

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Uloga bioinformatike u dizajniranju lijekova

Ivana Kajić

Voditelj: *Doc. dr. sc. Mile Šikić*

Zagreb, svibanj 2010.

Sadržaj

1	Uvod.....	3
2	Od aminokiseline do proteinskog kompleksa	4
3	Uvod u istraživanje lijekova	6
	3.1. Tradicionalni pristup	6
	3.2. Suvremeni pristup.....	6
4	Prvi koraci u dizajniranju lijekova	7
	4.1. Otkrivanje strukture receptora.....	8
	4.2. Otkrivanje strukture liganada	9
5	Virtual screening.....	10
	5.1. Dokiranje proteina	10
	Afinitet vezanja.....	11
	Scoring funkcije. Algoritmi pretraživanja	12
6	QSAR	14
	6.1. Matematička forma. Prve primjene.....	14
	6.2. 3D-QSAR	16
	CoMFA.....	16
7	Zaključak	18
8	Literatura.....	19
9	Sažetak.....	21

1 Uvod

Čovjekova potreba za istraživanjem i proizvodnjom lijekova stara je koliko i on sam. Borba za opstanak i preživljavanje prisilili su čovjeka da se suoči s bolestima kako bi svoj boravak na Zemlji učinio što ugodnijim i dugovječnijim. Već su se prvobitne zajednice okretale prirodi u potrazi za ljekovitim biljem koje bi obrađivali i koristili za liječenje bolesnih. Uloga liječnika pripadala je najstarijim iniciranim članovima koji su morali dobro poznavati biljke i njihova svojstva kako bi oboljelom mogli odrediti odgovarajuću „terapiju“. Prekretnicu u proizvodnji lijekova obilježilo je otkriće penicilina koji je bio prvi antibiotik korišten za liječenje bakterijskih infekcija. Razvoj tehnologije omogućio je unapređenje istraživanja na području znanosti što je dovelo do niza otkrića na području biologije, kemije, medicine i farmacije.

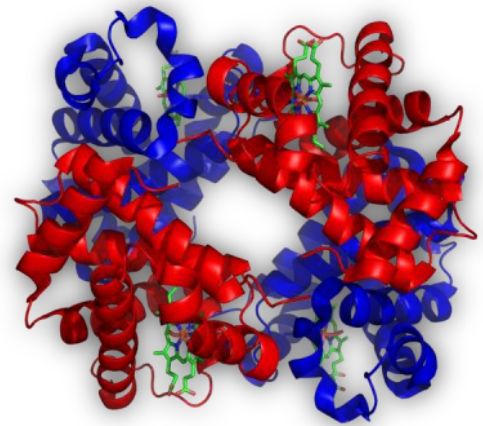
Otkriće DNK šezdesetih godina 20. stoljeća predstavljalo je početak razvoja molekularne biologije koja se bavi proučavanjem spojeva na mikrobiološkoj razini. U procesu molekularnobioloških istraživanja cirkuliraju velike količine podataka s kojima je potrebno manipulirati na optimalan način što je dovelo do potrebe suradnje dviju znanosti: računarske znanosti i molekularne biologije. Posljedica stapanja ovih dvaju znanstvenih područja rezultirala je novom interdisciplinarnom znanosti koja se danas naziva – bioinformatika.

Bioinformatika je polje znanosti koje koristi računala za pohranu, pristup, manipulaciju i distribuciju informacija koje opisuju biološke makromolekule kao što su: DNK, RNK i proteini. Bioinformatika obuhvaća tri bitna područja: razvoj novih algoritama koji se koriste za upravljanje velikim količinama podataka biološkog karaktera; analizu i interpretaciju različitih tipova podataka koji se odnose na nukleotide, sekvence aminokiselina i proteine; razvoj i implementaciju alata koji omogućuju efikasan pristup i upravljanje različitim tipovima podataka (npr. podaci o strukturi proteina)^[1]. Jedan od zadataka bioinformatike je da korištenjem *in silico*¹ metoda ubrza proces otkrivanja lijeka. Zahvaljujući računalima i računalnim modelima znanstvenici su u mogućnosti detaljnije prodrijeti u problem otkrivanja novih lijekova i cjepiva.

¹Termin *In silico* obuhvaća sav rad na računalu kojim se simulira neki biokemijski proces čije bi izvođenje u laboratoriju bilo dugotrajnije i kompleksnije.

2 Od aminokiseline do proteinskog kompleksa

Osnovne građevne jedinice proteina nazivaju se aminokiseline. Svaka od ukupno dvadesetak poznatih aminokiselina sastoji se od: centralnog atoma ugljika, amino skupine $-NH_2$, vodikovog atoma $-H$ i karboksilne skupine $-COOH$ na koju se nadovezuje kiselinski ostatak koji se označava sa slovom R. Aminokiseline se spajaju zajedno u velike lance polimera preko peptidnih veza između jedne aminokiseline i karboksilne skupine susjedne aminokiseline. Redoslijed vezivanja aminokiselina određen je slijedom gena zapisanih u genetskom kodu. Proces povezivanja aminokiselina naziva se sinteza proteina. Proteini su esencijalni spojevi u ljudskom organizmu; izgrađuju mišiće, krv, kožu, nokte, unutarnje organe, neke hormone... Za razumijevanje proteina i njihovih funkcija bitno je poznavati i strukturu proteina koja se definira načinom savijanja proteina (engl. *foldng*). Postoje četiri strukture proteina:



Slika 1. Kvartarna struktura hemoglobina

- Primarna struktura ili primarni oblik proteina je linearna sekvenca aminokiselina povezanih peptidnom vezom.
- Sekundarna struktura je lokalna prostorna organizacija aminokiselina povezanih vodikovim vezama u trodimenzionalnoj formi. Te strukture mogu biti α -helix (uzvojnica) ili β -sheet (ploča) ili β -turn (zavoji), ovisno o načinu na koji se aminokiseline saviju.
- Terarna struktura je trodimenzionalna struktura čitavog polipeptidnog lanca koja se dobije kombinacijom sekundarnih struktura
- Kvartarna struktura (proteinski kompleks) predstavlja prostorni raspored više polipeptida (svih terarnih struktura) koji tvore protein.

