

UNIVERZA V LJUBLJANI
FAKULTETA ZA FARMACIJO

GINA JARC

DIPLOMSKA NALOGA

UNIVERZITETNI ŠTUDIJ FARMACIJE

Ljubljana, 2011

UNIVERZA V LJUBLJANI
FAKULTETA ZA FARMACIJO

GINA JARC

**EVALUACIJA REZULTATOV RAČUNALNIŠKE APLIKACIJE ZA
NAPOVEDOVANJE MEST METABOLIZMA V MOLEKULAH**

**EVALUATION OF RESULTS OF COMPUTER APPLICATION FOR
THE SITE OF METABOLISM PREDICTIONS IN MOLECULES**

DIPLOMSKA NALOGA
UNIVERZITETNI ŠTUDIJ FARMACIJE

Ljubljana, 2011

Diplomsko nalogo sem opravljala na Fakulteti za farmacijo pod mentorstvom doc. dr. Lucije Peterlin Mašič, mag. farm.

Zahvala

Posebna zahvala je namenjena moji mentorici doc. dr. Luciji Peterlin Mašič, mag. farm. za pomoč in strokovno vodenje pri opravljanju diplomske naloge. Zahvalila bi se tudi svoji družini, ki mi je omogočila študij in mi tekom študija vedno stala ob strani.

Izjava

Izjavljam, da sem diplomsko nalogo samostojno izdelala pod mentorstvom doc. dr. Lucije Peterlin Mašič, mag. farm.

Gina Jarc

VSEBINA

POVZETEK	III
SEZNAM OKRAJŠAV	IV
1 UVOD	1
1.1 METABOLIZEM KSENOBIOTIKOV	1
1.1.1 Reakcije I. faze	1
1.1.1.1 Reakcije oksidacije in redukcije	4
1.1.1.2 Reakcije hidrolize	7
1.1.2 Reakcije II. faze	7
1.2 POMEN NAPOVEDOVANJA MEST METABOLIZMA V MOLEKULAH IN METABOLIČNE STABILNOSTI V ZGODNJIH FAZAH RAZVOJA UČINKOVIN ..	9
1.3 IN SILICO METODE ZA NAPOVEDOVANJE MEST METABOLIZMA V MOLEKULAH	11
1.3.1 MetaSite	11
1.3.2 QSAR	12
1.3.3 Doking	13
1.4 PROGRAM METAPRINT2D KOT ORODJE ZA NAPOVEDOVANJE MEST METABOLIZMA V MOLEKULAH	14
2 NAMEN DELA	19
3 MATERIALI IN METODE	21
3.1 IZBOR SPOJIN IN DOLOČANJE MEST METABOLIZMA	21
3.2 RAZLIKE IN PODOBNOSTI MED PROGRAMOMA MP2D (PRIMERJAVA PRVE VERZIJE PROGRAMA Z DRUGO)	22
3.3 NAPOVEDOVANJE METABOLIZMA	23
3.4 ANALIZA REZULTATOV TESTIRANJA	28
4 REZULTATI	32
4.1 REZULTATI GLEDE NA MESTO METABOLIZMA V MOLEKULAH.....	32
4.1.1 Prva verzija programa.....	32
4.1.2 Druga verzija programa.....	33
4.1.3 Primerjava druge verzije programa s prvo	34
4.2 REZULTATI GLEDE NA BARVO ATOMA.....	35
4.2.1 Prva verzija programa.....	35
4.2.2 Druga verzija programa.....	36
4.2.3 Primerjava druge verzije programa s prvo	37
4.3 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA TIP REAKCIJE	38
4.3.1 Glede na mesto napovedi metabolizma	38
4.3.2 Glede na barvo atoma	40
4.4 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA MOLEKULSKO MASO.....	42
4.4.1 Glede na mesto napovedi metabolizma.....	42
4.4.2 Glede na barvo atoma	44
5 RAZPRAVA	46
5.1 REZULTATI GLEDE NA MESTO METABOLIZMA V MOLEKULAH.....	46
5.2 REZULTATI GLEDE NA BARVO ATOMA.....	48
5.3 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA TIP REAKCIJE	51
5.4 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA MOLEKULSKO MASO MOLEKUL	53

6 SKLEP	55
7 LITERATURA	57

POVZETEK

Napovedovanje mest metabolizma v molekulah lahko močno vpliva na celoten potek razvoja novih učinkovin. V zadnjem času je velika pozornost usmerjena v razvoj in silico metod za napovedovanje mest metabolizma v molekulah, s katerimi že v zgodnjih fazah odkrivanja novih učinkovin izločimo potencialno reaktivne in metabolično nestabilne molekule. Pozno odkritje metabolične nestabilnosti povzroča visoke stroške odkrivanja novih učinkovin, in je časovno zelo potratno.

Program MetaPrint2D je zasnovan tako, da je njegova uporaba popolnoma avtomatizirana in omogoča združevanje z drugimi programskimi opremami. Temelji na SPORCalc (substrate product occurrence ratio calculator) pristopu, ki za svoje izračune uporablja prstne odtise atomov substratov, ki so objavljeni v podatkovni bazi Metabolite[®]. Pomanjkljivost programa MetaPrint2D je določevanje reakcijskega centra na katerem naj bi potekla metabolična pretvorba (predvsem pri reakcijah dealkiliranja in hidrolize). V drugi, modificirani verziji, so zato določili reakcijske centre za vse metabolične reakcije prve faze s SMARTS kemijskim jezikom.

Vsem naključno izbranim molekulam smo napovedali mesta metabolizma z obema verzijama programov. Pri tem smo upoštevali dve različni opciji programa, »standard« in »loose«. Pri opciji »standard« je okolica vsakega atoma bolj natančno definirana kot pri opciji »loose«. Ugotovili smo, da ima izbira opcij pomemben vpliv na rezultate. Po primerjavi rezultatov obeh verzij programa smo dokazali, da je optimizirana verzija programa bolj zanesljiva, potrdili smo tudi, da so pri reakcijah dealkiliranja in hidrolize reakcijski centri pravilno definirani. Iz dobljenih rezultatov je razvidno, da je z drugo verzijo programa mogoče zadovoljivo napovedati mesta metabolizma za reakcije hidrolize in oksidacije. Prav tako prihajamo do zaključka, da je z drugo verzijo programa možno narediti najbolj zanesljivo napoved, ne glede na izbiro opcije, za srednje velike molekule v območju molekulskih mas od 300 do 500 g/mol. Iz dobljenih rezultatov je razvidno, da je bila optimizacija programa smiselna, in da so nadaljnje nadgradnje možne in zaželeno.

SEZNAM OKRAJŠAV

ATP – adenzintrifosfat

CYP450 – citokrom P 450

FAD – flavin adenzin dinukleotid

FMO – flavinmonooksigenaze

Hidrol. - hidroliza

MAO – monoamin oksidaze

metab. - metabolizem

MP2D – MetaPrint2D

MW – molekulska masa

NADPH – nikotinamid adenin dinukleotid fosfat

NMR – nuklearna magnetna resonanca

Oks. - oksidacija

OR – occurrence ratio (razmerje na atomu)

Redukc. - redukcija

r(i) – normalizirane OR vrednosti

SPORCalc – substrate product occurrence ratio calculator

stand. – »standard«

1 UVOD

1.1 METABOLIZEM KSENOBIOTIKOV

Metabolizem je proces, ki v telesu poteka neprestano. Ksenobiotiki so telesu tuje snovi, brez hranilne vrednosti, zato si jih organizem prizadeva v čim krajšem času izločiti iz telesa. V ta namen so se razvili posebni encimski sistemi, ki pretvarjajo lipofilne snovi v bolj hidrofilne. S to pretvorbo organizem doseže hitrejše in lažje izločanje, ter na ta način preprečuje akumulacijo v telesu. Metabolizem spojin poteka primarno v jetrih. Poleg tega poteka tudi v drugih tkivih: v steni črevesja, v pljučih, v ledvicah, v krvi, v koži in v mišicah.

Metaboliti, ki nastanejo v procesu metabolizma, so lahko aktivni ali neaktivni. Zdravilne učinkovine, ki se metabolizirajo v neaktivne metabolite, imenujemo »soft drugs«. Po drugi strani je možno tudi, da učinkovina pridobi svojo biološko aktivnost šele v procesu metabolizma. V tem primeru govorimo o aktivnih metabolitih, učinkovine iz katerih le-ti nastanejo pa imenujemo predzdravila (prodrugs). Glavna naloga metabolizma ksenobiotikov je eliminacija fiziološko nepomembnih, in v nekaterih primerih telesu škodljivih snovi. (1)

Metabolizem ksenobiotikov je običajno dvofazni proces. Spojine se sprva biotransformirajo z reakcijami I. faze (reakcije funkcionalizacije), temu pa lahko sledijo reakcije II. faze (reakcije konjugacije). Potrebno pa je poudariti, da to zaporedje ne velja v vseh primerih. Na določenih ksenobiotikih potečejo le reakcije II. faze, brez predhodnih reakcij funkcionalizacije. Prav tako se vsi ksenobiotiki po prehodu reakcij I. faze ne biotransformirajo naprej z reakcijami II. faze. Znani so tudi primeri, kjer reakcije funkcionalizacije sledijo reakcijam konjugacije. (1)

V zadnjem času je metabolizem ksenobiotikov postal zelo pomembno področje raziskovanja, ki ima poseben pomen v farmacevtski kemiji, biologiji, toksikologiji kot tudi v drugih farmacevtskih znanostih. (1)

1.1.1 Reakcije I. faze

Imenujemo jih tudi reakcije funkcionalizacije, v katerih se v molekule uvedejo ali tvorijo polarne funkcionalne skupine, kot so hidroksilna, amino ali karboksilna skupina. Obsegajo

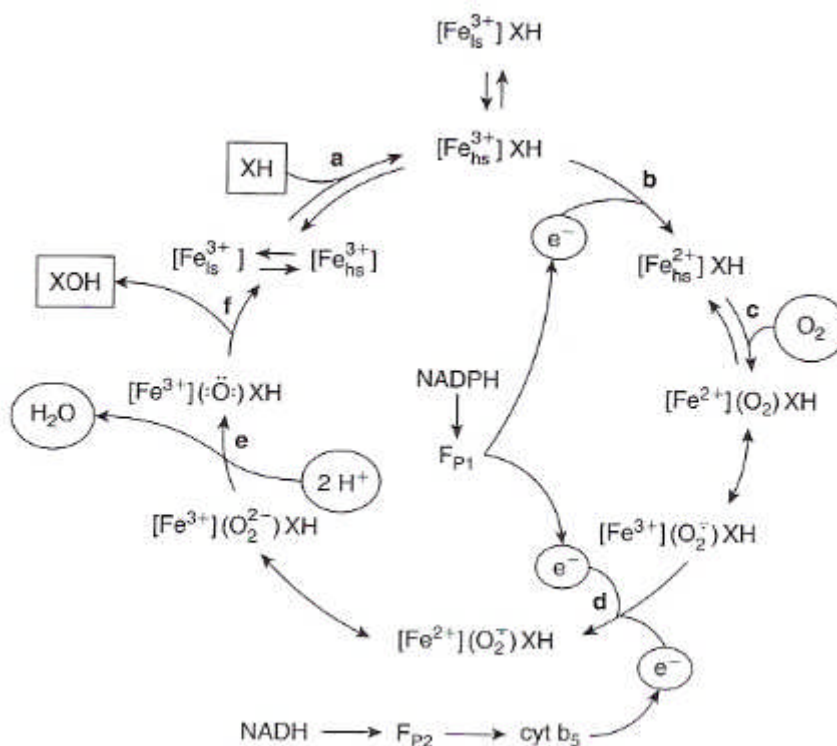
reakcije oksidacije, redukcije in hidrolize. Reakcije oksidacije in redukcije katalizirajo encimi, imenovani oksidoreduktaze, medtem ko reakcije hidrolize katalizirajo hidrolaze. Skupino oksidoreduktaz lahko razdelimo v dve skupini: monooksigenaze in ostale oksidoreduktaze. Pri metabolizmu ksenobiotikov imajo večjo vlogo monooksigenaze, med katere spadajo tudi encimi iz družine citokrom P450 (CYP450). (1)

Monooksigenaze

Med monooksigenaze prištevamo dve družini encimov. Encime iz družine citokrom P450 (CYP450) in encime iz družine flavinmonooksigenaz (FMO). Encimi se med seboj razlikujejo po svojih lastnostih, kot tudi po strukturi. (1) FMO so družina encimov, ki katalizirajo reakcije oksidacije na dušikovem, fosforjevem in žveplozem atomu (primer nastajanja *N*-oksidov, *S*-oksidov). Za svoje delovanje potrebujejo kisik, kofaktor NADPH in flavin adenozin dinukleotid (FAD). (2)

Ko govorimo o metabolizmu ksenobiotikov imajo CYP encimi glavno vlogo. Predstavljajo več kot 75% celotnega metabolizma učinkovin. (1) Najdemo jih v vseh živih organizmih, tako pri bakterijah, kot pri ljudeh. Pri ljudeh in ostalih sesalcih se CYP v najvišjih koncentracijah nahajajo v endoplazmatskem retikulumu v jetrih. V nižjih koncentracijah so prisotni tudi v ostalih tkivih (možgani, pljuča, ledvica,...). (2) Danes poznamo 57 različnih genov, ki kodirajo encime CYP. Obstaja 12 družin CYP, večina ksenobiotikov se metabolizira z CYP1, CYP2 in CYP3. Najpogostejši izocimi, ki so vključeni v metabolizem zdravilnih učinkovin so CYP3A4 in CYP2D6. (1)

Da bi razumeli delovanje CYP450, je potrebno razumeti njegov katalitični cikel (slika 1). (1, 2) Citokrome P450 imenujemo tudi hemproteini, ker imajo v svoji strukturi vezan hem, na katerega je vezano železo. Encim se v svojem oksidiranem stanju (feri oblika, Fe^{3+}) nahaja v ravnotežju med dvema spin stanjema (low-spin in high-spin stanje). Po vezavi substrata (HX) na encim nastane kompleks (high-spin stanje) (*reakcija a*). V naslednjem koraku vstopi v reakcijo elektron, ki reducira železov atom v encimu v fero (Fe^{2+}) obliko (*reakcija b*), ki tvori kompleks s kisikom $[\text{Fe}^{3+}](\text{O}_2)\text{XH}$ (*reakcija c*). Cikel se nadaljuje z vstopom naslednjega elektrona, ki reducira trikomponentni kompleks (*reakcija d*). Z vstopom dveh vodikovih protonov v sistem, se odcepi voda, v kompleksu pa ostane kisikov atom, ki ima v zunanji orbitali 6 elektronov (*reakcija e*). To je reaktivna oblika kisika, ki napade substrat. Iz sistema disocira produkt XOH (*reakcija f*), kar pa omogoči regeneracijo začetnega stanja encima. S tem je katalitični cikel zaključen. (1, 2)



Slika 1: Katalitični cikel delovanja CYP450.

Encimi CYP450 katalizirajo naslednje reakcije:

- alifatsko hidroksiliranje
- aromatsko hidroksiliranje
- oksidacija alkenov, alkinov
- dealkiliranje (*N*-, *O*-, *S*-)
- *N*-oksidacija, *S*-oksidacija
- oksidativno deaminiranje in
- dehalogeniranje.

Ostale oksidoreduktaze

Pri oksidaciji ali redukciji ksenobiotikov imajo večjo ali manjšo vlogo tudi ostale oksidoreduktaze. (1) Ena izmed njih je monoaminska oksidaza (MOA). Obstaja v dveh različnih oblikah, MAO-A in MAO-B. Nahaja se na zunanji membrani mitohondrijev. MAO-A se nahaja predvsem v placenti, črevesju in jetrih. MAO-B pa prevladuje v možganih, jetrih in trombocitih. Njena glavna funkcija v telesu je uravnavanje koncentracije aminskih nevrottransmitterjev, zato se jih izrablja tudi v namene zdravljenja

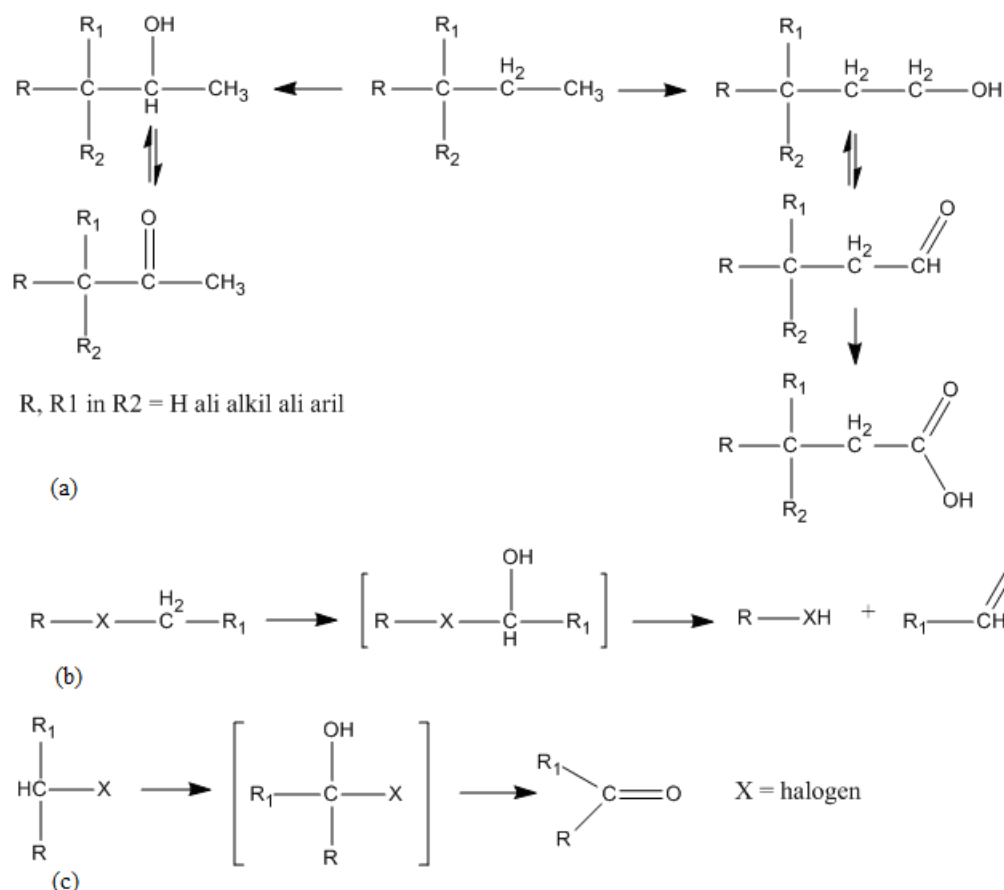
nevrodegenerativnih bolezni, kot je parkinsonova bolezen. (2) V procesu oksidacije sodelujejo tudi diamin oksidaze, aldehyd oksidaze, ksantin oksidaze, različne peroksidaze, alkohol dehidrogenaze, aldehyd dehidrogenaze in aldol-keto reduktaze. (1)

1.1.1.1 Reakcije oksidacije in redukcije

Reakcije oksidacije in redukcije potekajo tako na sp^3 , sp^2 , ali sp - hibridiziranih ogljikovih atomih. Najpomembnejša in najpogostejša reakcija, ki poteka na sp^3 C-atomih je CYP katalizirana alifatska hidroksilacija. Najpogostejše mesto napada je predzadnji C-atom, vendar pa lahko reakcija poteče tudi na zadnjem, ali katerem koli drugem C-atomu. Kje bo reakcija potekla je odvisno od steričnih ovir v molekuli, ter od regiospecifičnosti CYP450. Produkt hidroksilacije, ki pri tem nastane, se lahko v naslednjem koraku pod vplivom dehidrogenaz pretvori v ustrezen aldehyd ali keton. Dehidrogenaze lahko omenjeno pretvorbo katalizirajo v obeh smereh, zato je reakcija reverzibilna. Kadar poteka reakcija v obratni smeri, govorimo o redukcijah. Za aldehyde je značilno, da se pogosto dodatno oksidirajo do karboksilnih kislin, pri čemer gre za ireverzibilno reakcijo (slika 2a). (1)

V molekulah, ki vsebujejo nenasičen sistem ali heteroatom (N, O, ali S), poteče reakcija hidroksilacije na sosednjem C-atomu. Takšni metaboliti so navadno nestabilni, zato sledi hitra post-encimska eliminacija. V odvisnosti od začetnega substrata, lahko kot produkti nastanejo sekundarni ali primarni amini, alkoholi ali tioli. Alkilna skupina se sprosti v obliki aldehyda ali ketona. Opisane reakcije so reakcije *N*-dealkiliranja, *O*-dealkiliranja ali *S*-dealkiliranja (slika 2b). (1)

Oksidativno dehalogeniranje poteka na sp^3 hibridiziranih C-atomih, ki tvorijo vez s halogenom. Kot halogena atoma (X) sta največkrat prisotna klorov ali bromov atom. Reakcija poteka podobno, kot zgoraj opisane reakcije dealkiliranja, s hidroksilacijo in cepitvijo HX (slika 2c). (1)

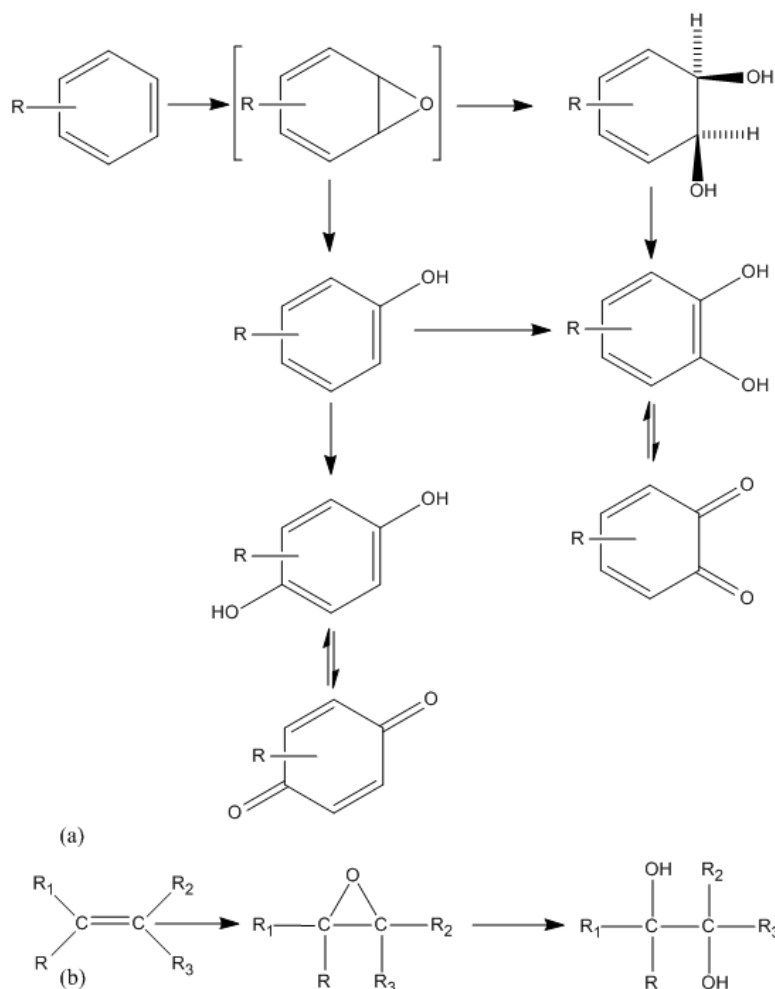


Slika 2: Reakcije oksidacije na sp^3 hibridiziranih C-atomih (a): CYP katalizirana hidroksilacija. (b): CYP katalizirana alfa hidroksilacija (N-dealkiliranje ali deaminiranje, O-dealkiliranje in S-dealkiliranje). (c): CYP katalizirana reakcija oksidativnega dehalogeniranja.

