

TARTU ÜLIKOOL
Füüsika-keemiateaduskond
Keemilise füüsika instituut

KALEV TAKKIS

HIV-1 PROTEAASI INHIBIITORITE MODELLEERIMINE

Magistritöö teoreetilises ja arvutikeemias

Juhendaja: Vanemteadur Ph.D. SULEV SILD

Tartu 2007

Sisukord

1	Sissejuhatus	3
2	Kirjanduse ülevaade	4
2.1	HIV	4
2.2	HIV proteaas	6
2.3	HIV-1 proteaasi inhibeerimine	8
2.4	Arvutusmeetodid	10
2.5	Lineaarsed mudelid	14
3	Andmekomplekt ja metoodika	19
3.1	Andmekomplekt	19
3.2	Metoodika	20
4	Tulemused ja analüüs	22
4.1	Mudel	22
4.2	Mudeli rakenduspiirid	24
5	Kokkuvõte	27
6	Conclusion	28
	Kirjandus	29
	Lisa A Andmekomplekt	36

1

Sissejuhatus

AIDS-i põhjustav HIV-viirus on tänu oma pikale peiteajale ja heale kohanemisvõimele suutnud võrdlemisi lühikese ajaga edukalt levida üle maailma. Kuigi pandeemia näitab mõningaid stabiliseerumismärke, kannab siiski hetkel infektsiooni ligi 40 miljonit inimest. Tingituna probleemi akuutsusest käib intensiivne uurimistöö HIV nakkusega võitlemisel ja üheks enamkasutatavaks meetodiks on kujunenud viiruse proteaasi inhibeerimine. Sobivate inhibiitorite leidmiseks on rakendatud kõiki teadaolevaid vahendeid, sealhulgas loomulikult ka arvutuskeemia meetodeid.

Käesolev töö keskendub ühele potentsiaalsele proteaasi inhibiitorite klassile, mille aluseks on tsükliline urea. Kirjanduse ülevaates kirjeldatakse lühidalt HIV-i elutsükli, proteaasi funktsioone ja selle inhibeerimist, ning antakse ülevaade seni avaldatud arvutuskeemia-alastest töödest tsükliliste ureate modelleerimise kohta. Kasutades kvantitatiivsete struktuur-omadus sõltuvuste meetodit, leitakse eksperimentaalses osas lineaarne seos molekuli suurust, kuju ja laengujaotust kirjeldavate molekulaardeskriptorite ja inhibeerimiskonstandi vahel. Erinevalt varasematest kirjanduses avaldatud uurimustest on käesoleval juhul kasutusel suurem ja mitmekesisem andmekomplekt.

Magistritöoga on seotud kaks publikatsiooni,

- U. Maran, S. Sild, I. Kahn, **K. Takkis**, „Mining of the chemical information in GRID environment”, *Future Generation Computer Systems*, 23, 2007, 76-83
- **K. Takkis**, S. Sild, „QSAR modeling of HIV-1 Protease inhibition on six- and seven-membered cyclic ureas”, *Bioorganic and Medicinal Chemistry Letters*, (kirjastuses retsenseerimisel)

Esimeses neist on modelleeritud üht osa käesolevas töös kasutatud andmekomplektist, teine on allpoolkirjeldatud tulemuste publitseerimine.