Slijed aminokiselina skoro u potpunosti određuje trodimenzionalnu strukturu proteina i njegovu funkciju. Ta je struktura stoga od iznimnog značaja za istraživanje lijekova. Poznata je 3D struktura 56% proteina i za predikciju 3D strukture se također koriste računalne metode:

- Homologija modeliranja (engl. *homology modelling*) je konstrukcija modela proteina na molekularnoj razini iz njegove sekvence aminokiselina i eksperimentalno dobivene strukture poznatih homolognih proteina (predložka).
- Ab initio metoda predikcije strukture proteina temelji se isključivo na sekvenci aminokiselina i ne koristi nikakve predložke. Koristi se uglavnom za predikciju strukture malih molekula i zahtjeva veliku količinu računalnih resursa.
- *Threading* ili prepoznavanje savijanja je metoda koja se koristi za modeliranje proteina koji se savijaju na jednak način kao neki od proteina kojima je poznata 3D struktura, ali nisu homologni.

Područje bioinformatike koje se bavi predikcijom 3D strukture bioloških makromolekula (proteini, RNK, DNK) naziva se strukturalna bioinformatika.

3 Uvod u istraživanje lijekova

Neovisno o metodama koje se koriste da bi se određeni lijek otkrio i testirao, postoji par zajedničkih obilježja koja karakteriziraju taj postupak; proces istraživanja je dugotrajan i osim mnogo vremena također zahtjeva velika novčana sredstva. Naravno, te zahtjeve je nemoguće zaobići, ali uz pomoć računala uvelike se može uštedjeti na vremenu podjednako kao i na budžetu korištenom u cjelokupnom procesu. Korištenjem *in silico* pristupa pri istraživanju lijekova moguće je ubrzati proces za dvije do tri godine (15 – 20 godina potrebno je za pronalazak lijeka korištenjem tradicionalnog pristupa) te uštedjeti oko 200 000€ (tradicionalan pristup zahtjeva bar dvostruko veću količinu novčanih sredstava)^[2].

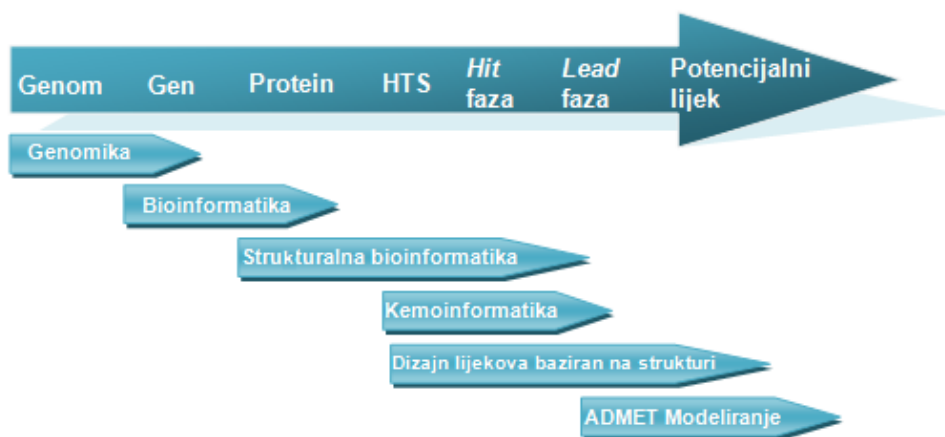
3.1. Tradicionalni pristup

Tradicionalni pristup obuhvaća proces istraživanja lijeka koji ne uključuje računalo za predviđanje potencijalnih spojeva i mogućih reakcija. Takav pristup započinje tradicionalnom medicinom odnosno sastojkom koji se izolira iz npr. ljekovite biljke^[4]. Produkt izolacije se naknadno nastoji sintetizirati u laboratorijskim uvjetima. Manipulacijom strukture spoja dobiva se novi spoj koji je u svojim ljekovitim svojstvima učinkovitiji (npr. nastoji se smanjiti količina popratnih simptoma) od prvotno izoliranog. Sam postupak otkrivanja proteina i modifikacije njegove strukture traje oko desetak godina i nakon toga slijede pretklinička i klinička testiranja te odobravanje od strane agencije za lijekove i medicinske proizvode.

3.2. Suvremeni pristup

Za razliku od tradicionalnog pristupa u kojem je polazišna točka potencijalni lijek, suvremeni pristup kreće od same bolesti. Nastoji se locirati biološka meta koju cilja određena bolest i izolira se glavni spoj zaslužan za razvoj bolesti. Nakon što je poznata meta bolesti uz pomoć računala se traži protein (ili neka druga molekula) koji bi po svojim karakteristikama mogao inhibirati razvoj bolesti te se testiraju razne kombinacije spojeva sve dok se ne nađe onaj par proteina koji će međusobno tvoriti kompleks. Struktura trodimenzionalnog kompleksa predstavlja potencijalni lijek ili cjepivo nakon čega slijede pretklinička testiranja. U suvremenom pristupu nužna je suradnja znanstvenika s više znanstvenih područja: genomika, bioinformatika, strukturalna bioinformatika, kemoinformatika (engl. *chemoinformatics*)^[4]... Najmanja jedinica od koje se počinje u dizajniranju lijeka je genom, a na posljeticu lijek mora zadovoljiti ADMET² kriterije što i grafički ilustrirano na slici 2.

² ADMET je skraćenica od sintagme *absorption, distribution, metabolism, excretion and toxicity*. Njome se definira skup elemenata na koje se posebno obraća pozornost pri dizajniranju lijekova: apsorpcija, distribucija, metabolizam, izlučivanje i toksičnost.