Produkti oksidacije sp^2 C-atomov v aromatskih obročih so lahko različni, vsem pa je skupen nastanek epoksida, kot vmesnega produkta. Epoksidi so reaktivne in nestabilne snovi, zato se v naslednji stopnji s pomočjo epoksid hidrolaz hidrolizirajo do dihidrodiolov, ali pa do fenolov. Dihidrodioli se v naslednji stopnji pod vplivom dihidrodiol dehidrogenaz pretvorijo v katehole. Nesubstituirani fenolni obroči se primarno oksidirajo na para mestu (nastanek dihidrokinonov), lahko pa se oksidirajo tudi na orto mestu, pri čemer nastanejo kateholi. V redkih primerih se kateholi in dihidrokinoni v naslednji stopnji oksidirajo do kinonov. Zadnja stopnja reakcije (oksidacija do kinonov) je reverzibilna (slika 3a). (1)

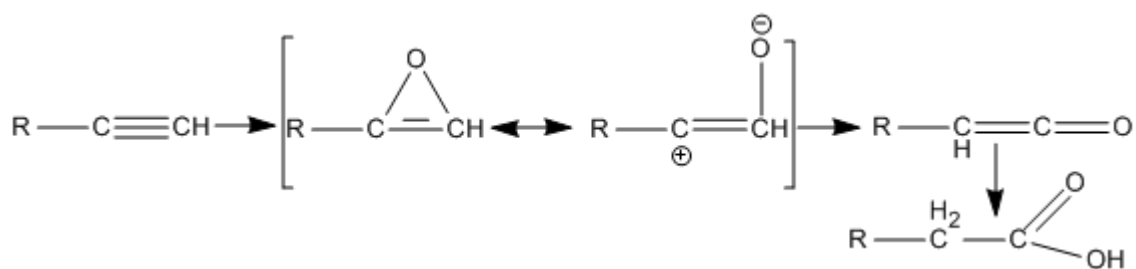
Alkeni so prav tako lahko tarča CYP450 katalizirane epoksidacije. V primerjavi z epoksidi, ki so produkt aromatske oksidacije, so epoksidi, ki nastanejo z oksidacijo alkenov bolj

stabilni in jih je možno izolirati. V večini primerov pa so samo substrat, iz katerega nastanejo dihidrodioli (slika 3b). (1)



Slika 3: Najpogostejše reakcije funkcionalizacije na sp^2 C-atomih.

Tarča CYP450 so tudi sp hibridizirani C-atom. Intermedijat, ki nastane kot posledica oksidacije trojne vezi, lahko reagira na več različnih načinov. Lahko tvori kovalentno vez z encimom, ali pa se pretvori do ketenov. Keteni so zelo reaktivne snovi, ki se po hidrataciji pretvorijo v derivate očetne kisline (slika 4). (1)



Slika 4: Oksidacija trojne vezi.

Reakcije oksidacije N ali S atomov katalizirajo CYP450 ali FMO, medtem ko encimi ki katalizirajo redukcije niso značilni za posamezne reakcije, in so lahko CYP450, NADPH-CYP reduktaze, ali pa kateri izmed ostalih reduktaz. Oksidacija terciarnih aminov, tako alifatskih kot aromatskih, je preprosta reakcija, katere metaboliti so *N*-oksidi. *N*-oksidacija poteka tudi na primarnih in sekundarnih aminih, metaboliti teh reakcij so hidroksilamini. Vse omenjene reakcije *N*-oksidacije so reverzibilne. V nasprotju z sekundarnimi in terciarnimi amini se lahko primarni amini v naslednji stopnji oksidirajo do nitrozo metabolitov. (1)

Nemalokrat zdravilne učinkovine vsebujejo žveplove atome. Žveplo vsebujoče učinkovine so lahko tioli, sulfidi (tioetri) ali tiookso spojine. Metaboliti oksidacije tiolov so v prvi stopnji sulfenske kisline, ki se nato lahko oksidirajo do sulfinske kisline, le-ta pa se v zadnji stopnji oksidira do sulfonske kisline. Te reakcije katalizirajo CYP450 ali FMO. Možna pa je tudi druga pot oksidacije tiolov do disulfidov. Omenjena reakcija poteka kot dehidracija med molekulama tiola in sulfenske kisline. Metabolizem sulfidov je precej enostaven. Kot produkt oksidacije v prvi stopnji nastanejo sulfoksidi, ki se v naslednji stopnji oksidirajo do sulfonov. Prva stopnja reakcije je reverzibilna, medtem ko je druga stopnja ireverzibilna. Produkta oksidacije tiokso spojin, *S*-okside (sulfine) in v naslednji stopnji *S*-diokside (sulfene), zaradi njihove visoke reaktivnosti ni mogoče izolirati. Za *S*-diokside je značilno, da so zelo elektrofilni, in lahko hitro reagirajo z nukleofili in nukleofilnimi deli bioloških makromolekul. Ta kovalentna vezava je vzrok toksikološkega delovanja številnih tioamidov. (1)

1.1.1.2 Reakcije hidrolize

Reakcije hidrolize potekajo ob prisotnosti molekul vode, pri čemer pride po adiciji molekule vode do cepitve vezi. Hidrolize najpogosteje potekajo na molekulah estrov in amidov. Kadar govorimo o hidrolizi epoksidov pogosto uporabimo izraz hidratacija epoksida, ki jo katalizirajo epoksid hidrolaze. (1)

1.1.2 Reakcije II. faze

Reakcije II. faze, oziroma reakcije konjugacije, imajo velik pomen pri metabolizmu endogenih molekul kot tudi pri metabolizmu ostalih ksenobiotikov. Za te reakcije je

značilno, da ponavadi sledijo reakcijam I. faze in da jih katalizirajo encimi transferaze. Pri reakciji sodelujejo kofaktorji, ki so nosilci endogenih molekul. Nastanejo kovalentne vezi med endogenimi molekulami in reaktivnimi skupinami ksenobiotikov. Metaboliti, ki pri tem nastanejo, so običajno neaktivni in bolj polarni. Reakcije ki jih uvrščamo med reakcije II. faze so:

- glukuronidacija
- metiliranje
- sulfatiranje
- acetiliranje
- konjugacija s koencimom A
- konjugacija z glutationom. (1)

GLUKURONIDACIJA

Je najpogostejša reakcija med konjugacijami. Reakcije glukuronidacije katalizira družina izoencimov, imenovanih glukuronil transferaze, ki katalizirajo prenos glukuronske kisline iz kofaktorja uridin-5'-difosfo- α -D-glukuronske kisline. Metaboliti, ki pri tem nastanejo, so bistveno bolj polarni. Za reakcije glukuronidacije je značilno, da imajo nizko afiniteto in visoko kapaciteto, ki je posledica zaloga glukuronske kisline v jetrih. (1)

METILIRANJE

Pri reakciji metiliranja pride do prenosa metilne skupine iz kofaktorja S-adenozil-L-metionina, pri čemer sodelujejo številne metiltransferaze. Reakcije metiliranja so značilne za fenole (katehole), amine in tiole. (1)

SULFATIRANJE

Tvorbo sulfatov katalizirajo različne sulfotransferaze. Reakcije potekajo na alkoholih, fenolih in aminih. Nukleofilna $-OH$ ali $-NH_2$ skupina substrata reagira z SO_3^- skupino kofaktorja. Konjugati so večinoma neaktivni in polarni. (1)

ACETILIRANJE

Reakcije acetiliranja potekajo predvsem na primarnih aminih, zato so najpogostejši encimi v sklopu teh reakcij *N*-acetilttransferaze. Reakcije na kateholih ali fenolih katalizirajo *O*-

acetiltransferaze. Kofaktor, ki pri omenjenih reakcijah prenese acetilno skupino na substrat je acetilkoencim A. (1)

KONJUGACIJA S KOENCIMOM A

Reakcija poteka na ksenobiotikih, ki imajo v molekuli prosto $-COOH$ skupino. Pri tem nastane acetyl-CoA tioester ($R-CO-S-CoA$). Za potek reakcije je potrebna energija v obliki ATP in prisotnost encimov acetyl-CoA sintetaz. (1)

KONJUGACIJA Z GLUTATIONOM

Glutation je tripeptid, ki se sintetizira iz glicina, glutaminske kisline in cisteina. Vsebuje nukleofilno tiolno skupino, ki ima pomembno vlogo v procesu detoksifikacije reaktivnih in potencialno toksičnih ksenobiotikov. V telesu se nahaja v ravnotežju med dvema oblikama, reducirano (GSH) in oksidirano (GS-SG). Zavira delovanje radikalov, zato ima pomembno vlogo pri celični zaščiti. (1)

1.2 POMEN NAPOVEDOVANJA MEST METABOLIZMA V MOLEKULAH IN METABOLIČNE STABILNOSTI V ZGODNJIH FAZAH RAZVOJA UČINKOVIN

Odkrivanje in razvoj novih učinkovin je dolgotrajen proces, ki ne vodi vedno do želenih rezultatov. V grobem ga lahko razdelimo v 3 faze. V prvi fazi je potrebno opredeliti, kakšen farmakološki učinek mora imeti novo razvita učinkovina. Zato je v tem koraku potrebno identificirati in validirati tarčo učinkovine. Pri tem je pomembno, da se določi patogeneza oz. potek bolezni, da zmanjšamo stranske učinke na delovanje ostalih fizioloških procesov v telesu. Ko je tarča odkrita in določena, je proces odkrivanja potrebno usmeriti na razvoj novih spojin vodnic. Spojine vodnice so molekule, ki kažejo zelene farmakološke aktivnosti in so osnova za razvoj končne učinkovine. Osnovno strukturo spojine vodnice je nato potrebno modificirati, oziroma optimizirati. S tem se izboljšajo njene lastnosti kot so učinkovitost, metabolična stabilnost, topnost, permeabilnost in varnost. Spojina gre nato v številne predklinične študije, ki se izvajajo *in vitro* ter *in vivo* na živalskih modelih. Spojine, ki prestanejo vse predklinične študije, se vključi v preizkuse na prostovoljcih, kar imenujemo klinične študije. Šele ko so dokazani

varnost, učinkovitost in kakovost učinkovine se lahko zaprosi za dovoljenje za njeno uporabo. Ugotavljanje njene varnosti pa poteka tudi po tem, ko je že na trgu. (3)

Številne molekule, ki so se prebile skozi predklinične študije, kliničnega preizkušanja ne prestanejo uspešno, zato so umaknjene iz procesa razvoja novih učinkovin. Pogosto je vzrok za ustavitev razvoja nove učinkovine prav njena toksičnost, za katero so pogosto vzrok reaktivni metaboliti, ki nastanejo v procesu biotransformacije. Pozno odkritje metabolične nestabilnosti povzroča visoke stroške odkrivanja novih učinkovin, in je časovno zelo potratno. (4)

Napovedovanje metabolizma v molekulah lahko močno vpliva na celoten potek razvoja novih učinkovin. (5) Ker so biološka razpoložljivost, učinkovitost, toksičnost, porazdelitev in tudi eliminacija molekul odvisne od poteka metabolizma, je zelo smotno, da se te informacije pridobijo že v zgodnjih fazah odkrivanja novih učinkovin. (6) Poznavanje metabolne poti učinkovine je lahko znanstvenikom v veliko pomoč pri odločitvah, ali naj izbrano spojino izločijo iz nadaljnjega procesa raziskav ali nadaljujejo z njeno optimizacijo, z namenom zagotavljanja stabilnosti in varnosti. (5, 7)

Če je molekula metabolično nestabilna, je pomembno, da se določi mesto metabolizma v molekuli. (8) Šele ko vemo, kje v molekuli poteka metabolična reakcija, lahko v nadaljnjem procesu molekulo optimiziramo ali del molekule odstranimo ali pa z uvedbo določenih funkcionalnih skupin onemogočimo metabolizem na tem mestu. (8, 9) Prepoznavanje mesta metabolizma pa je pomembno tudi pri razvoju predzdravil, kjer v procesu metabolizma nastane aktivna spojina z želenim delovanjem. (7) V zadnjem času je velika pozornost usmerjena v razvoj in silico metod za napovedovanje mest metabolizma v molekulah, s katerimi že v zelo zgodnjih fazah odkrivanja novih učinkovin izločimo potencialno reaktivne in metabolično nestabilne molekule. (4)

Če na kratko povzamemo, pogosto največji izziv v procesu odkrivanja novih učinkovin predstavlja faza odkrivanja novih spojin vodnic in njihova optimizacija. (3) Napovedovanje metabolizma in metabolične stabilnosti v zgodnjih fazah razvoja učinkovin omogoča, da že zgodaj iz procesa izločimo molekule, ki kažejo neugodne metabolične lastnosti v organizmu. S pomočjo dobrega in silico modela lahko to dosežemo, zato je vedno več raziskav usmerjenih v razvoj novih in izboljševanju že obstoječih in silico modelov.

V diplomski nalogi predstavljamo nekatere že obstoječe in silico metode z njihovimi prednostmi in slabostmi. Glavna tema naloge pa bo predstavitev računalniškega programa MetaPrint2D (MP2D) in rezultatov pridobljenih z evaluacijo tega programa.

1.3 IN SILICO METODE ZA NAPOVEDOVANJE MEST METABOLIZMA V MOLEKULAH

Uporaba in silico metod za določanje mest metabolizma v molekulah je začetna točka preučevanja metabolnih poti za večino novih molekul. V splošnem lahko metode razdelimo v dve skupini: tiste, ki temeljijo na tarčah (target-based) in tiste, ki temeljijo na osnovi podatkov o ligandih (ligand-based). (3, 8) Metode na osnovi tarče so usmerjene na interakcije med aktivnim mestom encima in posamezno spojino. Pri tej metodi je potrebno poznati 3D strukturo tarče. Za metode, ki temeljijo na podatkih o ligandih, je značilno, da temelji napoved metabolizma na osnovi podobnosti strukture preiskovane učinkovine z znanimi molekulami. Obe metodi imata pozitivne in negativne lastnosti, vendar pa se metode na osnovi liganda pogosteje uporabljajo, predvsem zaradi večje razpoložljivosti ustreznih podatkov. Metodi med seboj nista izključujoči, zato se v nekaterih primerih med seboj prepletata oziroma dopolnjujeta. (3) Primer metode, ki uporablja oba pristopa, je MetaSite. (7) Najpogosteje uporabljena metoda na osnovi tarče je sidranje v aktivno mesto encima (doking). (3) Vse pogosteje pa vstopajo v uporabo tudi »data mining« metode kot je na primer MetaPrint2D metoda. Omenjena metoda spada med ligand-based metode in bo osnova našega raziskovalnega dela v diplomski nalogi. Podrobneje je predstavljena v naslednjem poglavju.

1.3.1 MetaSite

Metoda MetaSite obsega izračun dveh sklopov deskriptorjev. En sklop deskriptorjev predstavljajo strukture CYP encimov, drugega pa molekula substrata. Deskriptorji, ki predstavljajo CYP encime temeljijo na GRID zapisu. (10, 11) Tudi vsi atomi v molekuli substrata so določeni z GRID probami in so razvrščeni glede na 4 osnovne lastnosti: hidrofobnost, donor vodikove vezi, akceptor vodikove vezi ali sposobnost tvorbe naboja. Oba sklopa deskriptorjev se nato primerjata med seboj. (7,11) Rezultati predstavljajo prileganje vsakega atoma substrata v aktivno mesto encima. Izračunani so po naslednji enačbi:

$$P_{SM,i} = E_i R_i$$

(enačba 1)

E_i v enačbi 1 ponazarja prileganje molekule substrata v encim, ko je atom i v molekuli substrata usmerjen proti hemu v encimu. Reaktivnost atoma i ponazarja R_i v zgornji enačbi. Kadar poznamo mehanizem poteka reakcije, lahko R_i predstavlja aktivacijsko energijo, ki je potrebna za nastanek reaktivnih intermediatov. Reaktivnost atoma i je odvisna od 3D strukture liganda in mehanizma reakcije. Ko sta oba, E_i in R_i , določena, je mesto metabolizma opisano kot verjetnostna funkcija P_{SM} in predstavlja verjetnost metabolizma na tem atomu v molekuli. Za en par liganda in CYP encima, so izračunane različne P_{SM} , za vsak atom liganda posebej. Atom, ki ima najvišjo izračunano vrednost P_{SM} , predstavlja mesto v molekuli, kjer je največja verjetnost, da bo potekel metabolizem. Metabolizem torej poteče na mestih, ki so dobro izpostavljeni in imajo visoko energijo. (7) Pogosto se MetaSite uporablja v kombinaciji z drugimi tehnikami, saj jih zelo dobro dopolnjuje. Rezultati pridobljeni z LC-MS-MS spektroskopijo nam pogosto ne dajejo dovolj podatkov, da bi lahko natančno določili, kje v molekuli bo reakcija potekala ter ne dobimo podatka o tem katera reakcija bo potekla in na katerem atomu. V kombinaciji z programom MetaSite lahko točno določimo mesto metabolizma v molekuli ter kateri encim reakcijo katalizira. Dobljeni rezultati so v pomoč raziskovalcem pri odločitvah o nadaljnjih modifikacijah v molekuli. (7)

Program MetaSite je bil zasnovan za delo s petimi najpomembnejšimi človeškimi citokromi. Kljub temu je z njim mogoče napovedati metabolizem katerega koli encima, zato se uporablja za napoved metabolizma tako v človeškem telesu, ko tudi v ribah, rastlinah in bakterijah. Celoten postopek je popolnoma avtomatiziran, poznana mora biti le 3D struktura encima. Metoda je enostavna za uporabo, veliko prednost pa predstavlja tudi hitrost procesa. (7)

1.3.2 QSAR

Pristop QSAR spada med ligand-based metode. S temi metodami pridobimo posredne informacije o aktivnem mestu proteina, ki temeljijo na obliki, elektronskih lastnostih in konformaciji substrata, inhibitorja ali metabolita. Odvisne so od razpoložljivosti eksperimentalnih podatkov dovolj velikega števila substratov in njihovih konformacij. QSAR predstavlja enostaven pristop napovedovanja aktivnosti in/ali vezavne sposobnosti

molekule. Poskuša povezati kemijsko strukturo molekule z njeno aktivnostjo, predstavljeno na kvantitativen način. Izmerjene ali izračunane lastnosti, imenovane deskriptorji, so med seboj povezane ter odvisne ena od druge. Deskriptorji opišejo površino molekule in receptorja, na kateri pride do vezave med njima, preračunane vrednosti pa predstavljajo energijo interakcije med molekulo in receptorjem. (12, 13) Pri pripravi modelov iz velikega števila deskriptorjev, zlasti kadar je na voljo malo eksperimentalnih podatkov, je izbor ustreznih deskriptorjev bistvenega pomena. Izbor deskriptorjev vpliva na hitrost celotnega procesa, od njega pa je tudi odvisno, ali bo model kompleksen, ali enostavno razumljiv. Pomembnost deskriptorja je močno odvisna od okolice v kateri ga določamo. V neki študiji so prišli do ugotovitve, da logP (eden od najbolj razširjenih deskriptorjev v QSAR) za predikcijo biološke aktivnosti, ni nič bolj pomemben, kot naključno izbrani deskriptorji. (3) Večina študij, ki so bile narejene na CYP450, so vključevale le tiste CYP450, ki imajo največji pomen za metabolizem ksenobiotikov. (12)