2

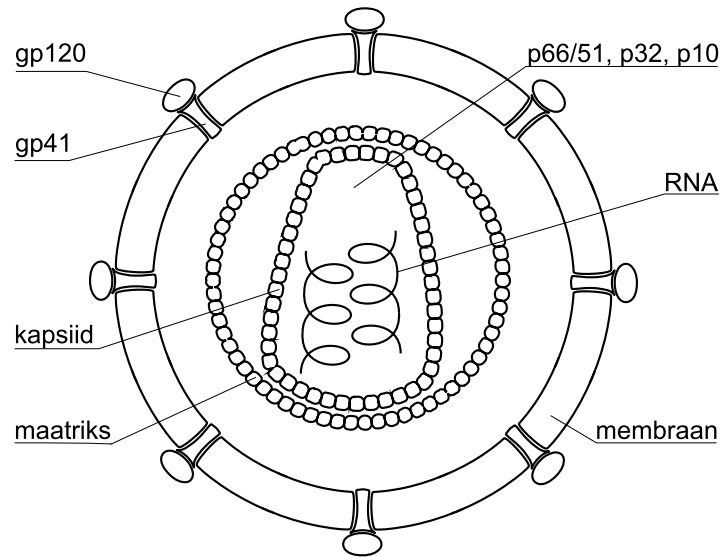
Kirjanduse ülevaade

2.1 HIV

Omandatud immuunpuudulikkuse sündroom (AIDS) on haigus, mis on tingitud immuunsüsteemi kahjustamisest. Hoolimata rohkem kui kahekümne aasta pikkusest intensiivsest uurimistööst ei ole sellele praeguse seisuga tõhusat ravi, võimalik on ainult tõsta nakatunud isiku oodatavat eluiga. AIDS-i tekitab inimese immuunpuudulikkuse viirus (HIV)[1] millest eksisteerib kaks inimesele ohtlikku liiki, HIV-1 ja HIV-2. HIV-1 pärineb tõenäoliselt Aafrikast, Kameruni aladel elanud šimpansitelt ja kandus inimestele millalgi 20. sajandil [2, 3]. Sellest alates on HIV-1 edukalt levinud, sellest on arenenud palju erinevaid tüvesid ja alamliike[4]. HIV-2 pärineb tõenäoliselt mangabeiks kutsutud pädiklaselt (*Cercocebus atys*), ta on tunduvalt vähem nakkav ja siiani piirdub selle levik peamiselt lääne-Aafrikaga[5]. Praeguse pandeemia taga on peamiselt HIV-1, ÜRO Aidsiprogrammi (UNAIDS) ja Maailma terviseorganisatsiooni (WHO) hinnangul oli maailmas 2006. aasta jaanuari seisuga 38.6 miljonit nakatunut kusjuures ainuüksi 2005. aasta jooksul liisandus 4.1 miljonit [6]. Kuna HIV-2 on tugevalt vähemuses, siis üldiselt rääkides HIV-ist viidatakse reeglina HIV-1-le. Seda tava on rakendatud ka käesolevas töös.

HIV on retroviirus, mis tähendab, et tema pärilikkusinfo on talletatud RNA kujul mis paljunemistsükli ajal kopeeritakse peremeesraku DNA-sse. Nagu nimi viitab, on peremeesrakkudeks immuunsüsteemi rakud, nagu CD4⁺T rakud, makrofaagid ja dendriittrakud. HIV-i geneetiline materjal sisaldab 9 geeni, mis kodeerivad 15 erinevat valku. Geenoom koosneb kahest RNA-ahelast mida ümbritseb kapsiid (tehtud proteiinist p24). Kapsiidi ümbritseb maatriks (proteiin p17), mida omakorda ümbritseb lipiidsest kaksikkihist membraan. Membraani pinnal on glükoproteiinid gp120 ja transmembraansed proteiinid gp41, mis on olulised peremeesrakkude äratundmisel. Terviklik viirusosake, nagu kujutatud joonisel 2.1, kannab nime virion.

Peremeesraku nakatamiseks peab HIV sisestama oma geneetilise materjali rakku. See