Slika 2. Suvremeni pristup istraživanju lijekova

4 Prvi koraci u dizajniranju lijekova

Osnovni predmet proučavanja bioinformatike u području istraživanja lijekova su proteini. Na proteine se vežu ligandi koji također mogu biti proteini; ali i neurotransmiteri, lipidi, nukleinske kiseline, polisaharidi^[4]. Proteini na koje se vežu ligandi nazivaju se receptori. Postoje dvije glavne vrste receptora: proteini koji slobodno plove u staničnoj citoplazmi i proteini koji su integrirani u staničnu membranu. Za posljednje navedenu skupinu proteina bitno je da lijek ne mora prodrijeti u stanicu nego je dovoljno inhibirati protein na površini čime se sprječava daljnja mogućnost uobičajenog funkcioniranja stanice.

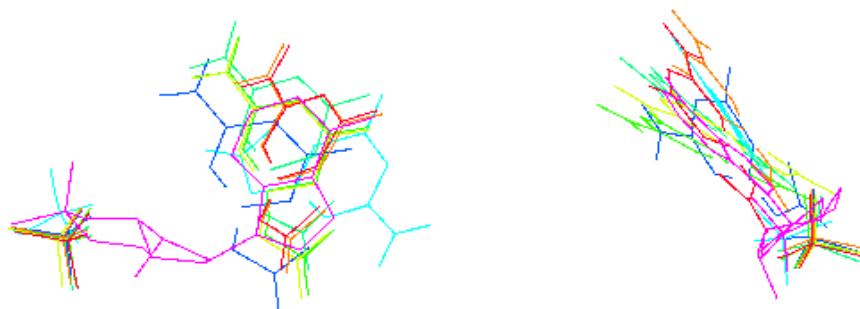
Ako ligand i receptor tvore kompleks tada se može govoriti o potencijalnom lijeku. Ovisno o nusproduktima koji nastaju njihovim vezanjem i novonastalim biokemijskim procesima odlučuje se da li taj protein može biti korišten kao lijek ili cjepivo. Postoje dva osnovna neovisna mehanizma na molekularnoj razini: što je veći afinitet liganda prema receptoru to će on bolje prijanjati (bolje prijanjanje predstavlja potrebu za manjim količinama) i ligand mora ili stimulirati ili inhibirati određene funkcije receptora kako bi mogao regulirati njegovu aktivnost. Oba mehanizma ovise o načinu na koji će se ligand pozicionirati na receptor. Najčešće je potrebno isprobati desetke tisuća spojeva kako bi se došlo do potencijalnih lijekova, dok ih tek par dođe do pretkliničkih testiranja^[5]. Preko 90% spojeva koji dođu do kliničkih testiranja nikad ne uspiju zadovoljiti daljnje uvjete i doći na police ljekarni^[4].

Istraživanje lijekova na racionalnim osnovama (engl. *Rational drug design*) obuhvaća dvostruku ulogu računala pri istraživanju lijekova. Prvo, računalima je moguće optimizirati farmakološku sliku postojećih lijekova sintezom novih spojeva. Drugo, računalima možemo „predvidjeti“ što bi se moglo desiti: s povećanjem količine podataka o strukturi proteina mogu se razviti novi terapijski koncepti u kojima će računalo na temelju poznatih podataka o proteinu

izračunati koja bi bila funkcija i struktura „nepoznatog“ proteina. Poznati podaci o proteinu obuhvaćaju sekvence aminokiselina iz baza podataka o istraženim proteinima i razumijevanje funkcije proteina na temelju njegove strukture. Ako je poznat biološki ili biokemijski mehanizam bolesti tada se može i pretpostaviti koje molekule taj mehanizam zahtjeva za promjenu funkcioniranja. Glavna pretpostavka jest da će prijanjanje liganda biti uspješnije što su strukturalno i kemijski komplementarniji spoj ligand-receptor. Ovisno o poznatim podacima o proteinu (bilo ligandu ili receptoru) primjenjuju se različite metode dizajniranja.

4.1. Otkrivanje strukture receptora

Poznavanje trodimenzionalne strukture receptora osnova je od koje se kreće u dizajniranju lijekova. Cilj je doći do strukture onog dijela proteina za koji se veže ligand. Proučavanjima temeljenim na statističkim metodama nastoji se saznati koje mjesto na proteinu je najizloženije utjecajima liganada. Ako je poznata sekvenca aminokiselina receptora, računalom se može pokušati predvidjeti izgled strukture nepoznatog dijela proteina. Pri tom postupku koristi se poznata struktura proteinskog predloška iz baze podataka o proteinima. Ako je 25-30% kiselinskih ostataka u dva proteina jednako, može se pretpostaviti da ti proteini imaju vrlo slične strukture^[5]. Ova vrsta predviđanja strukture naziva se homologija modeliranja. Metode koje se još koriste su rendgenska kristalografija i spektroskopija NMR (Nuklearna magnetska rezonancija). Rendgenska kristalografija je metoda određivanja trodimenzionalnog rasporeda atoma u kristalima rendgenskim zrakama i analizom njihovih difrakcijskih uzoraka^[8]. Spektroskopija NMR je metoda odabira za izučavanje interakcija biomolekula s potencijalnim ligandima. Ona daje podatke o 3D strukturi slobodnih proteina kao sirove podatke za molekularno modeliranje vezanja liganada i izravno može dati važne eksperimentalne informacije o samim biološki važnim kompleksima. Za razliku od rendgenske strukturne analize spektroskopijom NMR može se proučavati dinamika kako samih velikih biomolekula tako i dinamika kompleksa biomolekula i potencijalnih liganada^[9].



Slika 3. Dva prikaza predstavljaju moguće orijentacije i konformacije liganda na receptor: bočni prikaz i pogled odozdo (respektivno).

4.2. Otkrivanje strukture liganada

Ako je poznata 3D struktura biološke mete odnosno receptora može se krenuti s istraživanjem strukture liganda. Istraživanje strukture liganada je sinonim za istraživanje lijekova. Postoje dvije metode istraživanja liganada.