1.3.3 Doking

Metoda doking spada med najpogosteje uporabljene target-based metode. Preučuje vezavo liganda v aktivno mesto encima ali receptorja. (14) Prileganje v aktivno mesto se meri s funkcijo, ki temelji na interakcijah med ligandom in tarčo. Najpomembnejše so reakcije pri katerih je interakcija med ligandom in tarčo največja oziroma najmočnejša. Za takšne primere se predpostavi, da gre za biološko pomembno interakcijo. (3) Predpogoj za zanesljivo predikcijo je razpoložljiva kristalna struktura encima, ki sodeluje v reakciji. Pomembni pa so tudi drugi parametri, kot je fleksibilnost proteina in interakcije z molekulami vode, ki lahko zelo otežijo uporabnost metode. (14)

3D struktura proteina, ki mora biti poznana, se lahko določi z različnimi eksperimentalnimi tehnikami, kot so X-ray kristalografija, nuklearna magnetna resonanca (NMR) ali elektronska mikroskopija. Žal pa je mnoge proteine težko kristalizirati, zato kristalna struktura proteina ni vedno na voljo. V tem primeru lahko za določanje strukture proteina uporabljamo homologno modeliranje. (12) Osnova homolognega modeliranja temelji na rezultatih, da imajo proteini s podobnim aminokislinskim zaporedjem podobno 3D strukturo. 3D struktura proteina je tako določena na osnovi njegovega aminokislinskega zaporedja in 3D strukture proteina z podobnim zaporedjem. (15) Čeprav so ti modeli manj natančni od eksperimentalnih, so v primeru pomanjkanja eksperimentalnih podatkov

neprecenljive vrednosti. Do nedavnega je poznan model za človeški CYP450 temeljil na bakterijskem P450. Ko so uspeli pridobiti prvo kristalno strukturo sesalčevega CYP450 (zajčev CYP2C5), je to pomenilo velik korak naprej k določitvi ustreznega modela. (12) Slabost doking metode so neupoštevanje fleksibilnosti proteina in solvatacije. Študije so pokazale, da lahko vezane molekule vode povečajo, kot tudi zmanjšajo natančnost celotnega procesa. (14)

1.4 PROGRAM METAPRINT2D KOT ORODJE ZA NAPOVEDOVANJE MEST METABOLIZMA V MOLEKULAH

Razvitih je bilo veliko različnih metod za napovedovanje mest metabolizma ksenobiotikov. Mnoge med njimi so osredotočene le na transformacije, ki jih katalizirajo CYP450, ostalih metaboličnih reakcij pa sploh ne upoštevajo. Majhen delež metod je javno dostopnih in enostavnih za uporabo. Med javno dostopnimi pa se pogosto najdejo tiste, ki so plačljive. (3)

Z željo po izboljšanju omenjenih negativnih lastnosti že znanih programov, so znanstveniki razvili nov program, imenovan MetaPrint2D (MP2D), ki temelji na SPORCalc (substrate product occurrence ratio calculator) pristopu. Program je zasnovan tako, da je njegova uporaba popolnoma avtomatizirana in omogoča združevanje z drugimi programskimi opremami. (3) SPORCalc pristop za svoje izračune uporablja prstne odtise atomov substratov, ki so objavljeni v podatkovni bazi Metabolite[®]. (16)

METABOLITE DATABASE

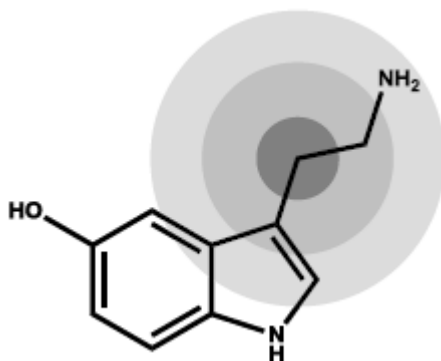
Metabolite[®] podatkovna baza je zbirka podatkov o metabolnih reakcijah, ter predstavlja vir metabolnih transformacij ksenobiotikov pridobljenih iz literarnih podatkov. Vsebuje datoteke, od katerih vsaka vsebuje eno ali več individualnih biotransformacij. Gre za biotransformacije reakcij I. in II. faze. (16) Obstoječa baza podatkov iz leta 2005 (ki smo jo uporabljali tudi pri naših raziskavah v tej diplomski nalogi) vsebuje več kot 73.000 reakcijskih parov. Vsako leto se baza izpopolni s približno 5.000 na novo dodanimi biotransformacijami. (3)

Z uporabo Metabolite[®] podatkovne baze lahko pridobimo informacijo o tem, ali je substrat, ki ga preučujemo, vključen v kakšno metabolno reakcijo ali ne. (9) Metoda je osredotočena

na odkrivanje vseh reakcij I. faze za vse živalske vrste z enim iskalnim nizom, lahko pa iskalni niz išče reakcije le za določeno živalsko vrsto ali določen tip reakcije (primer: vse reakcije hidroksilacije določene v podganah ali reakcije dealkiliranja pri ljudeh). Ne glede na to, kakšen iskalni niz smo izbrali, rezultati ki jih pri tem dobimo, predstavljajo celoten nabor podatkov v RD datoteki. Omenjena datoteka vsebuje reakcijske pare, kjer sta reaktant in produkt navedena v MOL formatu. (16)

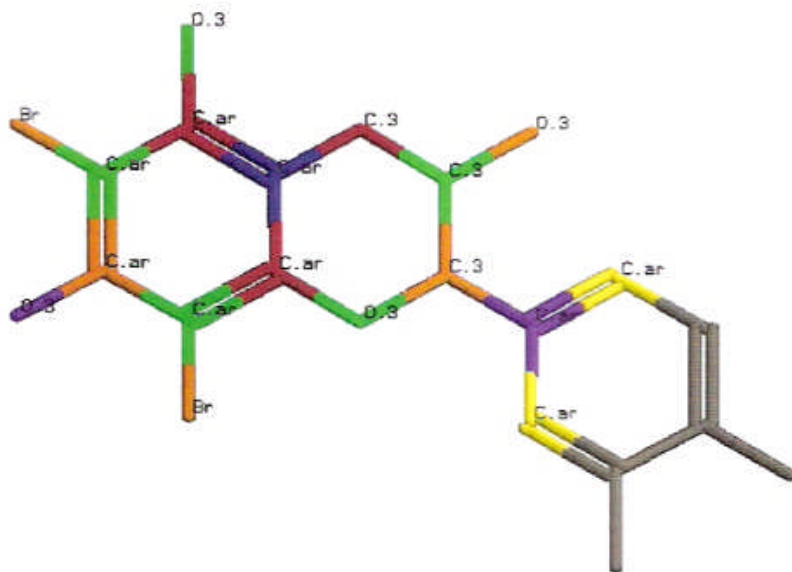
PRSTNI ODTISI

Deskriptor, imenovan prstni odtis (fingerprint), se uporablja za določanje podobnosti med molekulami (similarity). (17) Predstavlja hierarhičen opis atomov ter ga zaradi načina opisa imenujemo tudi krožni prstni odtis (circular fingerprint) (slika 5). (3, 18)



Slika 5: Grafičen hierarhičen opis atoma s krožnim prstnim odtisom.

Okolje, ki je pomembno za računalniško aplikacijo, je lahko predstavljeno s poljubnim številom stopenj oziroma s poljubno oddaljenostjo od preiskovanega atoma. MP2D opiše okolje vsakega atoma s šestimi stopnjami. Prva stopnja predstavlja atom, ki ga preiskujemo. V naslednjo stopnjo so vključeni vsi atomi, ki so od preiskovanega atoma oddaljeni za eno vez oziroma so nanj neposredno vezani. V naslednji stopnji so atomi, oddaljeni za dve vezi in tako naprej (slika 6). Iz poznanih prstnih odtisov se naredita dve zbirki podatkov. Prva zbirka podatkov vsebuje vse atome substratov v podatkovni bazi, v drugi (manjši) zbirki pa so le prstni odtisi reakcijskih centrov v molekulah. Reakcijski center molekule se predhodno določi. Metabolite[®] podatkovna baza vsebuje obe zbirki podatkov za vsak substrat. (16)



		Sybyl Atom Type																																			
		C.3	C.2	C.ar	C.1	N.3	N.2	N.1	O.3	O.2	S.3	N.ar	P.3	H	Br	Cl	F	I	S.2	N.pl3	L.P	Nn	K	Ca	Li	Al	Du	Si	N.am	S.O	S.O2	N.4	O.C	C.ent			
Fingerprint Level	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	2	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	3	1	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	4	1	0	1	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Slika 6: Matrica prstnega odtisa s šestimi stopnjami. Moder C – aromatski atom predstavlja osnovni atom raziskave. (16)

IZRAČUN RAZMERIJ

Vsako novo molekulo, ki jo program preučuje, najprej pretvori v MOL format. Določeni morajo biti tudi prstni odtisi za vse atome v molekuli substrata in metabolita. Prstni odtisi se nato primerjajo z obstoječimi prstnimi odtisi v bazi podatkov. Primerja se jih z obema zbirkama, tako s tisto, ki vsebuje vse atome substratov, kot tudi s tisto, ki vsebuje le prstne odtise reakcijskih centrov. Pomembno je ujemanje oziroma podobnost matric prstnega odtisa nove molekule z matricami prstnih odtisov v bazi podatkov. Dobljeni rezultat je odvisen tudi od tega kakšen način primerjanja je bil izbran. Na voljo so trije različni načini: »loose«, »standard« in »strict«. Način »strict« je najstrožji, zato je v tem primeru okolica

zelo pomembna in mora biti ujemanje med matricami največje. Najšibkejšo ujemanje med matricami je v načinu »loose«. (16)

Izračun razmerja sledi iz naslednje enačbe:

$$OR(i) = \Sigma t(i) / \Sigma s(i)$$

(enačba 2)

Pri čemer je $\Sigma t(i)$ v enačbi 2 vsota vseh dogodkov biotransformacij (v primeru da gre za hidroksilacijo torej vseh hidroksilacij) v bazi podatkov na atomu i , $\Sigma s(i)$ pa predstavlja vsoto vseh biotransformacij (npr. hidroksilacij) v celotni bazi podatkov na vseh atomih. $OR(i)$ (occurrence ratio), oziroma izračunano razmerje, je tako vedno manjše ali enako ena. Končni rezultat, ki ga program poda kot rešitev biotransformacije, so normalizirane $r(i)$ vrednosti (enačba 3). Vrednosti so normalizirane glede na najvišjo OR vrednost ($OR(max)$) v molekuli. Najvišja OR vrednost je normalizirana na 1,00 in predstavlja največjo verjetnost, da bo na tem mestu potekla biotransformacija. (9)

$$r(i) = OR(i) / OR(max)$$

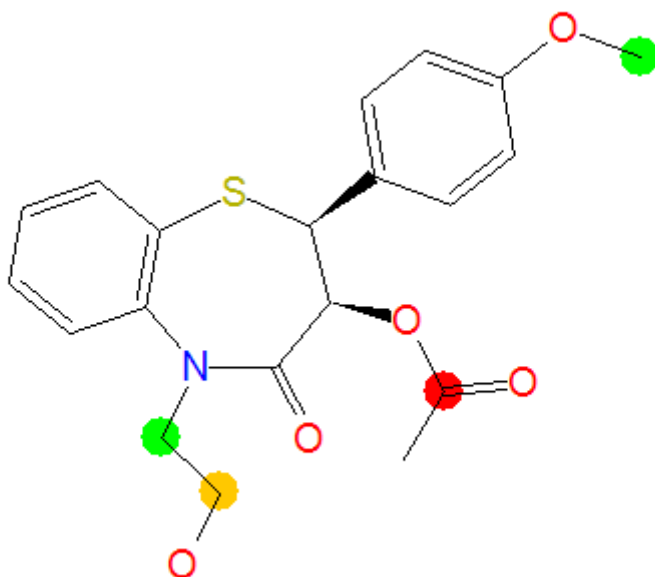
(enačba 3)

Atomi v preiskovani molekuli so nato glede na $r(i)$ vrednost obarvani v belo, zeleno, oranžno ali rdeče. Pri tem veljajo naslednja pravila:

- belo, $0 \leq r(i) < 0,15$
- zeleno, $0,15 \leq r(i) < 0,33$
- oranžno, $0,33 \leq r(i) < 0,66$
- rdeče, $0,66 \leq r(i) \leq 1,00$

(16)

Primer napovedovanja metabolizma z metodo MP2D prikazuje slika 7.

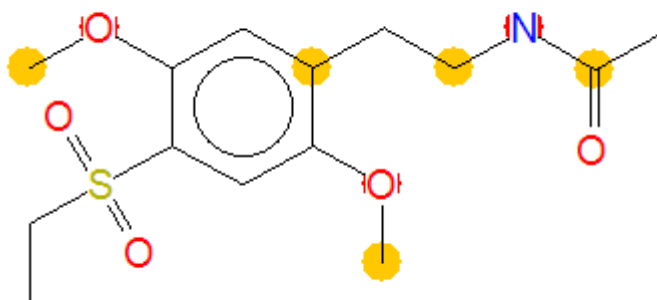


Slika 7: Primer napovedi metabolizma 5-(2-hidroksietil)-2-(4-metoksifenil)-4-okso-2,3,4,5-tetrahidro-1,5-benzotiazepin-3-il acetata v programu MetaPrint2D.

Na ta način lahko takoj vidimo, na katerih mestih bodo potekle transformacije. Metabolno najmanj stabilen je atom obarvan rdeče in predstavlja prvo mesto, na katerem bi naj z največjo verjetnostjo potekel metabolizem. Sledi mu oranžen, najmanjša verjetnost metaboliziranja pa je na zelenih in belih oziroma neoznačenih atomih. S pomočjo dobljenih rezultatov se lahko raziskovalci odločijo, katera mesta v molekuli bi bilo dobro zaščititi, zamenjati ali pa jih pustiti takšna kot so. Označena mesta v atomu pa ne dajo nobenih informacij o tem, katera reakcija bo potekla, ampak le kje bo potekla. (16)

2 NAMEN DELA

MetaPrint2D je »data mining« metoda za napovedovanje mest metabolizma v molekulah, ki jo razvijajo v AstraZeneci na Švedskem. Ena od pomanjkljivosti programa MP2D je določevanje reakcijskega centra na katerem naj bi potekla metabolična pretvorba. Predvsem pri reakcijah hidrolize in dealkiliranja sta kot reakcijska centra označena dva atoma. V primeru reakcije *O*-dealkiliranja sta pogosto kot reakcijska centra označena kisikov atom in sosednji ogljikov atom, ki je tudi dejansko reakcijski center (slika 8).



Slika 8: Primer reakcije O-demetiliranja v starem programu MP2D z opcijo »standard«.
Kot reakcijski center demetiliranja je označen tudi kisikov atom, na katerem reakcija sploh ne poteka.

Na naši fakulteti so določili reakcijske centre za vse metabolične reakcije prve faze s SMARTS kemijskim jezikom, s katerim se lahko opišejo reakcije. Namen te diplomske naloge je ovrednotiti izboljšavo z definiranjem reakcijskih centrov v metaboličnih reakcijah prve faze. Rezultate napovedovanja mest metabolizma bomo pridobili z obema verzijama programa. Namen dela naše diplomske naloge je primerjati drugo verzijo programa s prvo. Zanima nas, ali bodo rezultati, dobljeni z drugo verzijo programa bolj zanesljivi, kot rezultati prve verzije programa. Primerjali bomo rezultate dveh opcij programa, »standard« in »loose«. Osredotočali pa se ne bomo le na primerjavo med obema verzijama programa, ampak želimo boljše raziskati kako vpliva izbor različnih opcij programa na rezultate napovedovanja. Zato bomo primerjali rezultate dobljene z opcijo »standard« z rezultati opcije »loose«. S tem bomo tudi ugotovili, kako pomembno vpliva

okolica atoma na zanesljivost rezultatov. Ugotoviti želimo tudi, kakšen vpliv imata molekulska masa in tip reakcije na napovedovanje mest metabolizma v molekulah.

3 MATERIALI IN METODE

3.1 IZBOR SPOJIN IN DOLOČANJE MEST METABOLIZMA

Program MP2D uporablja podatkovno bazo metaboličnih reakcij Symyx[®], ki je ena najboljšežnejših zbirk ksenobiotikov in njihovih biotransformacij s poudarkom na zdravilnih učinkovinah in njihovih metabolnih poteh. Vsebuje tudi eksperimentalne podatke in vivo ter in vitro študij metaboličnih pretvorb. (19) Baza podatkov iz leta 2005 vsebuje 73.599 metaboličnih reakcij oz. reakcijskih parov, vsako leto pa je v bazo dodanih približno 5.000 novih reakcij. Vsaka reakcija je obrazložena z vrsto različnih informacij, vključno s sklicevanjem na literaturo, podrobnostih o živalskih vrstah in sistemih v katerih je prišlo do reakcije in klasifikaciji tipa reakcije. Vendar pa vsaka metabolična reakcija ni nujno obrazložena z vsemi omenjenimi informacijami. Reakcija sestoji iz ene molekule reaktanta (substrata) in ene molekule produkta (metabolita). Ena molekula substrata se lahko metabolizira z različnimi reakcijami, kar vodi do nastanka različnih produktov. (3)

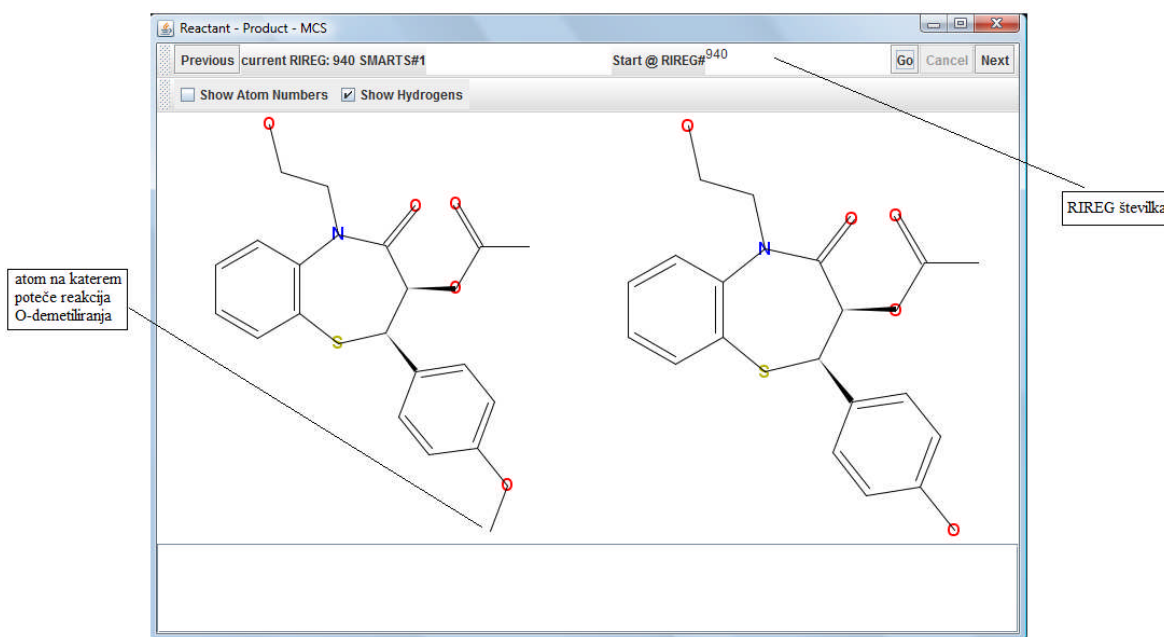
V MP2D program so bile vključene vse metabolične pretvorbe do vključno leta 2005. Pri vrednotenju izboljšave programa smo izbrali set spojin (testni set), ki so v bazi podatkov iz leta 2009, vendar jih ni v bazi iz leta 2005. Vse spojine, ki so bile v bazi do leta 2005, so že vključene v program kot njegova osnova, z točno določenimi mesti metabolizma in prstnimi odtisi atomov. Ti podatki, ki jih program MP2D že vsebuje, omogočajo napoved metabolizma na novih molekula, preveriti pa je potrebno, ali je napoved pravilna.

Med letoma 2005 in 2009 je bilo v bazo vpisanih 3152 novih molekul. Med temi smo s pomočjo programa Excel naključno izbrali 229 reakcijskih parov. Vsaka izhodna molekula (reaktant) je v podatkovno bazo vpisana pod točno določeno zaporedno RIREG številko.

V naslednjem koraku smo si pomagali z aplikacijo »Reactant – Product – MCS«. Ta aplikacija vsebuje vse reakcijske pare, ki jih vsebuje baza iz leta 2009. Z vpisom RIREG številke v iskalno okno, program poišče reakcijski par pod to zaporedno številko (slika 9). Tako smo za vseh 229 spojin, za vsako posebej vpisali RIREG številko in počakali, da je program poiskal ustrezen reakcijski par. Za vsak reakcijski par smo določili, ali gre za reakcijo I. ali II. faze. Pri reakcijah II. faze smo nato le še določili tip reakcije (konjugacija z glukuronsko kislino, s sulfatom,...), ter jih izločili iz nadaljnjih raziskav, ker nas v tej fazi razvoja programa zanimajo le reakcije I. faze.

Vsem reakcijam I. faze smo nato določili tip reakcije, ter atom, na katerem reakcija poteka (reakcijski center). Tako smo vsem reakcijam I. faze določili:

- tip reakcije (oksidacija, redukcija ali hidroliza)
- atom, na katerem reakcija poteče (npr. C3, O1, P,...)



Slika 9: Primer iskalnega niza reakcijskega para s pomočjo aplikacije »Reactant – Product – MCS«, za odgovarjajočo RIREG številko 940.

Na ta način smo ovrednotili 150 reakcij I. faze. Set teh 150 reakcijskih parov je postal naš testni set za testiranje napovedovanja mest metabolizma s programom MP2D.

