizovanje za multiplex PCR, gde se istovremeno traži više parova prajmera koji moraju zadovoljiti kompleksne uslove. Dobijena Pareto rešenja pokazuju da nije moguće pronaći „savršene“ prajmere za sve ciljeve odjednom, već se umesto toga formira spektar kompromisnih rešenja. To istraživaču daje mogućnost da, u zavisnosti od prioriteta (npr. stabilnost vs. razlika u amplitudama), odabere rešenje koje je najprikladnije za konkretan eksperiment.

Ovaj pristup predstavlja prvi korak u procesu dizajna prajmera: algoritamski generisani kandidati služe kao početna tačka koja značajno smanjuje broj potencijalnih kombinacija i vodi istraživača ka praktičnijim rešenjima. Dalje, svaki generisani prajmer mora biti laboratorijski testiran kako bi se potvrdila njegova funkcionalnost, specifičnost i efikasnost. Na osnovu tih rezultata, prajmeri se mogu po potrebi modifikovati, optimizovati ili odabrati najbolji iz skupa kompromisnih rešenja, čime se osigurava pouzdanost PCR eksperimenta u realnim uslovima.



## 5. Literatura

1. National Center for Biotechnology Information (NCBI). *Polymerase Chain Reaction (PCR) - Fundamentals and Applications*.  
<https://www.ncbi.nlm.nih.gov/books/NBK589663/>
2. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. *A fast and elitist multiobjective genetic algorithm: NSGA-II*. Computers & Operations Research, Volume 31, Issue 5, 2005, pp. 851–870. <https://www.sciencedirect.com/science/article/pii/S0377042705000774>
3. SantaLucia J. *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics*. PNAS, 1998.