Metoda dizajniranja lijekova u kojoj se počinje od poznatih molekula (liganada) koje se vežu za neki receptor je *metoda temeljena na ligandima*. Ideja leži u činjenici da sve molekule koje se vežu za određeni receptor imaju skup sličnih svojstava i na temelju tih svojstava nastoji se dizajnirati nova molekula. Iz poznatih molekula proizlazi farmakofor (engl. pharmacophore) koji definira minimum potrebnih strukturalnih svojstava koje molekula mora zadovoljavati da bi se vezala uz neki receptor^[6].

Ukoliko je poznata 3D struktura receptora tada se koristi drugi pristup koji se temelji na strukturi (engl. *structure-based*). U ovoj metodi dizajniramo ligand koji reagira s receptorom. Prvi korak sastoji se od otkrivanja aktivne strane proteina. Aktivna strana proteina je dio proteina koji se spaja s ligandom. Da bi se otkrila aktivna strana potrebno je poznavati 3D strukture oba proteina u PDB³ formatu, hidrofobne atome, donora vodikove veze, akceptora vodikove veze i polarni atom^[6].

³ PDB – Protein Data Bank je baza podataka koja sadrži informacije o velikim molekulama kao što su proteini, DNK i RNK. Redovito se upotpunjava novim podacima i koristi se diljem svijeta u raznim istraživanjima. PDB format je tekstualna datoteka koja sadrži podatke o molekuli kao što su: koordinate atoma, sekvence aminokiselina, podatke o istraživačima koji su definirali strukturu itd.

5 Virtual screening

Prvi način računalno potpomognutog dizajniranja lijekova (engl. *computer aided drug design*) obuhvaća pojam pod nazivom *virtual screening*. Uloga bioinformatike u ovom dijelu istraživanja je višestruka; razvoj programa koji obrađuju podatke i korištenje navedenih programa u istraživačke svrhe. *Virtual screening* koristi se za prepoznavanje pogodaka (engl. *hitova*) jednako na temelju poznatih struktura liganda kao i na temelju poznatih struktura receptora.

5.1. Dokiranje proteina

Dokiranje proteina je metoda molekularnog modeliranja koja predviđa orijentaciju molekule u prostoru nakon povezivanja s drugom molekulom. Poznavanje orijentacije molekula pruža informacije o tome koliko čvrsta veza će se uspostaviti između te dvije molekule. Ovaj model naziva se i modelom brava-ključ u kojem ključ predstavlja ligand koji se traži, a brava receptor. Dokiranje se provodi nad receptorima koji su proteini, dok su ligandi također proteini ali se bitno razlikuju u svojim dimenzijama. U dizajniranju lijekova posebna se pozornost obraća na manje proteine. Dva su osnovna pristupa koja se bave rješavanjem ovog problema, no oba ih povezuje temelj na kojem su zasnovani, a to je termodinamika intermolekularnih reakcija.

Prvi pristup se zasniva na komplementarnosti površina sastavnih dijelova kompleksa. Raspoznavanje strukture kompleksa protein-protein predstavlja problem koji rješavaju za tu svrhu posebno dizajnirani algoritmi. Generalna strategija za simulaciju dokiranja uključuje uspoređivanje komplementarnosti oblika. Neki pristupi se posebno oslanjaju na usklađivanje vanjskih površina proteina, dok postoje i oni pristupi koji poboljšavaju potragu za geometrijskom komplementarnošću usklađujući položaj površinskih sfera i površinskih normala^[11].

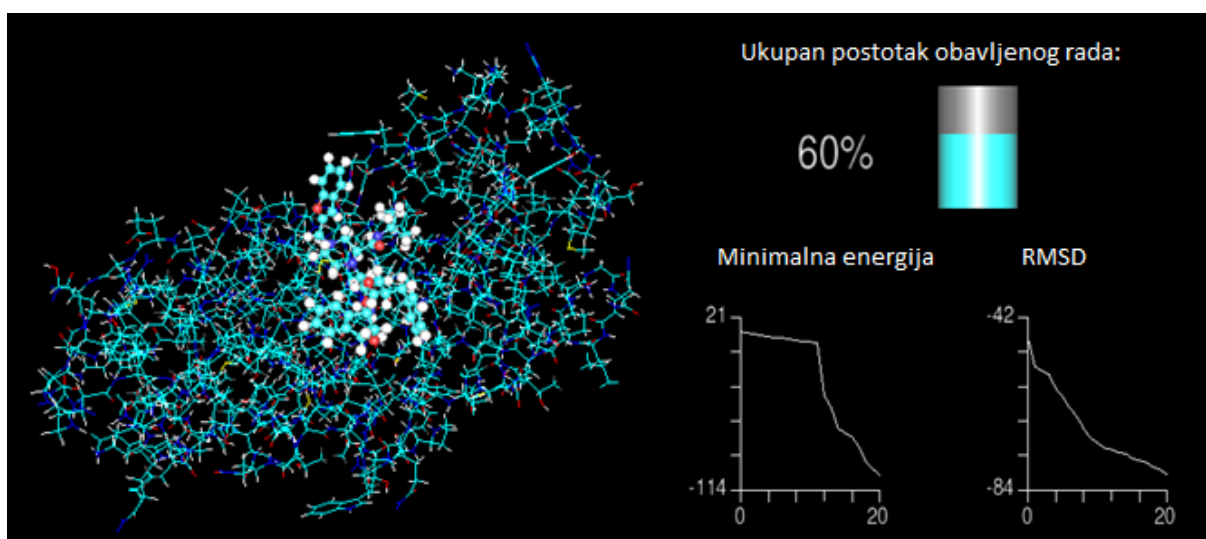
Drugi pristup zasniva se na traženju minimalne energije i što manjoj vrijednosti RMSD (engl. *Root mean square deviation*). RMSD je mjera prosječne udaljenosti koja se mjeri uz pomoć koordinata C α atoma koji su karakteristika tercijarne strukture proteina. RMSD se definira preko jednadžbe:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta^2}$$

Pri čemu δ predstavlja udaljenost između N parova ekvivalentnih atoma (uobičajeno C α ili ponekad C, N, O, C β). Vrijednost RMSD izražena je preko mjerne jedinice Ångströma[Å] i iznosi 10⁻¹⁰ m. U procesu dokiranja RMSD se koristi da bi se usporedile strukture proteina koji nije do

kraja savijen i struktura izvornog proteina. Kod simulacije u ovom pristupu ligand se postavi na određenu fizičku udaljenost od proteina, a potom ga se pusti da traži svoju poziciju na proteinu nizom konformacija preko translacija i rotacija, kao i internim promjenama strukture liganda koje uključuju promjenu kuta torzije. Svaka konformacija liganda rezultira apsorpcijom energije sistema pa se za svaku konformaciju ta energija nanovo računa.