3.2 RAZLIKE IN PODOBNOSTI MED PROGRAMOMA MP2D (PRIMERJAVA PRVE VERZIJE PROGRAMA Z DRUGO)

Obe različici programa, s katerima smo operirali, sta vključeni v Bioclips, ki predstavlja grafično podlago za naravoslovne vede in vsebuje orodja za različne farmacevtsko-kemijske aplikacije. Z njim je mogoče iskanje in pretvorba med datotekami različnih formatov, urejanje kemijskih struktur, interaktivna vizualizacija 2D ali 3D strukture in urejanje zbirk in knjižnic spojin. Primeren pa je tudi za načrtovanje novih spojin. (20)

Z vključitvijo programa MP2D v sistem Bioclips, je na voljo veliko novih možnosti vizualizacije kemijskih struktur. Sistem podpira datoteke različnih formatov: MDL Molfile, Chemical Markup Language (CML), SMILES in Structure Data File (SDF). (20)

Pred pričetkom dela smo v programa vnesli vseh 229 struktur molekul iz testnega seta v sdf formatu. Program omogoča, da lahko na namizju vidimo vse molekule zaporedno, in da si nato izberemo tisto, za katero želimo izvesti napoved metabolizma. Oba programa, tako prva verzija kot tudi druga verzija programa, omogočata izbiro različnih opcij: »loose«, »standard« in »strict«. Kot je bilo omenjeno že v uvodnem delu, pri izbiri opcije »loose« okolica ni zelo pomembna, za razliko od opcije »strict«, kjer je okolica atoma zelo natančno določena. Specifična orodja omogočajo uporabniku možnost izbire med različnimi podatkovnimi bazami, ki vsebujejo podatke za različne vrste. Mi smo pri delu uporabljali podatke zbrane v vseh podatkovnih bazah, ki so na voljo.

Pomembno razliko med programoma predstavlja način zapisa poteka reakcij v podatkovni bazi. Osnova obeh programov so prstni odtisi vseh atomov reaktantov in produktov, in reakcijskih centrov, ter preračunana razmerja za vsak atom. Na ta način je možno primerjati razliko v strukturi med reaktantom in produktom. Velika pridobitev druge verzije programa pa je ta, da so vse reakcije I. faze definirane s SMARTS kodami.

SMARTS je jezik, s katerim je možno zapisati molekulske vzorce z uporabo določenih pravil. Vsi atomi in vezi v molekuli so označeni s točno določenimi simboli. Okolje vsakega atoma v molekuli je možno opisati s SMARTS jezikom. Potek reakcije je opisan tako, da so opisi reaktantov ločeni od opisov produkta z znakom » > «. Gre za zelo natančne opise, tako da lahko program za novo molekulo primerja njene SMARTS kode z SMARTS kodami v bazi podatkov in ugotavlja kakšna je podobnost med simboli. Glede na podobnosti in ujemanje simbolov določi reakcijski center v molekuli. (21)

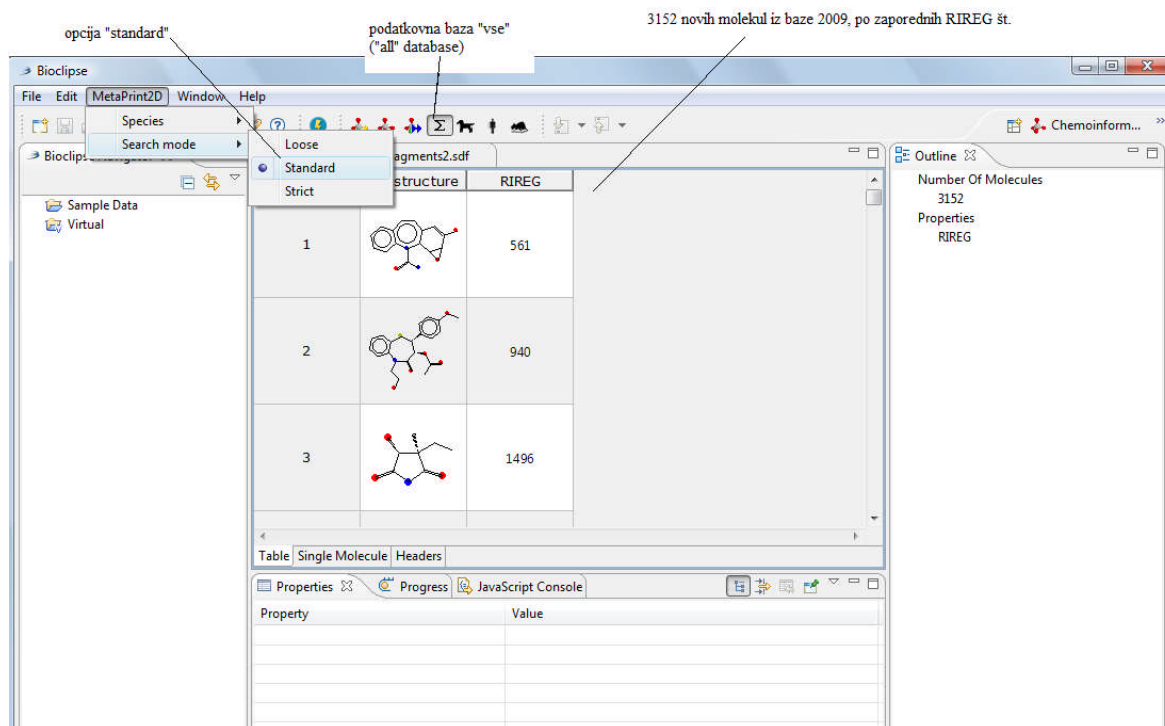
3.3 NAPOVEDOVANJE METABOLIZMA

S pomočjo seta vseh 150 spojin, reakcij I. faze, smo testirali napovedovanje metabolizma z uporabo obeh verzij programa. Spojine pri obeh verzijah smo testirali dvakrat. Prvič pri opciji »standard«, nato še pri opciji »loose«.

Pri uporabi prve verzije programa smo sledili naslednjim korakom:

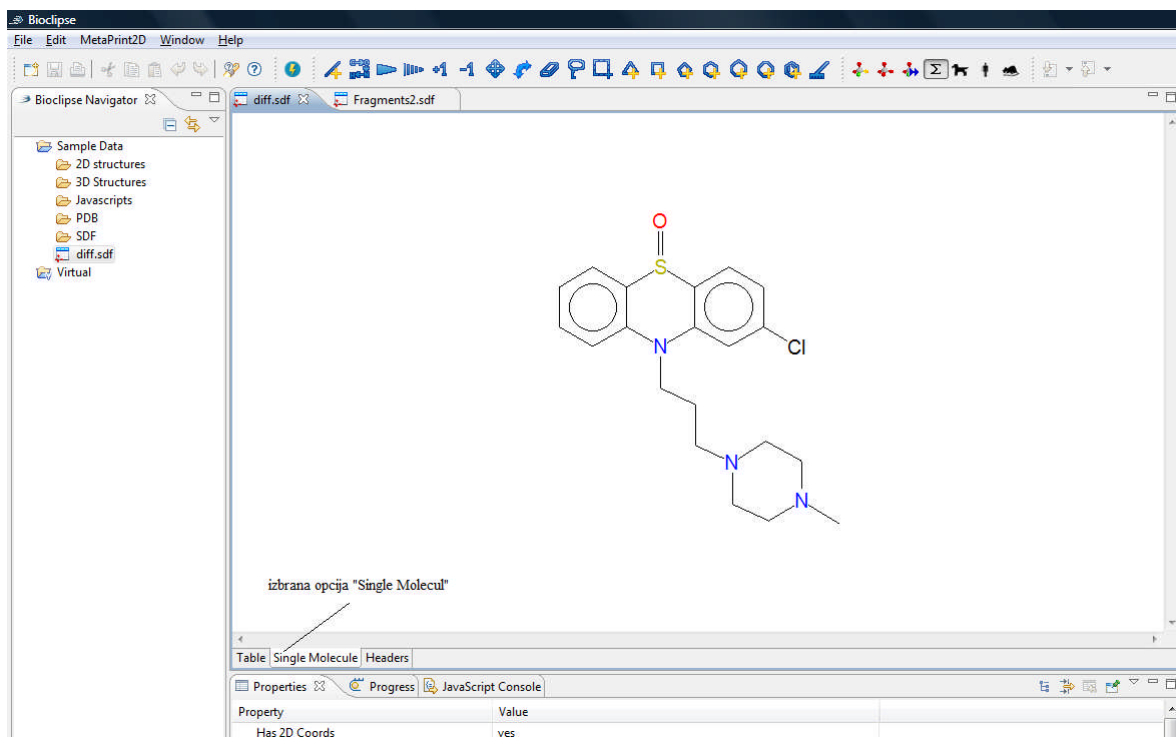
- v orodni vrstici izberemo ikono »MetaPrint2D«
- med ponujenimi možnostmi izberemo »search mode«

- izberemo opcijo »standard«
- nato v orodni vrstici poiščemo ikono »all database« tako da bo program operiral z vsemi podatkovnimi bazami, ki jih ima na voljo (slika 10)



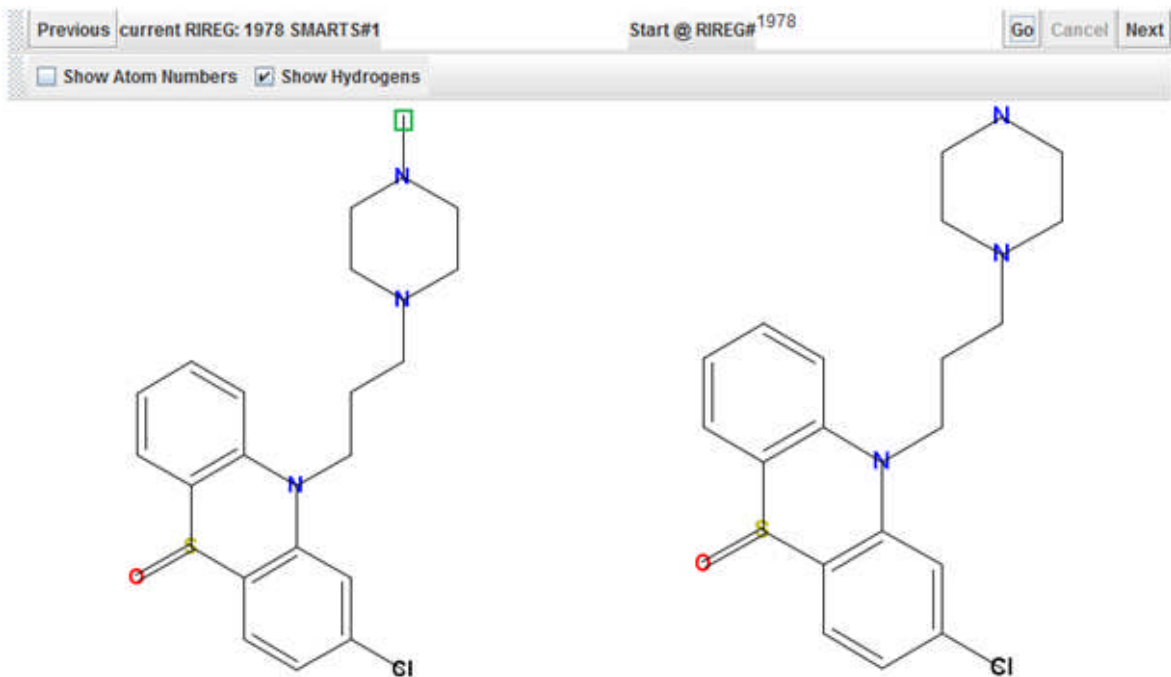
Slika 10: Izbor nastavitev prve verzije programa pred pričetkom testiranja.

Ko imamo zelene opcije izbrane, v podatkovni bazi (izmed 3152 novih molekul, ki so v bazi 2009, a jih ni v bazi 2005) po zaporedni RIREG številki poiščemo spojino, ki jo želimo, da ji program določi mesta metabolizma. S pritiskom na molekulo se vrstica obarva, kar pomeni da je zelena molekula izbrana, nato pritisnemo ikono »Single Molecule«. Na zaslonu se prikaže le molekula, s katero bomo operirali (slika 11).



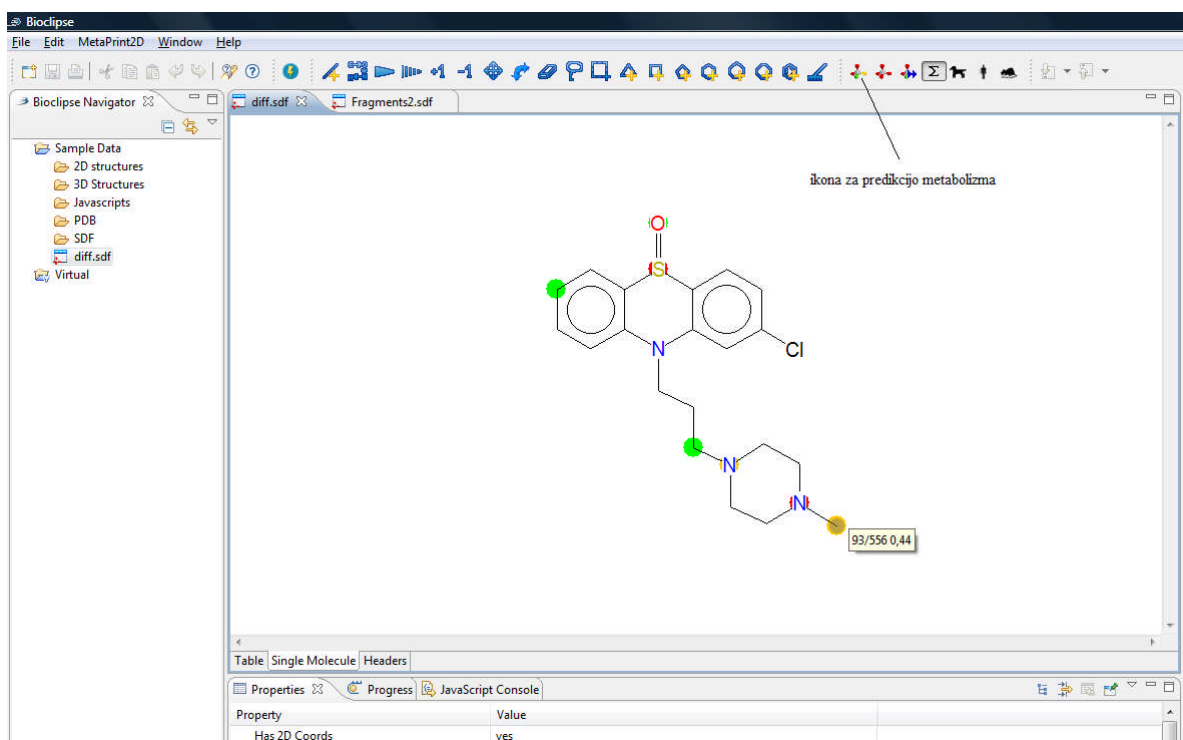
Slika 11: Prikaz posamezne molekule v prvi verziji programa, pred napovedjo metabolizma.

V naslednjem koraku damo programu ukaz, da izvede napoved metabolizma, in atomi v molekuli se obarvajo glede na to, kakšna je verjetnost da bo na tem mestu potekla reakcija metabolizma. Ko se s tabulatorjem pomaknemo na določen atom, vidimo tudi, kakšno je normalizirano razmerje (»normalized occurrence ratio«). V naslednjem koraku preverimo kakšna je napoved metabolizma na atomu, ki smo ga določili kot mesto metabolizma (reakcijski center) na podlagi eksperimentalnih podatkov z uporabo aplikacije »Reactant – Product – MCS«. Primer molekule in atom, na katerem poteče reakcija je označen z zelenim kvadratom in je prikazan na sliki 12. Sedaj ta isti atom poiščemo na molekuli v programu MP2D in pogledamo kako je program napovedal verjetnost metabolične reakcije na tem atomu (slika 13). Na ta način vidimo obarvanost iskanega atoma in verjetnost, da na tem atomu poteče reakcija. Na podlagi tega ocenimo katero mesto program MP2D napove eksperimentalno določenemu reakcijskemu centru. Pri tem se orientiramo po razmerjih vseh atomov (razmerja so preračunana po enačbi 3), in sicer preštejemo atome, ki imajo višje preračunano razmerje od našega atoma, naš atom tako zaseda naslednje mesto (če imata 2 atoma višje razmerje, pripišemo našemu atomu tretje mesto).



- I. faza
- N-demetiliranje
- atom C25

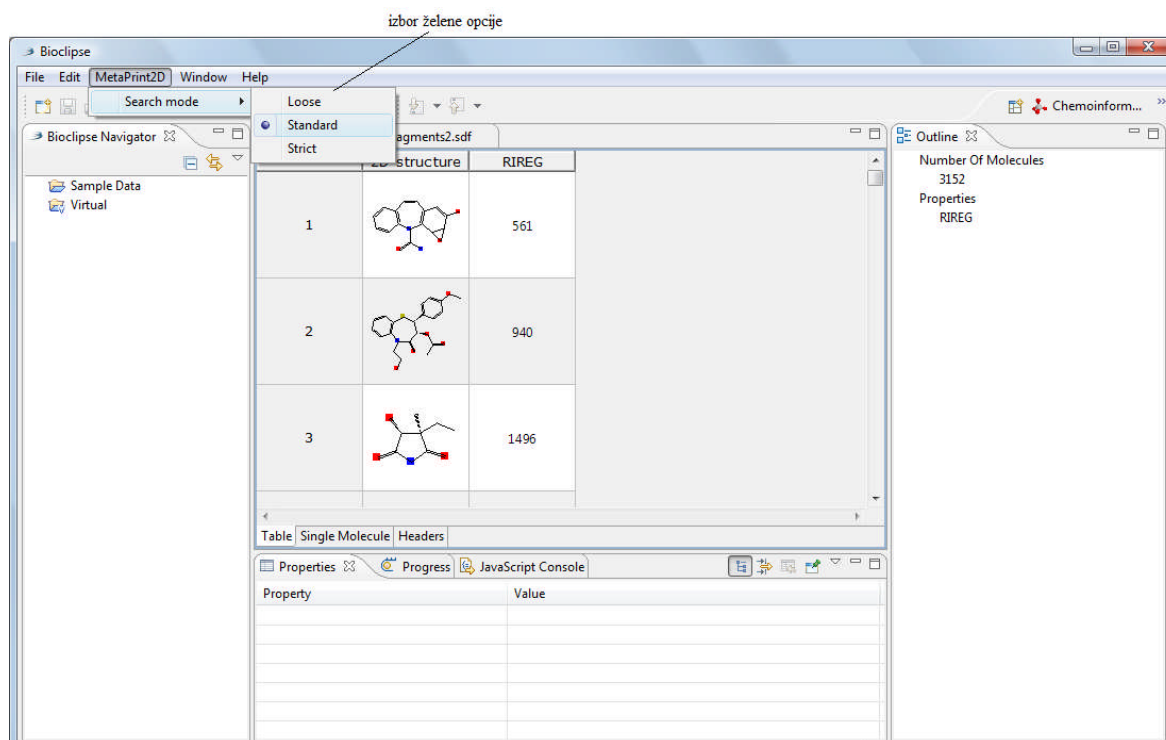
Slika 12: Atom, ki smo ga označili kot mesto metabolizma, ima zeleno oznako.



Slika 13: Molekula v prvi verziji programa po napovedi metabolizma. Atom, ki smo ga mi predhodno določili kot mesto metabolizma je obarvan oranžno in ima razmerje 0,44.

Na ta način smo testirali vseh 150 molekul. Nato smo enak postopek še enkrat ponovili, z eno razliko, da smo v ikoni »MetaPrint2D«, »search mode«, izbrali opcijo »loose«. Izbira vseh ostalih opcij in potek dela so ostali enaki. Rezultate smo sprti vpisovali v pripravljene tabele, kar pa je podrobneje opisano v naslednjem poglavju.

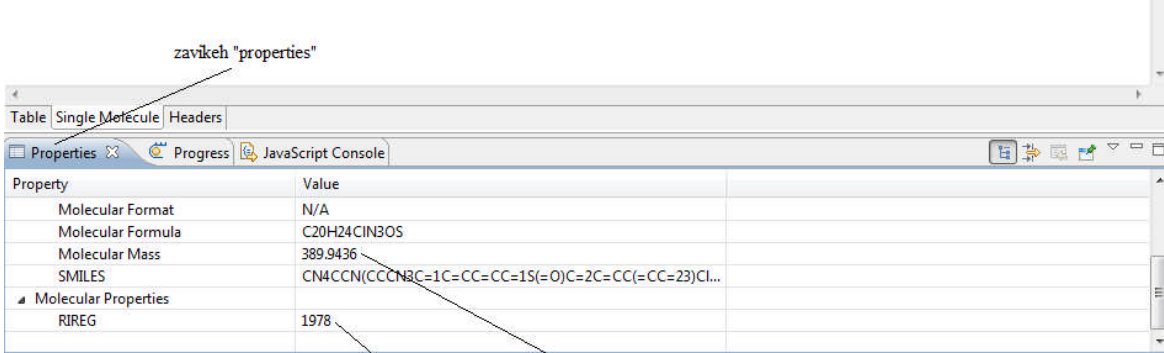
Testiranje z drugo verzijo programa je potekalo zelo podobno. Prav tako smo testirali vseh 150 molekul, vsako posamezno. Tudi tokrat smo vse molekule sprva testirali pod opcijo »standard«. V ikoni »MetaPrint2D« je tukaj na voljo le »search mode« in s pritiskom na to opcijo smo izmed vseh možnih opcij izbrali »standard« (slika 14). Baze podatkov tukaj nismo izbirali, saj program te možnosti še nima. Avtomatsko namreč izbira iz vseh podatkovnih baz, ki jih ima na voljo.



Slika 14: Izbora opcij »loose«, »standard« ali »strict« v drugi verziji programa.

V naslednjem koraku smo izbrali zeleno molekulo po zaporedni RIREG številki, podali ukaz za napoved metabolizma in pogledali, kako je atom, ki smo ga določili kot mesto metabolizma obarvan, kakšno razmerje ima in katero mesto metaboliziranja zaseda.

Program ima za vsako molekulo, poleg vseh ostalih podatkov o molekuli, zabeležen tudi podatek o njeni molekulski masi. Ta podatek se nahaja v spodnjem delu okenca, v zavihku »properties« (slika 15). Vzporedno z napovedjo metabolizma smo v tem programu zabeležili tudi molekulsko maso vsake molekule. Tako smo lahko dobljene rezultate po obdelavi rangirali tudi glede na molekulsko maso.



zavikeh "properties"

Property	Value
Molecular Format	N/A
Molecular Formula	C20H24ClN3OS
Molecular Mass	389.9436
SMILES	CN4CCN(CCCN3C=1C=CC=CC=1S(=O)C=2C=CC(=CC=23)Cl...
▲ Molecular Properties	
RIREG	1978

RIREG številka molekule molekulska masa

Slika 15: Podatek o molekulski masi molekule se nahaja v zavihku »properties«.

Kot zadnji korak testiranja smo vse molekule testirali še enkrat z drugo verzijo programa, tokrat z izbiro opcije »loose«. Tudi tokrat smo vse dobljene rezultate vpisovali v pripravljene tabele.

3.4 ANALIZA REZULTATOV TESTIRANJA

Ker je bil namen diplomske naloge ugotoviti izboljšave v napovedovanju mest metabolizma z obema verzijama programa MP2D, smo pridobili rezultate za vsak program posamezno, in jih nato primerjati med seboj. Pripravili smo štiri različne razporednice. Za vsako izmed opcij »standard« in »loose« po dve, eno za rezultate glede na barvo atoma, in eno za rezultate glede na to katero zaporedno mesto metaboliziranja atom zaseda. V pripravljene razporednice smo nato vpisovali rezultate.

Prvi stolpec vsake razporednice so zasedale RIREG številke, v naslednjem stolpcu je bil prostor za molekulske mase. Oznake naslednjih štirih stolpcev so bile odvisne od tega, ali smo gledali barvo atoma ali mesto metabolizma. Kadar smo ugotavljali barvo atoma, so bili stolpci označeni: rdeča, oranžna, zelena, ni oznake. Pri rezultatih glede na mesto metabolizma, pa so si oznake stolpcev sledile: 1. mesto metabolizma, mesto 2 ali 3, mesto 4 ali več, ni metabolizma. V šesti stolpec smo vpisovali tip reakcije. Pri rezultatih glede na

barvo atoma, je bil to tudi zadnji stolpec. Za rezultate glede na mesto metabolizma pa smo dodali še en stolpec, v katerega smo vpisovali kakšno je razmerje na atomu (tabela 1).

Tabela 1: Primer razpredelnice z rezultati glede na mesto metabolizma.

RIREG	Mol. masa	1. mesto metab.	Mesto 2 ali 3	Mesto 4 ali več	Ni metab.	Tip reakcije	Razmerje
940	387,4513		X			Oks.	0,27
1847	268,6979				X	Oks.	0,00
1978	389,9436		X			Oks.	0,84
1983	389,9436	X				Oks.	1,00
2541	322,4829			X		Oks.	0,02
5333	249,2664				X	Hidrol.	0,00
5334	236,2245				X	Hidrol.	0,00
6380	392,1291		X			Oks.	0,77
6392	392,1291	X				Oks.	1,00
6397	392,1291		X			Oks.	0,27
7353	358,472	X				Redukc.	1,00

Za vsako molekulo smo s križcem (X) označili ustrezen stolpec. Pri rezultatih glede na barvo atoma smo v stolpec *ni oznake* vpisali vse molekule, katerih atom ni bil obarvan. Torej tudi, če je bilo razmerje večje od 0,00, saj so brez oznake vsi atomi z razmerji manjšimi od 0,15. Za rezultate glede na mesto metabolizma, pa smo določili pravilo, da v stolpec *ni metabolizma* vpisujemo le tiste molekule, ki imajo na eksperimentalno določenem atomu oznako razmerja 0,00. Če je bilo razmerje večje od 0,00, a manjše od 0,15, in je bil atom na četrtem ali še višjem mestu po vrsti, smo ga označili kot *mesto 4 ali več*. Prav tako smo imeli primere, kjer sta imela dva ali več atomov enako razmerje. V tem primeru smo vse atome uvrstili na enako mesto. V kolikor pa smo določali mesto metabolizma za atom, ki je imel nižje razmerje, smo jih šteli kot vsak atom posebej. Tako smo na primer atom z razmerjem 0,95 uvrstili na tretje mesto, če sta imela dva atoma v molekuli razmerje 1,00.

Ko smo imeli zbrane vse rezultate v vseh razpredelnicah, smo jih obdelali tako, da smo prešteli križce za posamezne stolpce. Tako smo dobili rezultate koliko eksperimentalno določenim atomom (reakcijskim centrom) je pripadala oznaka rdeče, oranžno, zeleno, ali pa so bili brez oznake, ter koliko atomom je pripadalo prvo mesto metabolizma, koliko atomom mesto 2 ali 3, mesto 4 ali več, ali pa ni bilo metabolizma. Sešteli smo tudi atome,

ki so zasedali mesta od 1 do 3. Sedaj smo lahko primerjali rezultate pridobljene za prvo verzijo programa z rezultati druge verzije programa. Pomembno je, da primerjamo med seboj rezultate, ki so bili pridobljeni pod enakimi pogoji (med seboj primerjamo »standard« prve verzije programa, s »standard« druge verzije programa). Rezultate pridobljene pod različnimi pogoji, pa smo lahko primerjali znotraj programa. Zanimalo nas je torej, kako se razlikujejo rezultati pridobljeni s »standard« prve verzije programa, z rezultati »loose« prve verzije programa. Enako primerjavo smo naredili tudi za drugo verzijo programa (»standard« druga verzija programa v primerjavi z »loose« druge verzije programa).

Pri testiranju druge verzije programa smo sproti za vsako molekulo pripisali še njeno molekulsko maso. Z pridobljenimi podatki o molekulske maso smo nato dopolnili še ostale razpredelnice. Ker smo bolj podrobno želeli proučiti drugo verzijo programa, nas je zanimalo tudi, kakšen vpliv ima na rezultate molekulska masa (MW) molekule. Zato smo razdelili molekule v tri skupine:

- $MW \leq 300$
- $300 < MW < 500$
- $MW \geq 500$

Sedaj smo pogledali in sešteli, koliko molekul v posamezni skupini ima eksperimentalno določen atom obarvan rdeče, oranžno, zeleno, ali pa je brez oznake. Enako smo naredili tudi glede na mesto metabolizma – koliko molekul ima eksperimentalno določen atom na prvem napovedanem mestu, koliko na mestu 2 ali 3, mestu 4 ali več, ter koliko eksperimentalno določenih atomov nima napovedi metabolizma. Pregledali smo rezultate tako za opcijo »standard«, kot tudi za opcijo »loose«. Ker pa število molekul med posameznimi skupinami ni bilo enako, smo morali dobljene rezultate (seštevek molekul) pretvoriti v odstotke, da smo dobili primerljive rezultate.

Podobno smo pridobili rezultate tudi glede na tip reakcije. Tudi te rezultate smo preučili samo za drugo verzijo programa. Metabolične reakcije smo razdelili v štiri skupine:

- oksidacije
- redukcije
- hidrolize
- ostalo

Tudi tukaj smo sešteli molekule znotraj skupine, ki imajo eksperimentalno določen atom označen rdeče, oranžno, zeleno, ter koliko molekul atoma nima označenega. Pri rezultatih glede na mesto metabolizma, tudi enako kot prej, znotraj skupine seštejemo molekule, ki imajo eksperimentalno določen atom na 1. napovedanem mestu metabolizma, na mestu 2 ali 3, mestu 4 ali več, ter nimajo napovedi metabolizma. Dobljene rezultate smo nato preračunali v odstotke.

4 REZULTATI

Rezultate napovedi metabolizma molekul smo pridobili glede na mesto metabolizma v molekulah. Primerjali smo rezultate dobljene z opcijo »standard« z rezultati opcije »loose« znotraj vsake verzije programa, prav tako smo primerjali tudi rezultate pridobljene z drugo verzijo programa z rezultati prve verzije programa. V drugi verziji programa smo pridobili tudi rezultate vpliva molekulske mase in tipa reakcije na napovedovanje mest metabolizma v molekulah. Rezultate smo predstavili v tabelah in grafično. Rezultati predstavljajo seštevek molekul glede na mesto metabolizma v molekuli in seštevek molekul glede na barvo mesta metabolizma v molekuli. Pri prikazu molekul uvrščenih na prva tri mesta so rezultati predstavljeni v odstotkih. Prav tako so pri molekulske masi in tipu reakcij vsi rezultati v odstotkih.

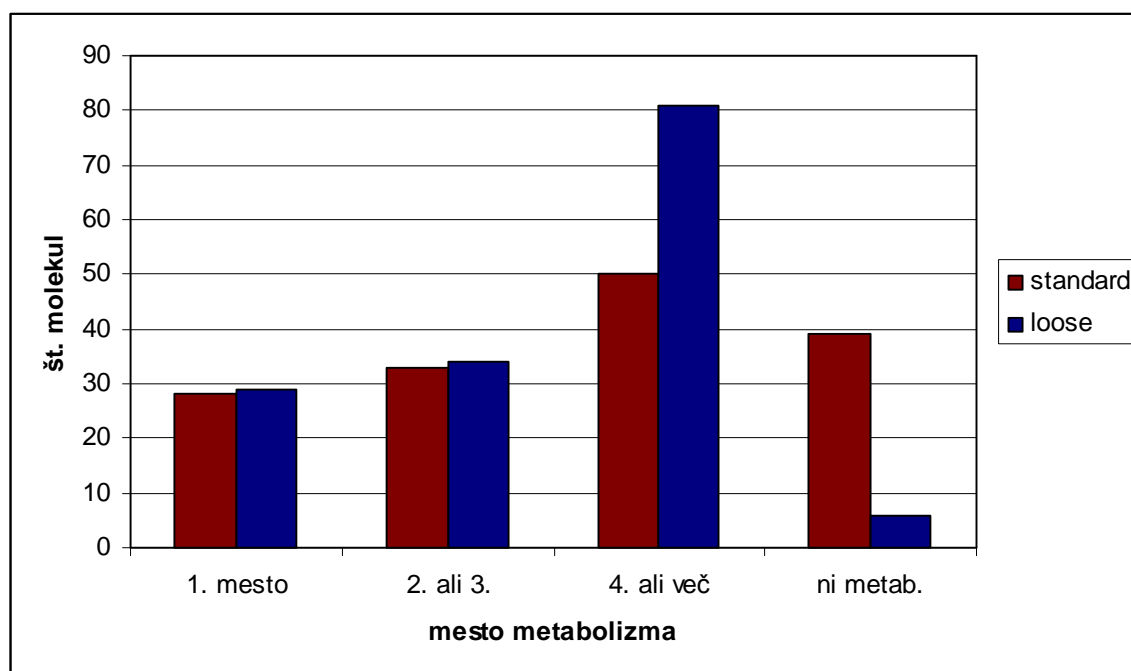
4.1 REZULTATI GLEDE NA MESTO METABOLIZMA V MOLEKULAH

Glede na verjetnost poteka metabolične reakcije na eksperimentalno določenem atomu (reakcijskem centru) smo molekule uvrstili na 1. mesto (mesto napovedi je eksperimentalno določeno mesto metabolizma, preračunano normalizirano razmerje na reakcijskem centru je 1,00), 2. ali 3. mesto, 4. mesto ali več, ter ni metabolizma za vsa preračunana razmerja 0,00 na reakcijskem centru. Rezultati so prikazani za prvo verzijo programa, drugo verzijo programa, ter kot primerjava med programoma.

4.1.1 Prva verzija programa

Tabela 2: Rezultati prve verzije programa za opciji »standard« in »loose« in so prikazani kot vsota molekul za določeno mesto, ter kot odstotek molekul na prvih treh mestih.

	1. mesto	2. ali 3.	4. ali več	ni metab.	1.-3. mesto
standard	28	33	50	39	40,67%
loose	29	34	81	6	42,00%

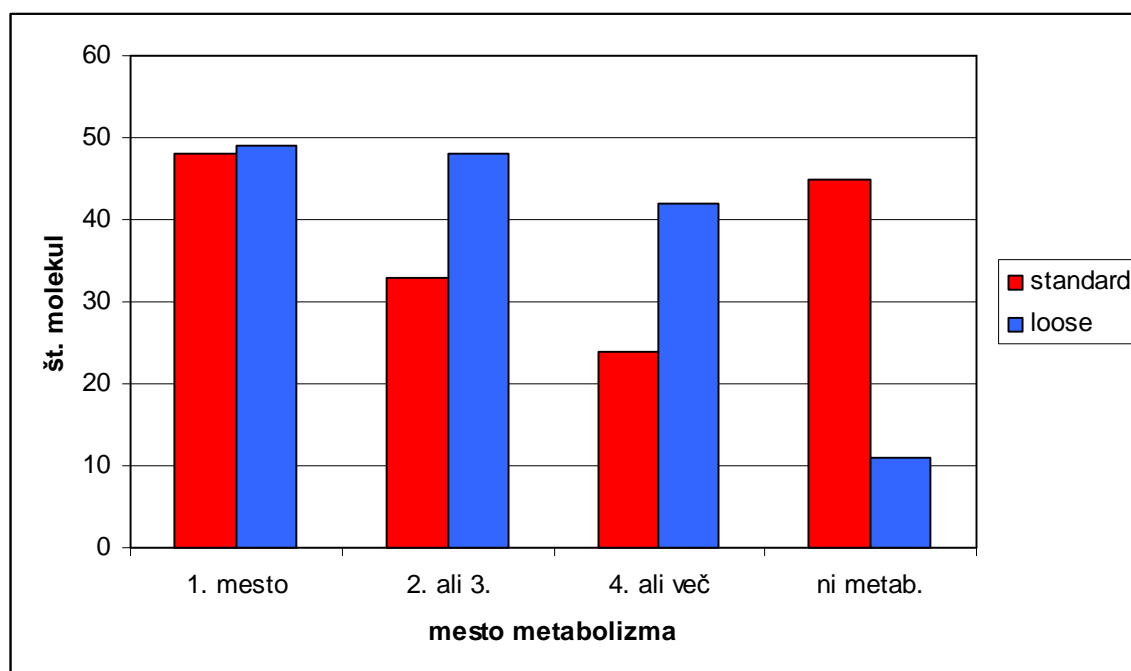


Graf 1: Primerjava opcije »standard« z opcijo »loose« za prvo verzijo programa.

4.1.2 Druga verzija programa

Tabela 3: Rezultati druge verzije programa za opciji »standard« in »loose« in so prikazani kot vsota molekul za določeno mesto, ter kot odstotek molekul na prvih treh mestih.

	1. mesto	2. ali 3.	4. ali več	ni metab.	1.-3. mesto
standard	48	33	24	45	54,00%
loose	49	48	42	11	64,67%

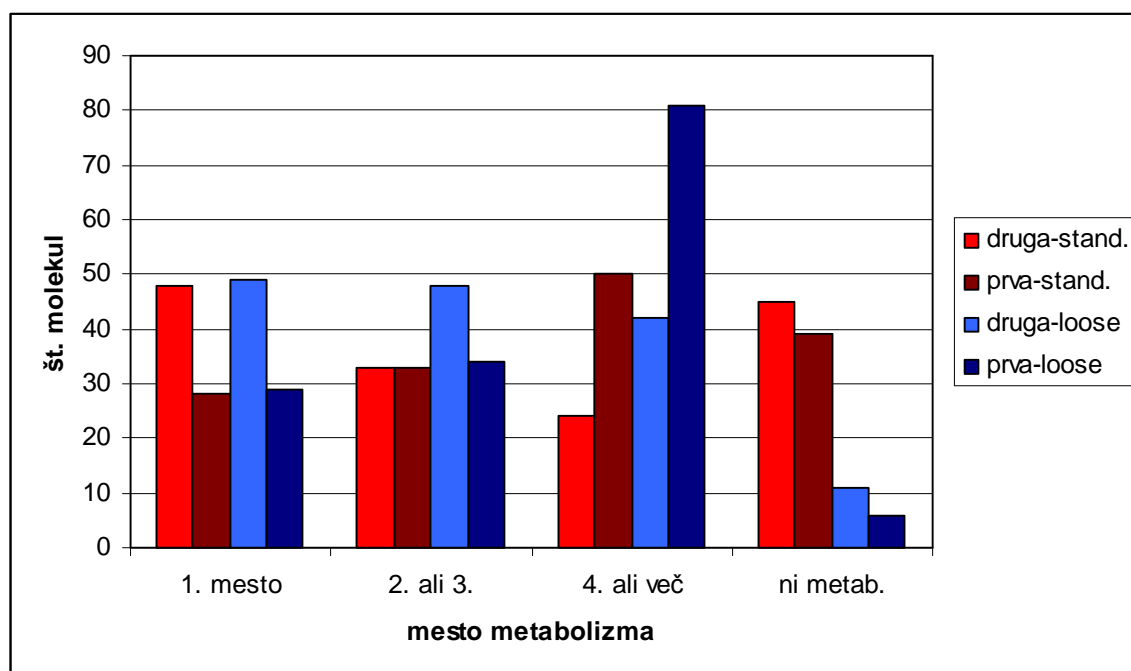


Graf 2: Primerjava opcije »standard« z opcijo »loose« za drugo verzijo programa.

4.1.3 Primerjava druge verzije programa s prvo

Tabela 4: Primerjava rezultatov druge verzije programa s prvo pri opciji »standard« in opciji »loose«.

	1. mesto	2. ali 3.	4. ali več	ni metab.	1.-3. mesto
druga-stand.	48	33	24	45	54,00%
prva-stand.	28	33	50	39	40,67%
druga-loose	49	48	42	11	64,67%
prva-loose	29	34	81	6	42,00%



Graf 3: Primerjava druge verzije programa s prvo glede na mesto metabolizma pri opciji »standard« in opciji »loose«.

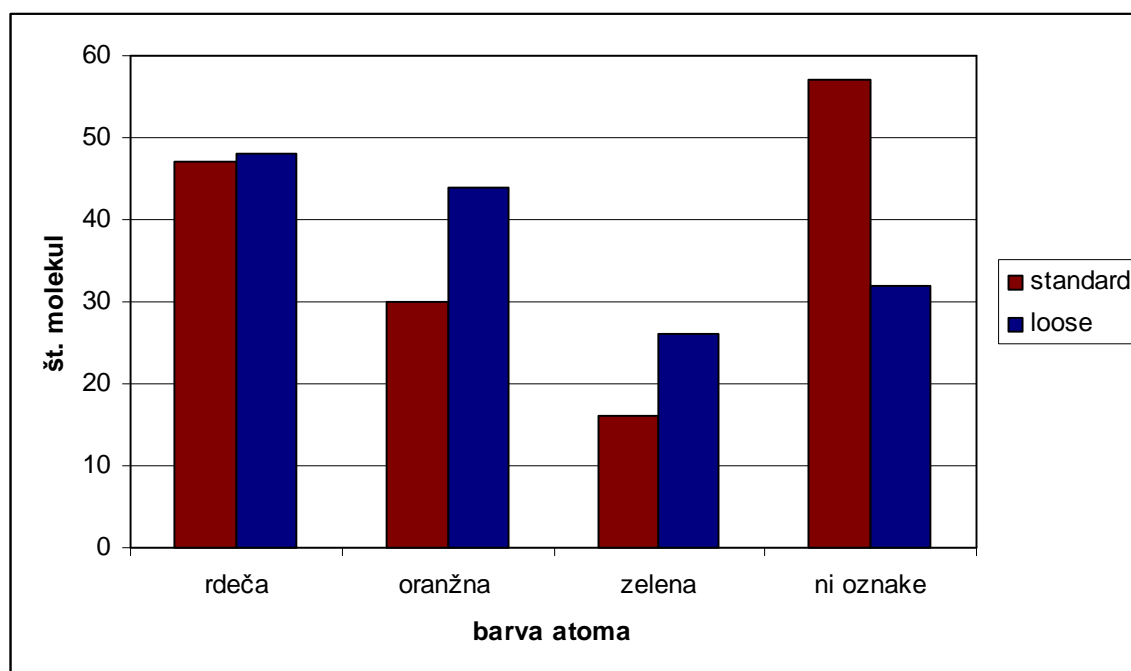
4.2 REZULTATI GLEDE NA BARVO ATOMA

Na eksperimentalno določenem atomu smo pogledali barvno oznako, in molekulo glede na to oznako uvrstili med rdeča, oranžna, zelena ali ni oznake. Rezultati so prikazani za prvo verzijo programa, drugo verzijo programa, ter kot primerjava med verzijama programa, in predstavljajo seštevek molekul za posamezno barvo.