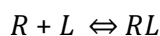
Na slici 4. prikazana je simulacija dokiranja programom za dokiranje. Grafički je prikazan proces dokiranja pri čemu je receptor prikazan štapičastim modelom, a ligand modelom kuglica-štap. Pored tog prikaza nalaze se statistički podaci o utrošku energije i promijeni vrijednosti RMSD tijekom vremena. Prikazani kompleks predstavlja dokiranje enzima HIV proteaze.



Slika 4. Dokiranje receptora enzima HIV proteaze

Afinitet vezanja

Vezanje liganda na receptor opisuje se jednadžbom:



receptor + ligand ⇌ receptor - ligand kompleks

Visoki afinitet vezanja liganda je fizikalno važan zbog dijela energije koja se oslobađa pri vezanju i uzrokuje konformacijske promjene receptora. Ligand koji se veže na receptor promijeni njegovu funkciju i uzrokuje fiziološku promjenu se još naziva i agonist. Antagonisti su ligandi koji se vežu na receptor, ali pritom ne uzrokuju fiziološku promjenu. Ligandi s visokim afinitetom vezanja nazivaju se i potpuni agonisti. Za veliku uspješnost vezanja dovoljna je mala koncentracija takvih liganada. Ako se povećava količina liganada na konstantan broj receptora moguće je uzrokovati zasićenje vezanja. Zasićeno i visokoafinitetno vezanje se nazivaju i specifično vezanje koje je bitno pri promatranju djelovanja lijekova. Koncentracija liganada koja

usprije zauzeti polovicu od ukupnog broja aktivnih strana na receptoru naziva se konstanta disocijacije – K_d . Vezanje receptor-ligand ovisi o koncentracijama interaktivnih komponenti i opisano je konstantom ravnoteže:

$$K_a = \frac{[RL]}{[R][L]} = \frac{1}{K_d}$$

K_a je konstanta asocijacije. Koristeći Scatchardovu analizu može se ustanoviti konstanta disocijacije K_d i broj aktivnih strana na receptoru u zadanom pripravku. Kada proces vezanja postigne ravnotežu ukupan broj mogućih aktivnih strana B_{max} jednak je zbroju slobodnih strana $[R]$ i vezanih strana na ligandu, tj. $B_{max} = [R] + [RL]$. Iz te jednadžbe proizlazi broj slobodnih strana: $[R] = B_{max} - [RL]$. Izraz za ravnotežu se može zapisati kao:

$$K_a = \frac{[RL]}{[L](B_{max} - [RL])}$$

Omjer liganada vezanih na receptor i slobodnih izražen je preko^[13]:

$$\frac{[vezani]}{[slobodni]} = \frac{[RL]}{[L]} = K_a(B_{max} - [RL]) = \frac{1}{K_d}(B_{max} - [RL])$$

Scoring funkcije. Algoritmi pretraživanja

Koliko će dokiranje biti uspješno ovisi o scoring funkcijama i algoritmu pretraživanja (engl. *search algorithm*). *Scoring* funkcije predstavljaju skup matematičkih metoda koje služe za eliminaciju netočnih struktura kompleksa dobivenih eksperimentalnim metodama. Njima se može ustanoviti afinitet vezanja između liganda i receptora, a temelje se na električnim karakteristikama proteina, empirijski definiranim poljima sila, hidrofobnosti površina te raznim drugim kemijskim i mehaničkim karakteristikama interakcija.

Algoritam pretraživanja pronalazi strukture za koje *scoring* funkcija daje visoku vjerojatnost vezanja. Zbog toga što su obje molekule fleksibilne i imaju velik stupanj slobode (engl. *degrees of freedom* - DOF) broj mogućih položaja molekula je astronomskih veličina. Za razliku od konformacije koja predstavlja prikaz molekule preko relativnih odnosa atoma u trodimenzionalnoj strukturi neovisno o koordinatnom sustavu, položaj molekule strogo je definiran koordinatama u koordinatnom sustavu^[14]. Stariji programi za dokiranje u obzir su uzimali samo translacijske i rotacijske stupnjeve slobode liganda u odnosu na receptor, dok se u novije doba zahvaljujući sofisticiranijim algoritmima koriste i podaci o stupnju slobode unutar

liganada. Očigledno, što se povećava broj stupnjeva slobode povećava se i prostor pretraživanja što zahtjeva brze i efikasne algoritme kako bi se dokiranje uspješno izvelo.

Algoritmi za dokiranje odnosno pretraživanje dijele se u više skupina: stohastičko pretraživanje, *incremental construction* algoritme i višekonformacijsko poravnavanje (engl. *multiconformer alignment*). Primjer jednostavnog evolucijskog algoritma prikazan je sljedećim pseudokodom^[15]:

1. stvori inicijalnu populaciju rješenja
2. koristeći funkciju energije izračunaj prikladnost svakog rješenja unutar populacije
3. koristeći operator mutacije stvori potomka iz svakog roditelja
4. izračunaj prikladnost potomka
5. za sve potomke u turniru
 - a. slučajnim odabirom izaberi N ostalih potomaka kao protivnike u turniru
 - b. zabilježi pogodak svaki puta kada odabrani potomak više odgovara nego protivnik
 - c. rangiraj odabranog potomka prema broju pogodaka u njegovom turniru
6. odaberi najbolje rangirana rješenja kao novu populaciju rješenja
7. ako je broj generacija premašio korisničke zahtjeve stani
inače skoči na 3.