4.2.1 Prva verzija programa

Tabela 5: Rezultati prve verzije programa za opciji »standard« in »loose« glede na barvo eksperimentalno določenega reakcijskega centra v molekuli.

	rdeča	oranžna	zelena	ni oznake
standard	47	30	16	57
loose	48	44	26	32

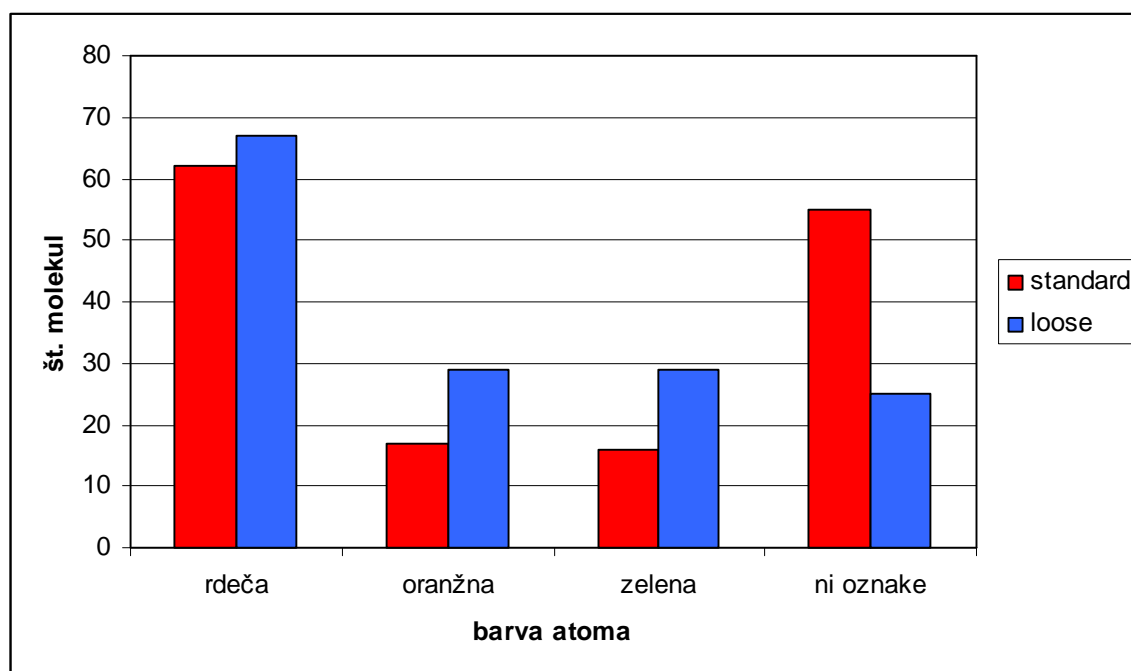


Graf 4: Rezultati tabele 5 predstavljeni v grafični obliki. Primerjava opcije »standard« z opcijo »loose« za prvo verzijo programa.

4.2.2 Druga verzija programa

Tabela 6: Rezultati druge verzije programa za opciji »standard« in »loose« glede na barvo eksperimentalno določenega reakcijskega centra v molekuli.

	rdeča	oranžna	zelena	ni oznake
standard	62	17	16	55
loose	67	29	29	25

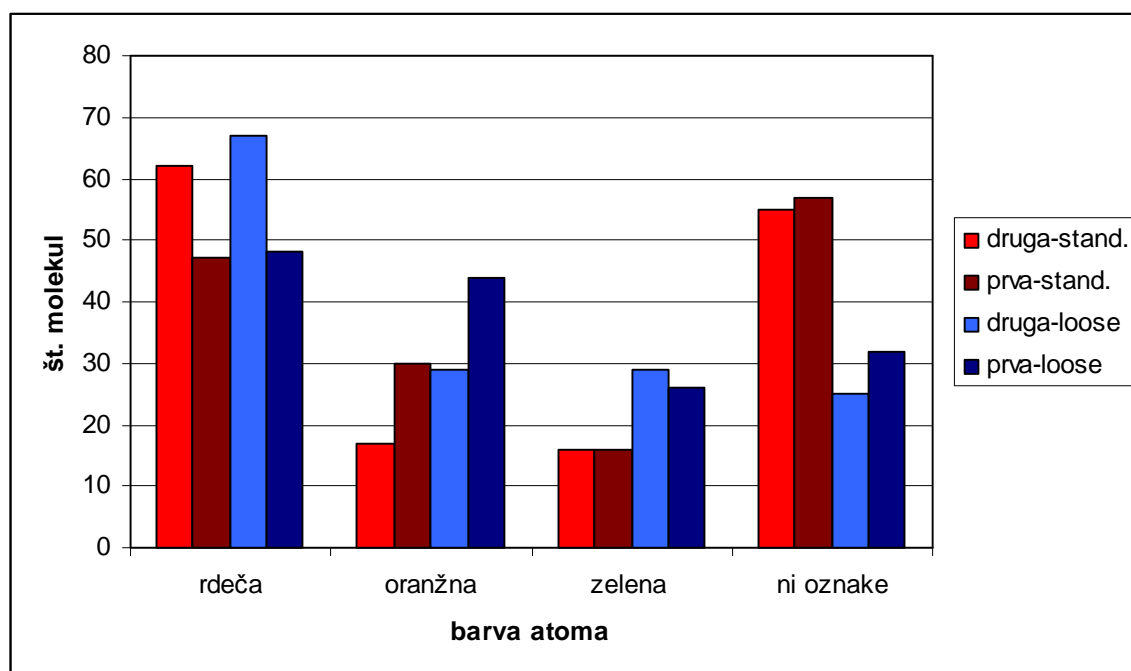


Graf 5: Rezultati tabele 6 predstavljeni v grafični obliki. Primerjava opcije »standard« z opcijo »loose« za drugo verzijo programa.

4.2.3 Primerjava druge verzije programa s prvo

Tabela 7: Primerjava druge verzije programa s prvo pri opciji »standard« in »loose« glede na barvo eksperimentalno določenega reakcijskega centra v molekuli.

	rdeča	oranžna	zelena	ni oznake
druga-stand.	62	17	16	55
prva-stand.	47	30	16	57
druga-loose	67	29	29	25
prva-loose	48	44	26	32



Graf 6: Primerjava druge verzije programa s prvo glede na barvo eksperimentalno določenega reakcijskega centra.

4.3 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA TIP REAKCIJE

Molekule smo razdelili glede na tip reakcije med oksidacije, redukcije, hidrolize in ostale reakcije. Največ molekul je pripadalo reakcijam oksidacije – 82 molekul, najmanj pa reakcijam redukcije – 9 molekul. 39 molekul se hidrolizira, 20 molekul pa spada med ostale reakcije (reakcije aromatizacija, *N*- ali *O*-deglikozidacije, defosfatacije,...). Znotraj vsakega tipa reakcij smo jih nato razvrstili glede na mesto metabolizma in glede na barvo reakcijskega centra. Obe raziskavi smo opravili pri opciji »standard« in nato še pri opciji »loose«. Dobljeno število molekul za posamezno mesto in barvo atoma smo pretvorili v odstotke.

4.3.1 Glede na mesto napovedi metabolizma

- OPCIJA »STANDARD«

Tabela 8: Odstotek molekul znotraj določenega tipa reakcije, glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »standard«.

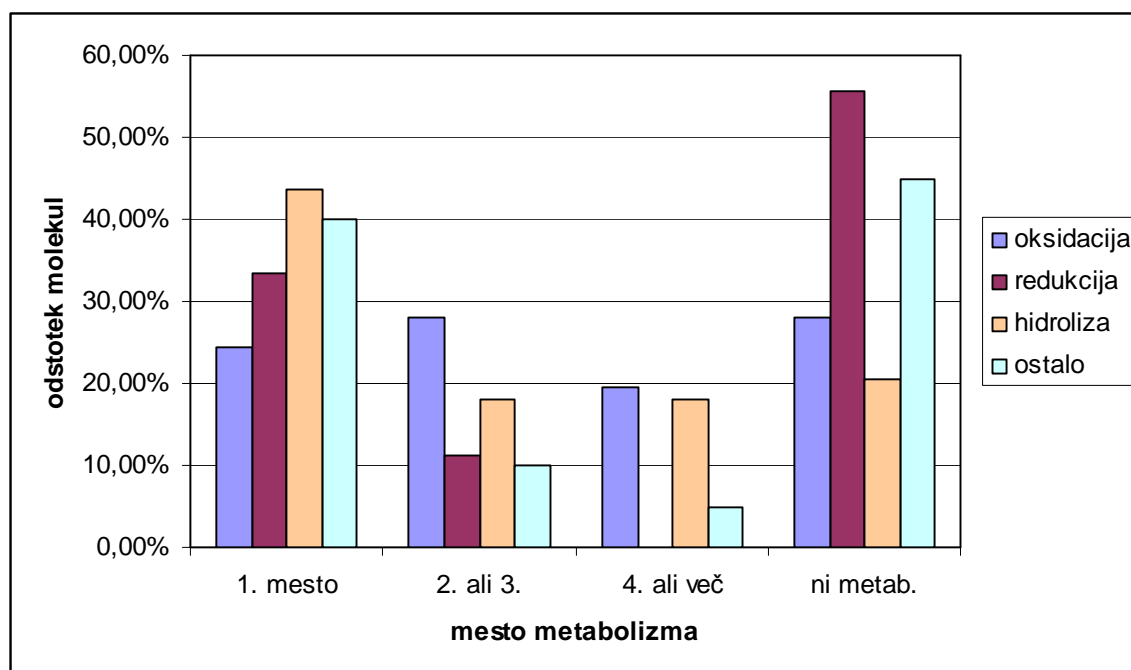
tip reakcije	1. mesto	2. ali 3.	4. ali več	ni metab.	1.-3. mesto
oksidacija	24,39 %	28,05 %	19,51 %	28,05 %	52,44%

redukcija	33,33 %	11,11 %	0,00 %	55,56 %	44,44%
hidroliza	43,59 %	17,95 %	17,95 %	20,51 %	61,54%
ostalo	40,00 %	10,00 %	5,00 %	45,00 %	50,00%

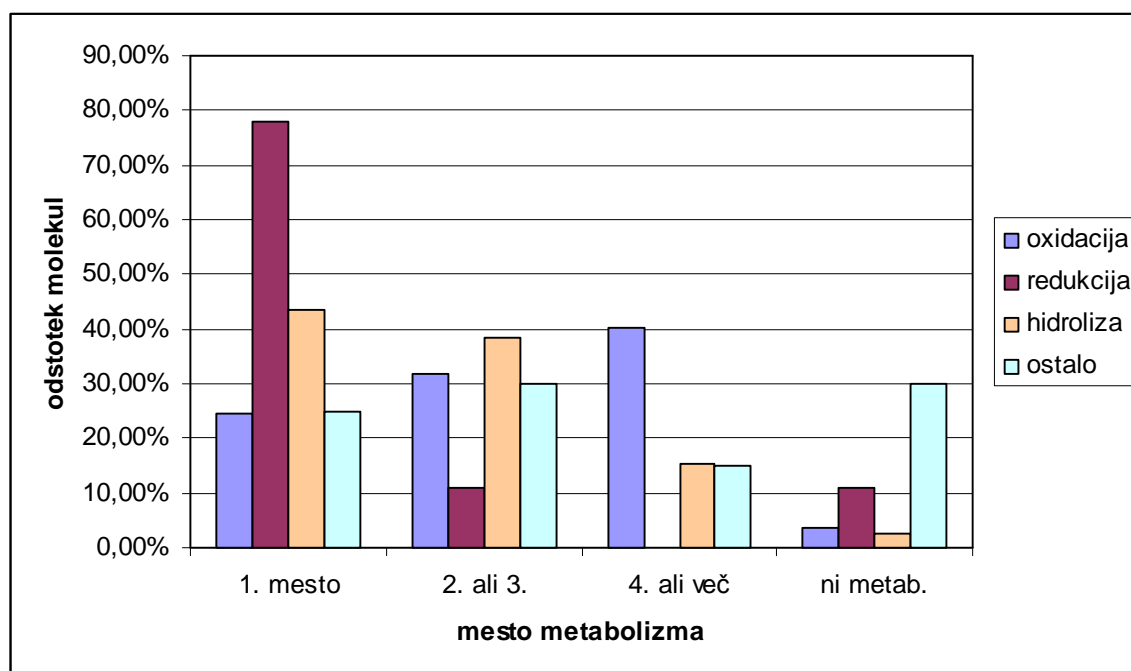
- OPCIJA »LOOSE«

Tabela 9: Odstotek molekul znotraj določenega tipa reakcije, glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

tip reakcije	1. mesto	2. ali 3.	4. ali več	ni metab.	1.-3. mesto
oksidacija	24,39 %	31,71 %	40,24 %	3,66 %	56,10%
redukcija	77,78 %	11,11 %	0,00 %	11,11 %	88,89%
hidroliza	43,59 %	38,46 %	15,38 %	2,56 %	82,05%
ostalo	25,00 %	30,00 %	15,00 %	30,00 %	55,00%



Graf 7: Grafičen prikaz odstotka molekul znotraj določenega tipa reakcije, glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »standard«.



Graf 8: Grafičen prikaz odstotka molekul znotraj določenega tipa reakcije, glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

4.3.2 Glede na barvo atoma

- OPCIJA »STANDARD«

Tabela 10: Odstotek molekul znotraj določenega tipa reakcije, glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »standard«.

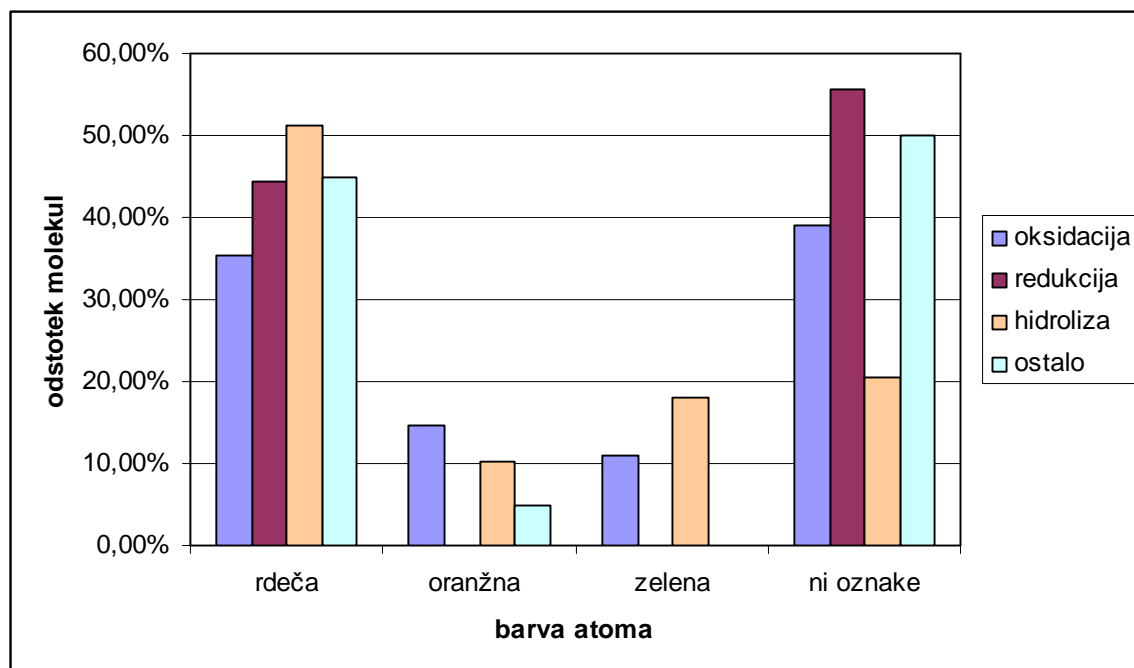
tip reakcije	rdeča	oranžna	zelena	ni oznake
oksidacija	35,37 %	14,63 %	10,98 %	39,02 %
redukcija	44,44 %	0,00 %	0,00 %	55,56 %
hidroliza	51,28 %	10,26 %	17,95 %	20,51 %
ostalo	45,00 %	5,00 %	0,00 %	50,00 %

- OPCIJA »LOOSE«

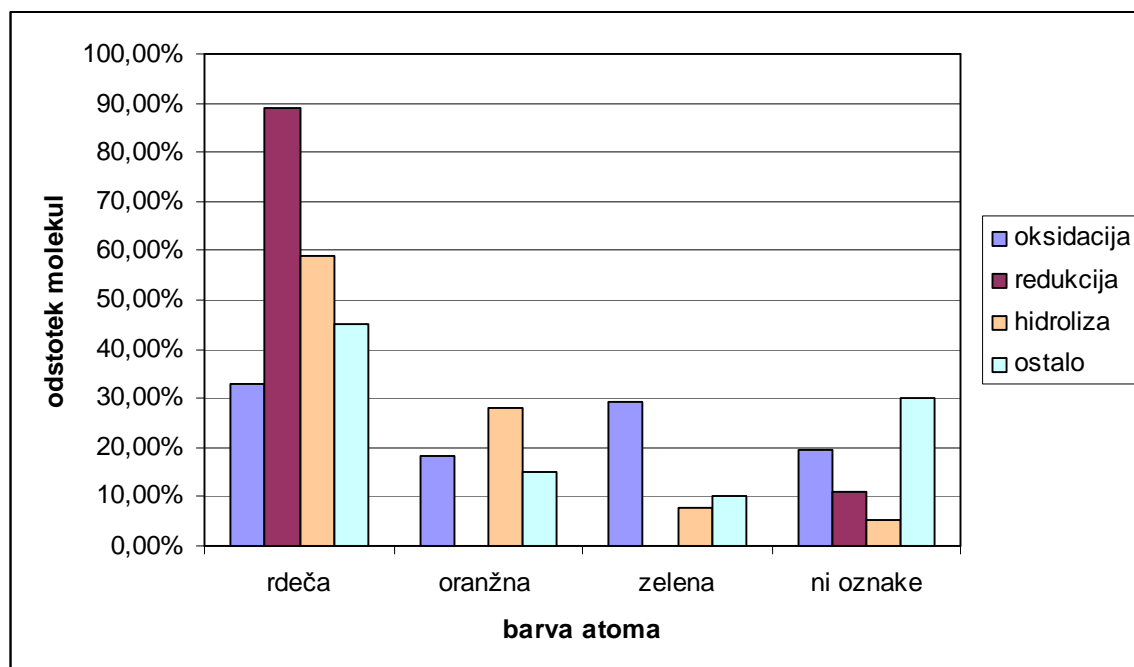
Tabela 11: Odstotek molekul znotraj določenega tipa reakcije, glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

tip reakcije	rdeča	oranžna	zelena	ni oznake
oksidacija	32,93 %	18,29 %	29,27 %	19,51 %
redukcija	88,89 %	0,00 %	0,00 %	11,11 %

hidroliza	58,97 %	28,21 %	7,69 %	5,13 %
ostalo	45,00 %	15,00 %	10,00 %	30,00 %



Graf 9: Grafičen prikaz odstotka molekul znotraj določenega tipa reakcije glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »standard«.



Graf 10: Grafičen prikaz odstotka molekul znotraj določenega tipa reakcije, glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

4.4 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA MOLEKULSKO MASO

Molekule smo razdelili v tri skupine, glede na molekulske mase. V prvi skupini so bile molekule z molekulsko maso do 300 g/mol, v kateri je bilo 37 molekul. V naslednjo skupino smo uvrstili molekule z molekulsko maso 300 – 500 g/mol, ki je bila tudi največja skupina z 83 molekulami. V zadnjo skupino so bile uvrščene molekule z molekulsko maso višjo od 500 g/mol, teh molekul je bilo 30. Znotraj vsake skupine smo jih nato uvrstili glede na mesto metabolizma in glede na barvo eksperimentalno določenega reakcijskega centra. Obe raziskavi smo opravili pri opciji »standard« in nato še pri opciji »loose«, dobljene rezultate pa smo pretvorili v odstotke.