6 QSAR

U nastojanju znanstvenika da za sintetizirani spoj saznaju afinitet vezanja i ostala relevantna svojstva spoja kao kandidata za lijek (ADMET), koriste se hipotezom da se slične molekule ponašaju na sličan način. Ovaj princip poznat je kao skraćenica SAR (engl. *structure-activity relationship*). SAR obuhvaća skup postupaka u medicinskoj kemiji koji modificiraju kemijsku strukturu spoja. S ciljem smanjenja tako sintetiziranih spojeva za koje se naknadno ispostavi da su neupotrebljivi pri daljnjem istraživanju, koristi se skup matematičkih i statističkih alata objedinjenih u procesu pod nazivom QSAR (engl. *quantitative structure-activity relationship*). QSAR se koristi da bi se za određeni spoj ustanovilo: hoće li on biti aktivan ili ne i kolika će ta aktivnost biti, da li donosi neke nove informacije korisne za daljnje istraživanje, zadovoljava li ADMET svojstva i koristi li boljem razumijevanju SAR-a^[21].

6.1. Matematička forma. Prve primjene

QSAR se temelji na količinskom (kvantitativnom) odnosu između biološke aktivnosti molekularnog sustava i njegovih geometrijskih te kemijskih svojstava. Nastoji se pronaći konzistentan odnos između biološke aktivnosti i molekularnih svojstava kako bi se uočene pravilnosti mogle iskoristiti za evaluaciju biološke aktivnosti novih spojeva. Biološka aktivnost izražava se sljedećom matematičkom formulacijom:

$$\text{Biološka aktivnost} = f(\text{deskriptor molekule ili dijela molekule})$$

Deskriptor (engl. *descriptor*) predstavlja neko svojstvo molekule kojim ju pobliže opisuje. Deskriptori mogu biti: fizičko-kemijski, topologijski, geometrijski, elektronski. Statistički koncept QSAR-a prikazan je tablicom u nastavku u kojoj se u prvom redu nalazi: biološka aktivnost iskazana konkretnom vrijednošću i n deskriptora označenih P_1, \dots, P_n .

Tablica 1. Primjer prikaza statističkih podataka spojeva korištenih za QSAR

	Biološka aktivnost	P_1	P_2	...	P_n
Spoj 1	0.34	3.7	1.1		
Spoj 2	3.2	9			
...					
Spoj n					

Zadatak QSAR-a je da pronade vrijednosti koeficijenata C_0, C_1, \dots, C_n takve da zadovoljavaju jednadžbu^[16]:

$$\text{Biološka aktivnost} = C_0 + C_1P_1 + \dots + C_nP_n$$

Korelacijski model (engl. *partial least squares*, skraćeno PLS) je statistička metoda korištena za računanje koeficijenata strukturalnih deskriptora.

Prva primjena QSAR-a pripisuje se C. Hanschu i T. Fujiti koji su šezdesetih godina 20. stoljeća prvi formulirali jednadžbu kojom su iskazali odnos između biološke aktivnosti i električkih svojstava te hidrofobnosti skupa struktura:

$$\log \frac{1}{C} = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

Pri čemu C predstavlja minimalnu efikasnu količinu lijeka (molarna koncentracija), P je koeficijent particije vode, σ Hamettova jedinična konstanta, a k_i konstante dobivene analizom. $\log P$ predstavlja relativnu topivost lijeka u oktanolu (zamjena za lipidnu dvoslojnu staničnu membranu) i vodi (koja se nalazi u stanici i krvi). Vrijednost $\log P$ ili koeficijent particije je posebno važna pri dizajniranju lijekova jer definira koliko je određeni lijek „ljekovit“ prema Lipinskijevom pet pravilu⁴. Određene vrijednosti se mogu dobiti eksperimentalno, ali i izračunati algoritmom koji koristi poznate izračunate vrijednosti. Izračunate vrijednosti koje se koriste pri QSAR-u su: površina plohe, volumen molekule, zakrivljenost, HOMO i LUMO energija (energije zauzetih i slobodnih orbitala), naponi na atomu, maksimalan pozitivan i negativan naboj... Neke od mjerljivih vrijednosti fizikalnih svojstva koja se koriste pri QSAR-u su: gustoća, energija ionizacije, molarna masa, točka vrenja, indeks refrakcije, dipolni moment (μ). Iako se istraživanjima ustanovi korelacija između više deskriptora, u statističkim se analizama nastoji izdvojiti samo one varijable koje najbolje opisuju promatranu biološku aktivnost tako da završni QSAR uključuje najviše 3 do 5 najvažnijih deskriptora. Skup podataka o potencijalnom lijeku mora biti puno veći od skupa podataka o spojevima u deskriptoru kako bi se eliminirale lažne visoke korelacije među spojevima^[16].

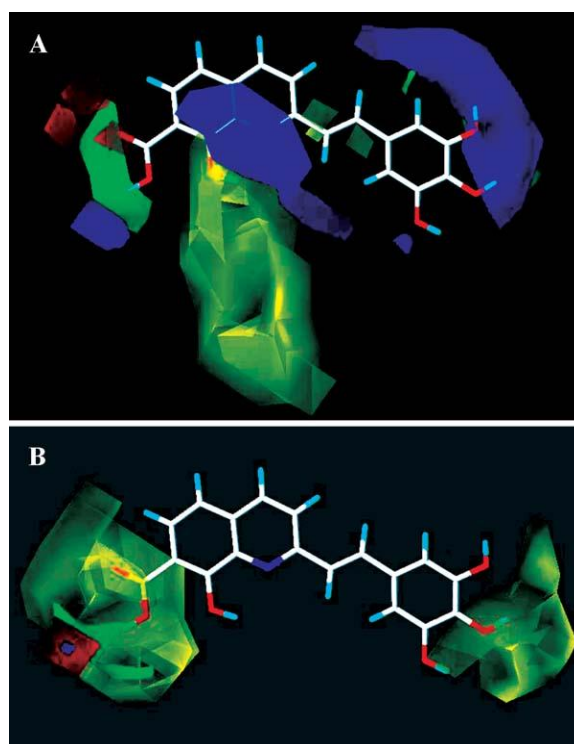
⁴ Lipinskijevo pet pravilo je heuristička metoda kojom se određuje koliko određeni spoj može biti povoljan kao lijek za oralnu primjenu kod ljudi. Pravilo se sastoji od pet točaka i glasi: lijek ne smije imati više od 5 donora vodikove veze, ne smije imati više od 10 akceptora vodikove veze, mora imati molekularnu masu manju od 500 daltona i $\log P$ manji od 5

6.2. 3D-QSAR

Strukturalni deskriptori koji se odnose na trodimenzionalne strukture liganada i receptora se izučavaju u okviru posebnih metoda QSAR-a koje se nazivaju 3D-QSAR. Najčešći strukturalni deskriptori su farmakofori i površinska svojstva molekule. Važna područja biološki aktivnih molekula (npr. područja hidrofobnosti, hidrofilnosti, donora vodikovih veza, akceptora vodikovih veza) na poseban se način mapiraju u 3D obliku. Tako se dobiva trodimenzionalni uzorak funkcionalno značajnih područja za lijekove. Za razliku od običnog QSAR-a, 3D-QSAR preferira izračunate vrijednosti pored onih dobivenih eksperimentalno. Najzastupljenija metoda korištena u 3D-QSAR-u je komparativna molekularna analiza polja (CoMFA).

CoMFA

Komparativna molekularna analiza polja (CoMFA) je 3D-QSAR metoda koja pri analizi kvantitativnih odnosa između biološke aktivnosti i deskriptora kao deskriptore koristi sterička (interakcije prouzročene van der Waalsovima silama) i elektrostatička svojstva molekule^[17]. Ti podaci o molekuli dobiveni su pretraživanjem trodimenzionalne rešetke molekule i nizom superponiranih konformacija svakog spoja. Uz pomoć CoMFA-e računaju se koeficijenti QSAR jednadžbe. Na slici 6. prikazana je molekula liganda korištenjem CoMFA metode: pod A) označena su sterička i elektrostatička polja na molekuli, pod B) su označeni dijelovi molekule koji mogu sudjelovati u vodikovim vezama.



Slika 5. CoMFA model liganda

Princip funkcioniranja CoMFA metode^[18] opisan je u četiri koraka:

- Aktivne molekule se smjeste u trodimenzionalnu mrežu tako da obuhvaćaju sve molekule
- Za svaki vokal⁵ računa se sterička (Lennard-Jonesov potencijal) i elektrostatska energija uz pomoć ispitivačkog atoma (engl. *probe atom*) koji je najčešće sp^3 ugljikov atom sa nabojem +1.0
- Kako bi se minimizirala dominacija područja velikih steričkih i elektrostatskih polja, sve energije koje premašuju određenu vrijednost se postavljaju na vrijednosti koje se u daljnjem postupku ne uzimaju u obzir
- CoMFA koristi PLS analizu za računanje koeficijenata i predviđanje biološke aktivnosti iz vrijednosti energija pojedinog voksla

Problem kod CoMFA metode predstavlja superpozicija molekula. Da bi se metoda uspješno provela svi spojevi moraju biti optimalno poravnati^[16]. Nadalje, zbog velike fleksibilnosti molekula moguće su stalne promjene koordinata voksla u mreži pa se velika količina podataka mora nanovo računati.

⁵ Kao što u dvodimenzionalnom prikazu piksel predstavlja najmanju jedinicu površine, tako u trodimenzionalnom prikazu vokal predstavlja najmanju volumnu jedinicu.

7 Zaključak

Istraživanje lijekova je područje znanosti koje je čovjeku oduvijek predstavljalo izazov. Nepremostivi problemi s bolestima prenosili su se s generacije na generaciju dok taj lijek ne bi bio otkriven i time produžio život ponekad i milijunima ljudi. Suvremeno doba sa sobom donosi i nove izazove koje je nemoguće riješiti bez suradnje znanstvenika s raznih područja. Tako je nastala i bioinformatika, simbiozom molekularne biologije i računarske znanosti.

Bioinformatika je omogućila veću efikasnost u obradi, organizaciji i korištenju podataka koji se odnose na kemijske spojeve i njihova svojstva. U kraćem roku može se testirati veća količina potencijalnih lijekova, smanjiti broj nuspojava i omogućiti oboljelima da što prije prime terapiju novootkrivenim lijekom. *In silico* metode zamijenile su veći dio *in vivo* i *in vitro* istraživanja. Pri dizajniranju lijekova bioinformatika svoj doprinos ostvaruje prvenstveno *virtual screeningom* i QSAR-om. Koristeći podatke o strukturi proteina dobivenim preko rendgenske kristalografije i spektroskopije NMR nastoji se pronaći molekula koji najbolje „sjeda“ u receptor. Podaci o molekulama koje predstavljaju kandidate za lijek uzimaju se iz PDB baze podataka koja je također predmet istraživanja bioinformatike.

Dokiranjem se nastoji pronaći optimalna konformacija liganda u odnosu na receptor, pritom uzimajući u obzir i unutarnje promjene položaja atoma. U tu svrhu se koriste heuristički algoritmi za procjenu koliko koja molekula odgovara receptoru na temelju poznatih podataka. QSAR metoda zasniva se na steričkim i elektrostatskim svojstvima molekule iz kojih nastoji predvidjeti aktivne strane koje sudjeluju u reakciji s ligandom tvoreći vezu ligand-receptor.