4.4.1 Glede na mesto napovedi metabolizma

- OPCIJA »STANDARD«

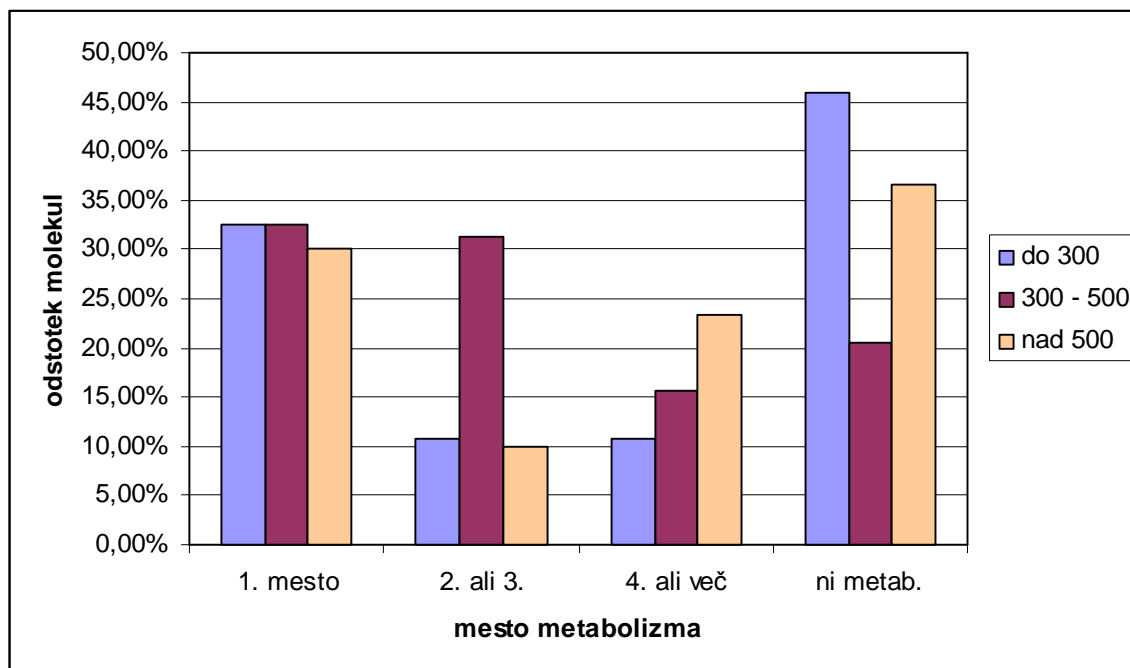
Tabela 12: Odstotek molekul znotraj posameznega intervala molekulske mase, glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »standard«.

mol. masa	1. mesto	2. ali 3.	4. ali več	ni metab.	1.-3. mesto
do 300	32,43 %	10,81 %	10,81 %	45,95 %	43,24%
300 - 500	32,53 %	31,33 %	15,66 %	20,48 %	63,86%
nad 500	30,00 %	10,00 %	23,33 %	36,67 %	40,00%

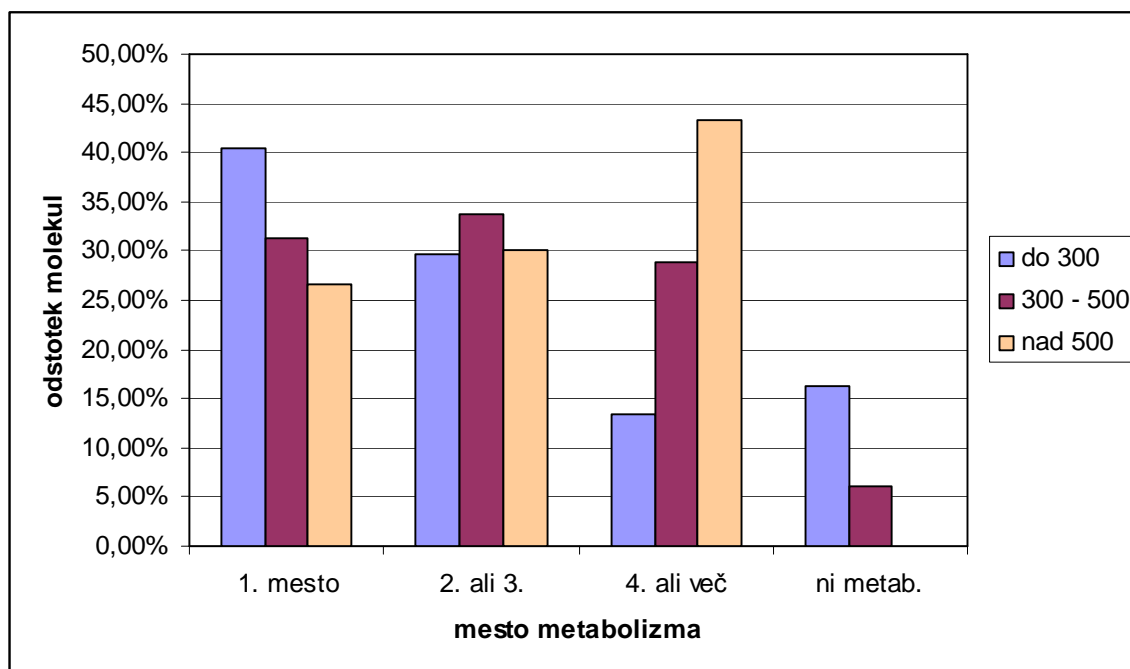
- OPCIJA »LOOSE«

Tabela 13: Odstotek molekul znotraj posameznega intervala molekulske mase, glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

mol. masa	1. mesto	2. ali 3.	4. ali več	ni metab.	1.-3. mesto
do 300	40,54 %	29,73 %	13,51 %	16,22 %	70,27%
300 - 500	31,33 %	33,73 %	28,92 %	6,02 %	65,06%
nad 500	26,67 %	30,00 %	43,33 %	0,00 %	56,67%



Graf 11: Grafičen prikaz odstotka molekul znotraj posameznega intervala molekulske mase glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »standard«.



Graf 12: Grafičen prikaz odstotka molekul znotraj posameznega intervala molekulske mase glede na mesto metabolizma eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

4.4.2 Glede na barvo atoma

- OPCIJA »STANDARD«

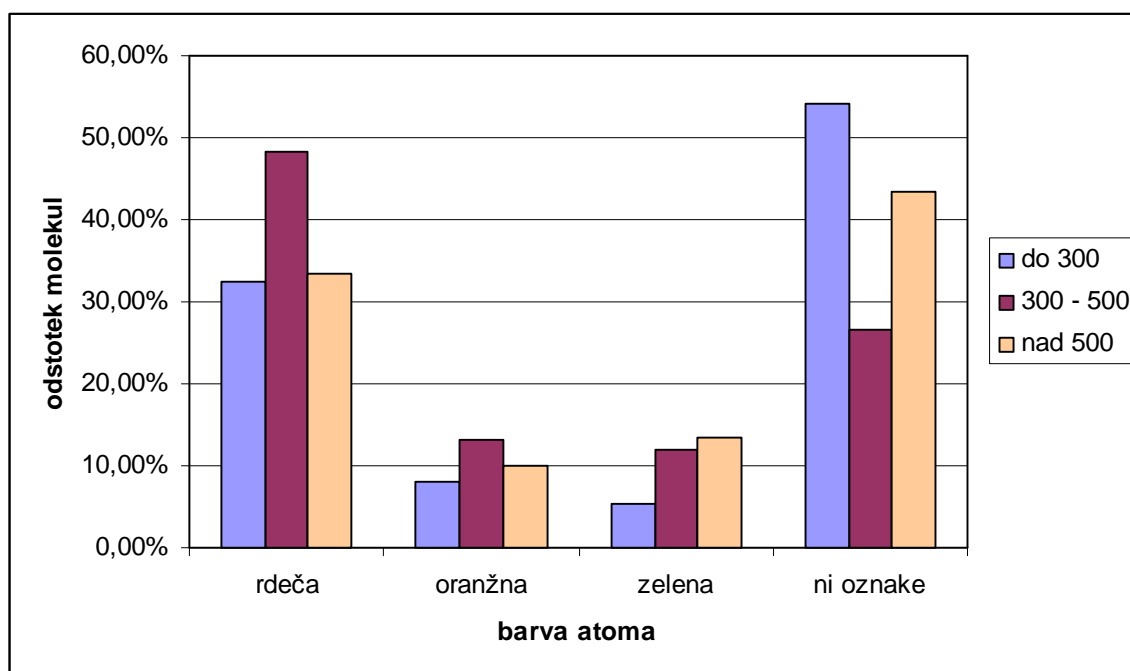
Tabela 14: Odstotek molekul znotraj posameznega intervala molekulske mase glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »standard«.

mol. masa	rdeča	oranžna	zelena	ni oznake
do 300	32,43 %	8,12 %	5,41 %	54,05 %
300 - 500	48,19 %	13,25 %	12,05 %	26,51 %
nad 500	33,33 %	10,00 %	13,33 %	43,33 %

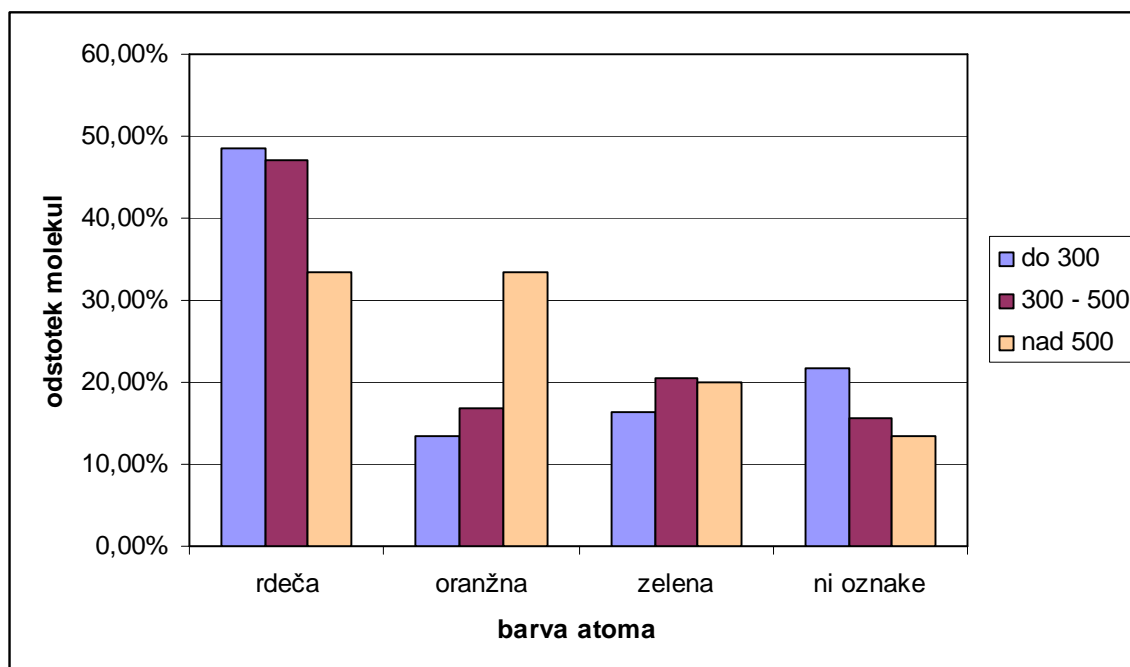
- OPCIJA »LOOSE«

Tabela 15: Odstotek molekul znotraj posameznega intervala molekulske mase glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

mol. masa	rdeča	oranžna	zelena	ni oznake
do 300	48,65 %	13,51 %	16,22 %	21,62 %
300 - 500	46,99 %	16,87 %	20,48 %	15,66 %
nad 500	33,33%	33,33 %	20,00 %	13,33 %



Graf 13: Grafičen prikaz odstotka molekul znotraj posameznega intervala molekulske mase, glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »standard«.



Graf 14: Grafičen prikaz odstotka molekul znotraj posameznega intervala molekulske mase, glede na barvo eksperimentalno določenega reakcijskega centra, pri opciji »loose«.

5 RAZPRAVA

MP2D je »data mining« metoda za napovedovanje mest metabolizma v molekulah, ki temelji na SPORCalc (substrate product occurrence ratio calculator) pristopu. (3) SPORCalc pristop za svoje izračune uporablja prstne odtise atomov substrata, ki so objavljeni v podatkovni bazi Metabolite[®]. (16) Program je zasnovan tako, da je njegova uporaba popolnoma avtomatizirana in omogoča združevanje z drugimi programskimi opremami. Prav tako je javno dostopen in enostaven za uporabo. (3)

5.1 REZULTATI GLEDE NA MESTO METABOLIZMA V MOLEKULAH

Izbrane molekule (150 molekul) smo v prvi fazi testirali s prvo verzijo programa. Predpostavljali smo, da izbor opcij »standard« ali »loose« vpliva na dobljene rezultate, zato smo želeli dano hipotezo preveriti. Pri izbiri opcije »loose« je okolje manj natančno definirano kot pri izbiri opcije »standard«. Pomembno je ujemanje med matricami prstnih odtisov, ki mora biti pri opciji »standard« večje kakor pri opciji »loose«, kjer je ujemanje med matricami najmanjše.

Pri izbiri opcije »standard« se je na *1. mesto* uvrstilo 28 molekul, pri opciji »loose« pa le ena molekula več, torej 29 molekul. Majhna razlika je tudi na *mestih 2 ali 3*, kamor je pri opciji »standard« uvrščenih 33 molekul, pri opciji »loose« pa 34 molekul. Iz dobljenih rezultatov je razvidno, da ima izbira opcij le majhen vpliv na rezultate prvih treh mest metabolizma v molekuli. Preračunano v odstotke, je bilo na prvih treh mestih pri opciji »standard« 40,67% vseh preiskovanih molekul, pri opciji »loose« pa le za dober odstotek več, 42,00%. Večja razlika je opazna na *mestu 4 ali več*, ter na mestu kjer *ni metabolizma* (razmerje na atomu 0,00). Opcija »loose« uvršča največ molekul (81 molekul) na *mestu 4 ali več*, ter najmanj na mestu *ni metabolizma*, 6 molekul. Medtem ko je pri opciji »standard« na mestu *ni metabolizma* uvrščenih največ molekul, kar 39 molekul. Iz tega je razvidno, da ima izbor opcij in s tem okolica atoma, za napovedovanje mest metabolizma v molekulah pomembno vlogo. Kadar je izbrana opcija »loose«, in s tem okolica atoma ni tako pomembna, ampak je pomemben le sam atom in nanj neposredno vezane skupine, program skoraj na vsakem atomu preračuna vsaj minimalno vrednost razmerja in s tem nakaže na neko verjetnost (čeprav zelo majhno), da na atomu poteče metabolna reakcija. Posledično zaradi tega na mestu *ni metabolizma* prihaja do tako velikih razlik med obema opcijama.

Za oceno zanesljivosti programa smo se osredotočili na rezultate prvih treh mest. Pri obeh proučevanih opcijah je bil odstotek molekul na prvih treh mestih manjši od 50,00%. Iz tega je razvidno, da so za uporabo programa v praksi, v raziskovalne namene, potrebne izboljšave.

Z drugo verzijo programa je bilo pri opciji »standard« na *1. mesto* uvrščenih 48 molekul, pri opciji »loose« pa 49 molekul. Tudi s tem programom je bila največja razlika med opcijama opažena na mestu *ni metabolizma*, kamor je bilo z opcijo »standard« uvrščenih 45 molekul, z opcijo »loose« pa le 11 molekul. S tem smo ponovno dokazali, da je izbira opcij, in s tem okolica atoma, pomemben faktor pri napovedovanju mest metabolizma v molekuli. Bolj kot je okolica atoma definirana, več oddaljenih atomov in funkcionalnih skupin vpliva na preiskovani atom, (16) in ujemanje med matricami prstnih odtisov nove molekule in prstnimi odtisi molekul v bazi podatkov more biti večje. Posledično pri takšnih ukazih program najde manj identičnih prstnih odtisov v bazi podatkov in predpostavlja, da metabolizem na tem atomu ne poteka. Vendar pa izbira opcij tudi pri tej verziji programa nima pomembnega vpliva na rezultate *1. mesta* metabolizma. To je posledica tega, da na *1. mesto* uvrščamo le molekule, ki imajo na eksperimentalno določenem atomu razmerje 1,00 (enačba 3), torej kadar gre za popolno ujemanje med matricami prstnih odtisov. Tako imajo posledično vse molekule, z razmerjem 1,00 pri opciji »standard«, razmerje 1,00 tudi pri opciji »loose«, saj se tukaj primerja manj stopenj prstnih odtisov, in so avtomatsko uvrščene na *1. mesto*. Večji vpliv pa zaznamo že na mestih *2 ali 3*, kamor smo pri opciji »standard« uvrstili 33 molekul, pri opciji »loose« pa 48 molekul. Odstotek molekul na prvih treh mestih pri opciji »standard« znaša 54,00%, medtem ko je pri opciji »loose« dokaj višji, in sicer 64,67%. Pri obeh opcijah je odstotek višji od 50,00%, kar pomeni, da je program za več kot polovico molekul naredil dobro napoved metabolizma. Z izbiro opcije »loose« smo še nekoliko bližje pravih napovedim.

Druga verzija programa vsebuje drugačen način zapisa poteka reakcij v podatkovni bazi kot prva verzija. Vse reakcije I. faze so v drugi verziji programa definirane s SMARTS kodami. Okolje vsakega atoma v molekuli je možno opisati s SMARTS jezikom. To je velika pridobitev, zato smo želeli preveriti, ali izboljšave v drugi verziji programa vplivajo na večjo zanesljivost rezultatov. To smo storili tako, da smo dobljene rezultate med seboj primerjali.

Tabela 4 in graf 3 prikazujeta primerjavo obeh verzij programa pri obeh opcijah. Druga verzija programa pri opciji »standard« največ molekul uvršča na *1. mesto*, 48 molekul, medtem ko prva verzija programa na *1. mesto* uvršča le 28 molekul. Na mesti *2 ali 3* obe verziji uvrščata 33 molekul. Na prva tri mesta je tako z drugo verzijo programa uvrščenih 54,00% vseh molekul, s prvo verzijo programa pa 40,67%. Iz grafa 3 je razvidno, da je največja razlika med programoma pri opciji »standard« na mestu *4 ali več*. V prvi verziji programa je bilo na to mesto uvrščenih največ molekul, 50. Predpostavljamo, da je to posledica nekonsistentnosti pri določanju reakcijskih centrov v molekulah.

Do podobnih zaključkov smo prišli tudi pri primerjavi verzij programov z opcijo »loose«. Tudi tokrat je bilo s prvo verzijo programa največ molekul uvrščenih na mesto *4 ali več* (81 molekul), medtem ko je bilo z drugo verzijo programa največ molekul uvrščenih na *1. mesto* (49 molekul). Odstotek molekul, ki zasedajo prva tri mesta v drugi verziji programa, je 64,67%, medtem ko je v prvi verziji programa ta delež le 42,00%.

Da bi prišli do zaključka, ali daje optimizirana verzija programa bolj zanesljive rezultate, se moramo osredotočiti predvsem na primerjavo *1. mesta* in prvih treh mest. Tako pri opciji »standard«, kot tudi pri opciji »loose«, je druga verzija programa na *1. mesto* uvrstila največ molekul, 20 več kot prva verzija program. Tudi prva tri mesta so bila več molekulam določena z drugo verzijo programa, kot s prvo. Lahko bi rekli, da je razlika na prvih treh mestih med obema verzijama programa očitna, zlasti pri opciji »loose« (druga verzija 64,67%, prva verzija 42,00%).

Iz dobljenih rezultatov prihajamo do zaključka, da je optimizirana verzija programa bolj zanesljiva, ter da se pravilnim rezultatom lažje približamo z opcijo »loose«. Razlog za bolj zanesljive rezultate v drugi verziji programa pripisujemo optimizaciji programa, saj je z uvedbo SMARTS kod vsak atom in njegovo okolico možno veliko bolj natančno definirati. Največ napak kot posledica nenatančne definiranosti okolice atomov smo zaznali v prvi verziji programa pri reakcijah *N-* in *O-* dealkiliranja, kjer sta kot reakcijski center označena dušikov ali kisikov atom, kar je še bolj razvidno pri rezultatih glede na barvo atoma (slika16). To je tudi eden izmed glavnih vzrokov nenatančnih rezultatov v prvi verziji programa.

5.2 REZULTATI GLEDE NA BARVO ATOMA

S pomočjo programa MP2D lahko hitro in enostavno, na vizualen način, predpostavimo katera mesta v molekuli so metabolno nestabilna. To nam omogoča funkcija programa, s

pomočjo katere program atome v molekuli obarva v različne barve. Barva atoma je odvisna od razmerja na atomu, in nakazuje, kolikšna je verjetnost, da bo na atomu potekla reakcija. (16) Za razvoj novih učinkovin sta znanstvenikom v največjo pomoč rdeča ter bela barva (ni oznake). Atom, ki ima najvišjo verjetnost, da na njem poteče metabolizem, je obarvan rdeče, na atomih, ki so brez oznake, reakcija ne poteka. Običajne zahteve so takšne, da je potrebno atome, označene rdeče, zaščititi ali nadomestiti s kakšnimi drugimi fragmenti, da se s tem zmanjša verjetnost metabolizma, in tako podaljša terapevtsko delovanje učinkovine. Drugačne pa so zahteve za učinkovine, ki so v obliki predzdravil. Pri teh je zelo pomembno, da metabolizem poteče hitro, zato mora biti atom, ki je odgovoren za potek metabolizma obarvan rdeče, in v kolikor je brez barve, je potrebno molekulo modificirati.

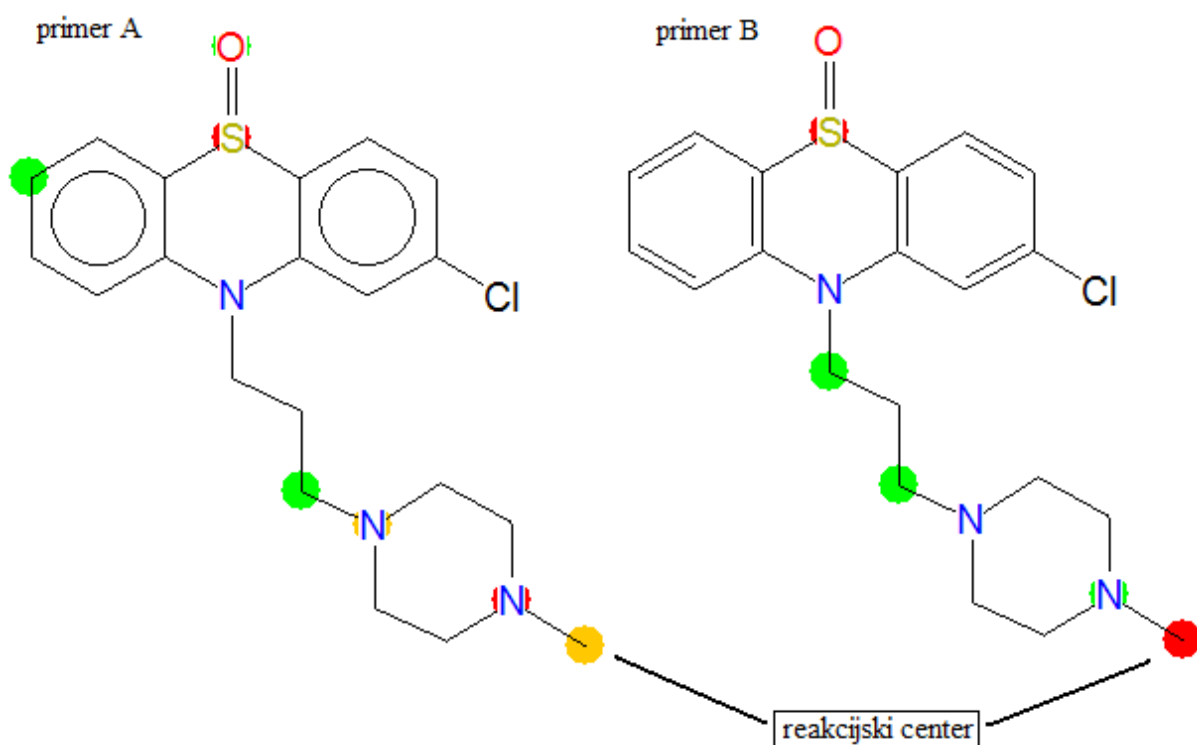
V diplomski nalogi smo za izbran set molekul preverili, kako dobro lahko glede na barvo atoma, na vizualen način, napovemo metabolizem v molekulah. Tudi tokrat smo preverjali, ali so napovedi z drugo verzijo programa bolj zanesljive.