Iako je uz pomoć računala proces dizajniranja lijekova uvelike unaprijeđen, kao rješenje za niz problema koji se pritom javljaju jedino ostaje čovjekov um, znanje te etička i moralna načela. Uz pomoć suvremene tehnologije lakše i brže možemo doći do cilja, a nad čovjekom počiva dužnost da cilj koji odabere bude ispravan.

8 Literatura

- [1] Chitty, M.: *Overviews: Bioinformatics, Cheminformatics & beyond*.
<http://www.genomicglossaries.com/content/chapterinfosourcetest.asp> (18. 4 2010)
- [2] Villoutreix, B: *Drug Design, Bioinformatics*
http://www.vls3d.com/courses_talk/Villoutreix_intro_drug_design.pdf (18. 4 2010)
- [3] Blundell, T. (1990). *Designer drugs head for the marketplace*, NewScientist, 1720:
<http://www.newscientist.com/article/mg12617204.700-designer-drugs-head-for-the-marketplace-pharmaceuticals-companies-spend-millions-searching-for-new-drugs-some-chemists-prefer-to-design-them-their-approach-could-help-us-to-fight-aids-diabetes-and-high-bl> (21. 04 2010)
- [4] Melissa, P.: *San Diego Supercomputer Center*,
www.sdsc.edu/pb/edu/pharm202/03/passino.ppt (21. 04 2010)
- [5] Altenhofen, W. (n.d.). *VSNS BioComputing Division*, <http://www.techfak.uni-bielefeld.de/bcd/ForAll/Introd/drugdesign.html> (21. 04 2010)
- [6] *Drug Design*, Wikipedia: http://en.wikipedia.org/wiki/Drug_design (19. 04 2010)
- [7] Božić, F.: *Mehanizmi djelovanja lijekova (opća načela)*, Farmakologija & Toksikologija:
http://www.jelena-suranc.com/joomla/images/stories/predavanja/frane/1._Mehanizmi_djelovanja_lijekova_-_opcenito.pdf (21. 04 2010)
- [8] *Kemijski rječnik i glosar*, Kemijsko-tehnološki Fakultet: http://www.ktf-split.hr/glossary/hr_o.php?def=rendgenska%20kristalografija (21. 04 2010)
- [9] Tepeš, P., & Novak, P. (2008.). *Interakcije receptora i liganada pod lupom spektroskopije NMR*, Hrčak - portal znanstvenih časopisa RH, Kemija u industriji, 57, 165-073:
<http://hrcak.srce.hr/file/37330> (21. 04 2010)
- [10] Funkhouser, T: *Structure-Based Drug design*, Princeton University - Department of Computer Science:
<http://www.cs.princeton.edu/courses/archive/fall05/cos597A/lectures/design.pdf> (21. 04 2010)
- [11] Henry A. Gabb, R. M.: *Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information*, Computer Graphics at Stanford University:

http://graphics.stanford.edu/courses/cs468-01-fall/Papers/gabb_jackson_sternberg.pdf (21. 04 2010)

[12] *Docking (molecula)*, Wikipedia: http://en.wikipedia.org/wiki/Docking_%28molecular%29 (21. 04 2010)

[13] Cox, D. L.: *Principles of Biochemistry*, 4. Izdanje, W. H. Freeman, 2004

[14] Choi, V.: *Yucca: An Efficient Algorithm for Small Molecule Docking*, Department of Computer Science, Virginia Tech: <http://people.cs.vt.edu/~vchoi/> (26. 04. 2010)

[15] David R. Westhead, D. E. (1997). A comparison of heuristic search algorithms for molecular docking. *Journal of Computer-Aided Molecular Design*, 11: 209-228, <http://array.bioengr.uic.edu/~yangdai/teach/bioe494-fall02/heuisticDocking.pdf>

[16] Dina Schneidman, A. S.-P. QSAR, BioInfo3D: http://www.google.hr/url?sa=t&source=web&ct=res&cd=5&ved=0CCIQFjAE&url=http%3A%2F%2Fbioinfo3d.cs.tau.ac.il%2Feducation%2FCS01a%2FGroupSeminar%2FQSAR.ppt&rct=j&q=qsar+typefile%3Appt&ei=2aLhS-vZKKO8mwP_hJ01&usg=AFQjCNFwsFFa7la_t80FTTQUbSkGwJFEPA

[17] *IUPAC Compendium of Chemical Terminology*, International Union of Pure and Applied Chemistry: <http://www.iupac.org/goldbook/CT06951.pdf> (26. 04. 2010)

[18] Ott, M.: *Comparative Molecular Field Analysis (CoMFA)*, CoMFA tutorial: <http://www.cmbi.kun.nl/edu/bioinf4/comfa-Prac/comfa.shtml> (30. 04. 2010)

[19] Xiao-hui MA: *Exploring binding mode for styrylquinoline HIV-1 integrase inhibitors using comparative molecular field analysis and docking studies*: <http://www.chinaphar.com/1671-4083/25/950.htm> (30. 04. 2010)

[20] Bevan, D. R.: *QSAR and Drug Design*, <http://www.netsci.org/Science/Compchem/feature12.html> (30. 04. 2010)

[21] Wermuth, C. G.: *The practice of Medicinal Chemistry*, 3. izdanje, Elsevier Ltd., 2008

9 Sažetak

U seminaru su obuhvaćene osnovne metode u bioinformatici koje se koriste pri dizajniranju lijekova. U svrhu razumijevanja tih metoda objašnjena su osnovna svojstva i funkcije proteina. Prikazana je usporedba tradicionalnog i suvremenog pristupa dizajniranju lijekova. Objašnjena je motivacija za korištenjem računala pri proučavanju kemijskih spojeva. Apstraktno je obuhvaćen i princip djelovanja lijeka u sklopu kompleksa receptor-ligand te metode kojima se otkrivaju njihove strukture. Objašnjeno je načelo dokiranja proteina i korištenih matematičkih te informatičkih metoda. U općenitom modelu prikazan je QSAR i 3D-QSAR te CoMFA kao najzastupljenija metoda unutar 3D-QSAR skupine metoda za dizajniranje lijekova.