Rezultate, ki smo jih dobili s prvo verzijo programa, prikazujeta tabela 5 in graf 4. Pri delu z opcijo »standard« je ostalo največ molekul, 57 molekul, brez napovedi. Z rdečo barvo je bilo označenih 47 molekul. Sledijo molekule označene z oranžno barvo, 30 molekul, ter 16 molekul, označenih z zeleno barvo. Pri opciji »loose« je bilo največ molekul označenih rdeče, 48 molekul, sledijo molekule označene z oranžno barvo, 44 molekul, brez oznake je ostalo 32 molekul. Razvidno je, da je tudi s tem načinom preverjanja največja razlika med obema opcijama med molekulami, ki so ostale brez oznake, ter najmanjša med tistimi, ki so bile obarvane rdeče. Ponovno smo potrdili, da izbira okolja (opcija programa) vpliva na dobljene rezultate.

Tudi z drugo verzijo programa smo to tezo še enkrat potrdili, saj je bila tudi tukaj največja razlika v številu molekul, ki so ostale brez metabolizma na eksperimentalno določenem atomu (graf 5). Z opcijo »standard« je bilo za to mesto določenih 55 molekul, z opcijo »loose« pa 25 molekul. Rdečo oznako je imelo pri opciji »standard« 62 molekul, pri opciji »loose« pa 67 molekul. Ker je pri izbiri opcije »loose« okolje manj natančno definirano kot pri izbiri opcije »standard«, obe verziji programa pustita pri opciji »loose« manj molekul nenapovedanih, kar kompenzirata s tem, da jih več napovevata z zeleno ali oranžno barvo.

Po tem, ko smo ugotovili kako vpliva definiranost okolice atoma na vizualne rezultate, smo preverili, ali s tem načinom tudi pri drugi verziji programa dobimo bolj zanesljive rezultate. Osredotočali smo se predvsem na primerjavo števila molekul označenih z rdečo barvo. Primerjavo med verzijama programa pri opciji »standard« in »loose« prikazujeta tabela 7 in graf 6. Z izbrano opcijo »standard« je več eksperimentalno določenih reakcijskih centrov označenih z rdečo barvo v drugi verziji kot v prvi. Z drugo verzijo je bilo rdeče označenih 62 atomov kot reakcijskih centrov, medtem ko je bilo s prvo verzijo rdeče označenih le 47 atomov. Prva verzija programa je ta primanjkljaj nadomestila z večjim številom oranžno označenih atomov. Le teh je bilo s prvo verzijo programa 30, z drugo pa le 17. Ta razlika ponovno nakazuje na pomanjkljivost prve verzije, ki kot reakcijski center v molekuli označuje dva ali več atomov hkrati. Največ takšnih napak smo zasledili pri reakcijah *N*- in *O*-dealkiliranja, kjer sta kot reakcijski center označena tudi dušikov in kisikov atom (slika 16). Razlika med številom neoznačenih atomov je zanemarljiva (2 molekuli več v starem programu), atomov označenih z zeleno barvo, pa je bilo v obeh programih enako (16 molekul).

Do podobnih rezultatov smo prišli tudi s primerjavo med verzijama pri izbiri opcije »loose«. Več atomov označenih z rdečo barvo je bilo v drugi verziji, oranžno označenih pa v prvi verziji. Razlika 19. atomov označenih z rdečo barvo (67 atomov prva verzija programa, 48 atomov druga verzija programa), podkrepi teorijo, da je izboljšava programa pripomogla k bolj zanesljivim rezultatom. Tudi tukaj lahko iz velikega števila oranžno označenih atomov v prvi verziji programa (44 molekul) sklepamo na nenatančno označevanje reakcijskih centrov (2 ali več), ki je bil glavni razlog za izboljšave in razvoj druge verzije programa.



Slika 16: Primer reakcije N-demetiliranja 2-kloro-10-[3-(4-metilpiperazin-1-il)propil]fenotiazin 5-oksida v prvi verziji programa (primer A) in v drugi verziji programa (primer B) z izbrano opcijo »standard«. Kot reakcijski center je v primeru A rdeče (prednostno) označen dušikov atom. V drugi verziji programa je reakcijski center označen rdeče (kot smo pričakovali).

5.3 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA TIP REAKCIJE

Molekule smo glede na tip reakcije razdelili v 4 skupine: oksidacije, redukcije, hidrolize in ostale reakcije. Oksidacijam je pripadalo 82 molekul, redukcijam 9 molekul, med reakcije hidrolize je bilo uvrščenih 39 molekul, 20 molekul je pripadalo ostalim reakcijam.

Pri delu z opcijo »standard«, je bilo pri reakcijah hidrolize največ molekulam na eksperimentalno določenem reakcijskem centru določeno razmerje 1,00. 43,59% molekul na katerih so potekle reakcije hidrolize je bilo uvrščenih na 1. mesto. Temu so sledile molekule, ki smo jih uvrstili med ostale reakcije, najnižji odstotek molekul na 1. mestu je bil med reakcijami oksidacije. Reakcije hidrolize ohranijo najvišji odstotek tudi ko preračunamo molekule na prvih treh mestih. Tokrat jim z 52,44% sledijo reakcije oksidacije.

Pri ostalih reakcijah je polovica molekul na prvih treh mestih, ter polovica na mestu 4 ali več in ni metabolizma. Podobna razporeditev molekul ostalih reakcij je tudi pri opciji »loose«. Tokrat je na prvih treh mestih 55,00% molekul, ostalih 45,00% je na mestih 4 ali več in ni metabolizma. Pri ostalih reakcijah vidimo torej v obeh primerih približno enako razporeditev – približno polovici molekul je reakcijski center pravilno določen, z razmerjem 1,00 na atomu, približno polovici pa program določi razmerje na reakcijskem centru 0,00. Takšno razporeditev pripisujemo dejstvu, da pod ostale reakcije spadajo zelo specifične reakcije, ki se ne dogajajo pogosto, zato veliko teh reakcij še ni vključenih v bazo podatkov s katero smo operirali. Za takšne reakcije je bilo zato razmerje na reakcijskem centru 0,00. Reakcije, ki pa so vključene v bazo podatkov, so toliko specifične in posebne, da je tudi ujemanje med matricami popolno, in tako imajo reakcijski centri preračunano razmerje 1,00.

Rezultati dobljeni z opcijo »loose« nakazujejo, da so najbolj zanesljive napovedi pri reakcijah redukcije. 77,78% molekulam, na katerih poteče redukcija, je bilo pripisano *1. mesto*. Prva tri mesta je zasedlo 88,89% vseh molekul redukcije. Pri opciji »standard« je bilo razmerje obratno, največ molekulam redukcije je bilo za reakcijski center namreč določeno, da reakcija tam sploh ne poteka. Prva tri mesta pa je zasedlo le 44,44% molekul, kar je bilo najmanj med vsemi tipi reakcij pri opciji »standard«. Zakaj prihaja pri reakcijah redukcije do takšnih razlik ne vemo, menimo pa, da je bil vzorec premajhen (le 9 molekul), da bi lahko predpostavili, da so rezultati zanesljivi. Najbolj verjetno je, da so dobljeni rezultati zgolj naključje, česar pa zaradi premajhnega vzorca ne moremo potrditi.

Če reakcij redukcij, zaradi premajhnega vzorca, ne upoštevamo, so rezultati dobljeni z obema opcijama med seboj zelo podobni in primerljivi. Tudi pri opciji »loose« je brez upoštevanja reakcij redukcije največ pravilno določenih molekul med reakcijami hidrolize. Sledijo ostale reakcije, ter na koncu reakcije oksidacije. Pri reakcijah hidrolize prva tri mesta zaseda kar 82,05% vseh molekul. Reakcije oksidacije prva tri mesta zasedajo z 56,10%.

Zaključke, do katerih smo prišli z rezultati glede na mesto metabolizma, smo podkrepili z rezultati glede na barvo atoma. Molekule, ki smo jih označili pod ostale reakcije, so pri obeh izbranih opcijah imela 45,00% molekul z rdeče označenim reakcijskim centrom. Pri opciji »standard« je bilo za ostale reakcije 50,00% molekul brez oznake, medtem ko je bilo pri opciji »loose« brez oznake 30,00% molekul, 10,00% molekul pa je imelo reakcijski

center označen zeleno. Rezultatov redukcije zaradi premajhnega vzorca molekul tudi tokrat nismo upoštevali, kažejo pa podobne rezultate kot prej. Z opcijo »standard« je med vsemi tipi reakcij bilo med reakcijami redukcije najmanj reakcijskih centrov označenih rdeče (44,44%), z opcijo »loose« jih je bilo rdeče označenih največ (88,89%). Če rezultatov redukcije ne upoštevamo, je največ rdeče označenih reakcijskih centrov med reakcijami hidrolize, tako pri opciji »standard«, kot tudi pri opciji »loose«. Sledijo jim ostale reakcije, ter na koncu reakcije oksidacije.

Zaradi majhnega vzorca reakcij redukcij dobljene rezultate težko posplošimo, lahko pa trdimo, da je z drugo verzijo programa možno narediti zadovoljivo napoved metabolizma za reakcije hidrolize. Zlasti pri opciji »loose«, kjer je bil odstotek pravilno določenih reakcijskih centrov na prvih treh mestih zelo visok, 82,05%. Zadovoljive rezultate dobimo tudi pri reakcijah oksidacije, saj je bil za prva tri mesta pri obeh opcijah odstotek molekul nad 50,00%. Pri ostalih reakcijah pa je pomembno za kakšno reakcijo gre, saj so nekatere reakcije zelo specifične in se redko pojavljajo, zato obstaja velika možnost, da baza podatkov kakšne reakcije še ne vsebuje. V kolikor pa je reakcija v bazi že prisotna, je verjetnost da je program ne bi pravilno označil zanemarljiva. V kolikor je program ne določi pravilno, obstaja velika verjetnost, da je to zaradi kakšnega fragmenta v molekuli, ki je prav tako zelo specifičen in vpliva na celoten izračun.

5.4 REZULTATI DRUGE VERZIJE PROGRAMA GLEDE NA MOLEKULSKO MASO MOLEKUL

Za preučevanje vpliva velikosti molekul na rezultate napovedovanja metabolizma smo molekule razdelili v tri skupine. V prvi skupini so bile molekule z molekulsko maso do 300 g/mol, v drugi z molekulsko maso med 300 g/mol in 500 g/mole, ter v zadnji skupini molekule z molekulsko maso večjo od 500 g/mol. Največ je bilo molekul z molekulsko maso med 300 in 500 g/mol, 83 molekul. Molekul z molekulsko maso manjšo od 300 g/mol je bilo 37; 30 je bilo molekul z molekulsko maso večjo od 500 g/mol.

Iz rezultatov, dobljenih glede na mesto napovedi metabolizma, smo prišli do zaključka, da program najmanj zanesljivo določi reakcijske centre molekulam večjim od 500 g/mol. To

smo potrdili z obema opcijama programa. Za ostali dve skupini molekul pa rezultati ene opcije programa niso potrjevali druge opcije programa. Pri opciji »standard« so bili najbolj točni rezultati za molekule velikosti med 300 in 500 g/mol. 63,86% teh molekul je imelo reakcijski center določen na enem izmed prvih treh mest. Molekule manjše od 300 g/mol so imele med prva tri mesta določen reakcijski center le pri 43,24% molekulah.

Pri delu z opcijo »loose« je program najbolj zanesljivo določil reakcijske centre molekulam manjšim od 300 g/mol. Kar 70,27% vsem molekulam manjšim od 300 g/mol so bili določeni reakcijski centri med prva tri mesta. Odstotek molekul med 300 in 500 g/mol, ki so imele določen reakcijski center med prva tri mesta, je bil skoraj enak kot pri opciji »standard«, 65,06%.

Iz dobljenih rezultatov je razvidno, da za srednje velike molekule izbira opcije nima velikega vpliva na zanesljivost rezultatov. Še največja je razlika, kadar je na reakcijskem centru označeno, da metabolizem na tem atomu ne poteka. Teh je bilo pri opciji »standard« 20,48%, pri opciji »loose« pa 6,02%. Vendar pa je to razlika, ki jo pričakujemo, in je prisotna pri vseh skupinah molekul in smo jo dokazali in potrdili že z prejšnjimi rezultati, pri primerjavi med opcijama znotraj programa. Bolj zanimiva je razlika med rezultati pri majhnih molekulah. Vidimo namreč, da je razlika že na 1. mestu (32,43% molekul pri opciji »standard« in 40,54% molekul pri opciji »loose«), še večja pa je po primerjavi prvih treh mest, skoraj 30,00%. Iz tega sklepamo, da program lažje pravilno obdela manjše molekule pri opciji »loose«.

Do podobnih zaključkov smo prišli tudi po obdelavi rezultatov glede na barvo atoma. Največji odstotek rdeče označenih reakcijskih centrov je bil pri majhnih molekulah (do 300 g/mol) pri delu z opcijo »loose«. Vendar, pa je bilo najmanj rdeče označenih reakcijskih centrov prav tako pri majhnih molekulah, tokrat pri delu z opcijo »standard«. Tudi iz teh rezultatov je razvidno, da ima izbira opcij največji vpliv na določanje reakcijskih centrov pri majhnih molekulah. Pri velikih in srednje velikih molekulah je namreč število rdeče označenih atomov pri obeh opcijah skoraj enako.

6 SKLEP

V okviru diplomske naloge smo se seznanili s programom za napovedovanje mest metabolizma v molekulah MP2D, s prvo verzijo kot tudi z drugo modificirano verzijo programa. S pomočjo programa smo napovedovali metabolizem novim molekulam. Raziskovali smo ali so s pomočjo izboljšane druge verzije programa rezultati zanesljivejši in uporabni za pomoč pri načrtovanju novih učinkovin. Ker program nudi tudi možnost izbire okolja atomov nas je zanimalo tudi, kakšne rezultate dobimo pri različni definiranosti le tega.

Iz dobljenih rezultatov smo prišli do ugotovitve, da izbira okolja pomembno vpliva na potek napovedi metabolizma v molekulah. Pri delu z opcijo »loose«, kjer je okolje atoma manj natančno definirano, je program za več molekul na eksperimentalno določenem reakcijskem centru pravilno napovedal metabolizem, kot pri delu z opcijo »standard«, ki ima za posledico bolj striktno definirano okolje vsakega atoma v molekuli. Tako so bili rezultati bolj primerljivi z našimi eksperimentalno dobljenimi rezultati kadar smo izbrali opcijo »loose«. To smo potrdili z obema verzijama programov.

Z primerjavo druge verzije programa s prvo smo dokazali, da so modifikacije vključene v drugo verzijo programa znatno vplivale na kakovost napovedi metabolizma. Z drugo verzijo programa so bili rezultati pri obeh izbranih opcijah bolj zanesljivi in pravilno preračunani. Odstotek molekul na prvih treh mestih napovedi metabolizma je bil v drugi verziji programa pri obeh opcijah, tako »standard« kot »loose«, večji od 50,00%, medtem ko je bil pri prvi verziji programa pri obeh opcijah manjši. Na podlagi tega si upamo trditi, da se lahko drugo verzijo programa vključuje v raziskovalne namene kot pomoč pri odkrivanju novih učinkovin.

Z drugo verzijo programa je bilo možno uspešno napovedati metabolizem tudi različnim tipom reakcij. Tukaj bi edino izpostavili reakcije redukcije, ki jih zaradi premajhnega vzorca pri končnih ugotovitvah nismo upoštevali. V kolikor bi želeli rezultate za ta tip reakcij posplošiti, bi morali načrtno povečati vzorec (vsaj okrog 30 molekul) in pogledati kakšne so napovedi metabolizma druge verzije programa samo za ta vzorec molekul. Najboljše napovedi metabolizma so bile pri reakcijah hidrolize.

Tako kot različni tipi reakcij vplivajo na napoved metabolizma, tako ima tudi velikost molekul pomemben vpliv na dobljene rezultate. Pri raziskovanju druge verzije programa smo izpeljali tudi poskus vpliva molekulske mase na pravilne napovedi metabolizma. Pri

tem smo prišli do zaključka, da izbira okolja atomov nima pomembne vloge pri srednje velikih molekulah. Največji vpliv okolja smo zaznali pri majhnih molekulah, kjer je program mnogo bolje napovedal metabolizem pri izbrani opciji »loose«.

Iz poteka celotnega dela lahko zaključimo, da nam je uspelo dokazati smiselnost modifikacij programa, kot tudi uporabnost druge verzije programa v raziskovalne namene. Smiselno bi bilo še razčleniti bazo podatkov v drugi verziji programa na različne vrste (posebej za ljudi, pse,...), tako kot je to storjeno že v prvi verziji. Prav tako menim, da bodo rezultati še bolj uporabni, ko bo v program vključena novejša baza podatkov in s tem več reakcij in molekul. Trenutno program s katerim smo operirali vključuje bazo iz leta 2005.

Program s katerim smo operirali je enostaven za uporabo, avtomatiziran in ima potencial za nadaljnje modifikacije. Razvoj tovrstnih programov je za farmacevtsko znanost zelo pomemben, saj omogoča hiter razvoj novih učinkovin in zmanjševanje stroškov razvoja.

7 LITERATURA

1. Camille Georges Wermuth, Bernard Testa: The practice of medicinal chemistry: Biotransformation Reactions and Their Enzymes. London, third edition 2008: 655-673.
2. Comprehensive Medicinal Chemistry II, Volume 5: ADME-Tox Approaches. W. F. Tager: Principles of Drug Metabolism 1: Redox Reactions. University of Washington, 2007: 87-132.
3. Samuel Edward Adams: Molecular Similarity and Xenobiotic Metabolism. Doctorate. Trinity College. University of Cambridge, 2010.
4. Andrej Bugrim, Tatiana Nikolskaya, Yuri Nikolsky: Early prediction of drug metabolism and toxicity: systems biology approach and modeling. Drug Discovery Today 2004; 9: 127-135.
5. Thomas A. Baillie: Metabolism and Toxicity of Drugs. Two Decades of Progress in Industrial Drug Metabolism. Chem. Res. Toxicol. 2008; 21: 129-137.
6. Sean Ekins, Sergey Andreyev, Andy Ryabov, Eugene Kirillov, Eugene A. Rakhmatulin, Andrej Bugrim, Tatiana Nikolskaya: Computational prediction of human drug metabolism. Expert Opin. Drug Metab. Toxicol. 2005; 1: 303-324.
7. Gabriele Cruciani, Emanuele Carosati, Benoit De Boeck, Kantharaj Ethirajulu, Claire Mackie, Trevor Howe, Riccardo Vianello: MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. J. Med. Chem. 2005; 48: 6970-6979.
8. Mingyue Zheng, Xiaomin Luo, Qiancheng Shen, Yong Wang, Yun Du, Weiliang Zhu, Hualiang Jiang: Site of metabolism prediction for six biotransformations mediated by cytochromes P450. Structural bioinformatics 2009; 25: 1251-1258.

9. Scott Boyer, Ismael Zamora: New methods in predictive metabolism. *Journal of Computer- Aided Molecular Design* 2002; 16: 403-413.
10. Goodford, P. J.: A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 1985; 28: 849-857.
11. Carosati E., Sciabola S., Cruciani G.: Hydrogen bonding interactions of covalently bonded fluorine atoms : From crystallographic data to a new angular function in the GRID force field. *J. Med. Chem.* 2004; 47: 5114-5125.
12. Marcel J. de Groot: Designing better drugs: predicting cytochrome P450 metabolism. *Drug Discovery Today* 2006; 11: 601-606.
13. De Groot M. J.: In silico methods for predicting ligand binding determinants of cytochromes P450. *Curr. Top. Med. Chem.* 2004; 4: 1803-1824.
14. Diansong Zhou, Lovisa Afzelius, Scott W. Grimm, Tommy B. Andersson, Randy J. Zauhar, Ismael Zamora: Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. *Drug metabolism and disposition* 2006; 34: 976-983.
15. Chothia C., Lesk A. M.: The relation between the divergence of sequence and structure in proteins. *EMBO J.* 1986; 5: 823-826.
16. Scott Boyer, Catrin Hasselgren Arnby, Lars Carlsson, James Smith, Viktor Stein, Robert C. Glen: Reaction Site Mapping of Xenobiotic Biotransformatios. *J. Chem. Inf. Model.* 2007; 47: 583-590.
17. Bender A., Mussa H. Y., Glen R. C.: Molecular Similarity Searching using Atom Environments, Information-Based Feature Selection and a Naive Bayesian Classifier. *J. Chem. Inf. Comput. Sci.* 2004; 44: 170-178.

18. Xing L., Glen R. C., Clark R. D.: Predicting pKa by Molecular Tree Structured Fingerprints and PLS. *J. Chem. Inf. Comput. Sci.* 2003; 43: 870-879.
19. <http://accelrys.com/products/pdf/dg-pharmacology-package-ds.pdf>, dostopno 8.3.2011
20. <http://www.biomedcentral.com/1471-2105/11/362>, dostopno 14.2.2011
21. <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>, dostopno 29.3.2011