Joonis 2.1: HIV virion

algab gp120 sidumisega CD4 retseptoriga peremeesraku pinnal. Sellest tulenev konformatsioonimuutus võimaldab gp120-l seostuda teise retseptoriga, tavaliselt CCR5 või CXCR4 kemokiini ko-retseptoriga. Viiruse ja raku membraanid lähenevad ja toimub fusioon, mille tagajärjel HIV-i pöördtranskriptaas, integraas, RNA ja teised valgud satuvad peremeesraku tsütoplasmasse. Siin tõlgitakse kõigepealt RNA DNA-ks. Võtmerolli mängib selles pöördtranskriptaas (*reverse transcriptase*, RT) ning protsess mida see katalüüsib on sama mis teiste DNA polümeraaside korral. Pöördtranskriptaas teeb RNA ahelat lugedes palju vigu, mille esinemissagedust on hinnatud kuni ühele paari tuhande aluspaari kohta[7]. Lisaks on RT RNA ahelaga nõrgalt seotud, mistõttu võib RT kogemata vahetada RNA ahelaid ja viia sellega rekombinantsete viiruste tekkeni. Selle tulemuseks on RT suuresti vastutav HIV-i suure geneetilise mitmekesisuse eest. RT kasutab DNA koostamisel raku tsütoplasmas leiduvaid nukleotiide ja neid ehituskive matkides on RT inhibiitoritel võimalus peatada RT katalüüsiv protsess. Pöördtranskriptaasi tulemusel saadud viiruse RNA transporditakse tuuma ja integraasi, järjekordse HIV-i ensüümi abil, sisestatakse raku genoomi.

Sisestatud geneetiline materjal käitub nagu iga teine geen, selle järgi kodeeritakse rakus leiduvate mehhanismide abil RNA, mis viiduna tuumast välja on aluseks valkude koostamisel. Ainult et antud juhul ei ole tulemuseks mitte rakule vajalikud valgud vaid viiruse proteiinide prekursorid, nimelt:

- gp160, mis sisaldab gp120-te ja gp41-te
- p55 (*gag*), mis sisaldab maatriksi- ja kapsiidi valke
- p160 (*gag-pol*), mis sisaldab, maatriksi- ja kapsiidivalke, proteaasi (p10), pöörd-

transkriptaasi (p66/51) ja integraasi (p32)

Valkude tootmise tulemusena sünteesitud viiruse eelvalgud kogunevad raku seina äärde ja algab uue virioni moodustumine. Raku proteaas lõikab gp160-ne lahti gp120-ks ja gp41-ks. Umbes 1500-st p55-st, 100-st p160-st ja kahest viiruse RNA koopiasst koostatakse uus viirus. Nukleokapsiid võtab kuju, RNA pakitakse sinna sisse ning viirus pungub rakust, moodustades välise membraani raku membraanist.

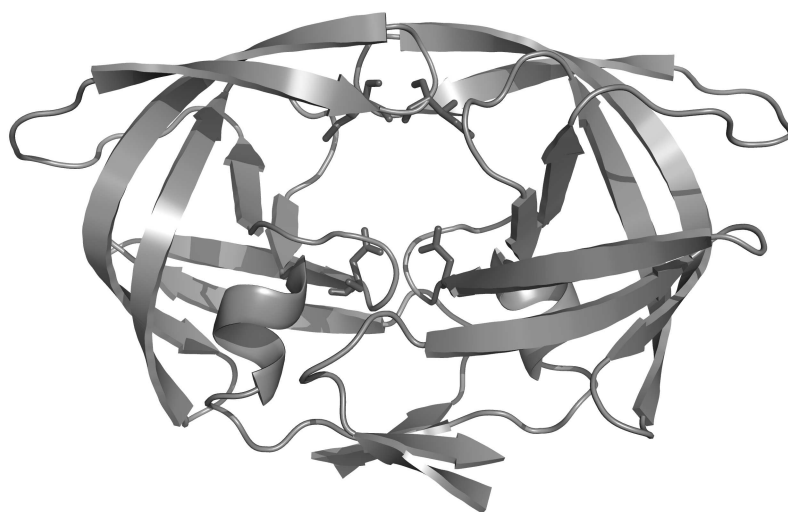
Rakust väljas iseseisvana ei ole HIV kohe veel valmis uusi rakke nakatama vaid enne tuleb läbi teha küpsemise faas. Polüproteiinide ahelate kõrge kontsentratsioon kutsub esile proteaasi aktiveerumise, kõigepealt lõikab see iseennast *gag-pol* ahelast välja ja töötleb siis ka ülejäänud proteiini prekursorid bioloogiliselt aktiivseteks valkudeks[8].

Kuna suur osa HIV -i elutsüklist tugineb peremeesraku mehhanismide ärakasutamisele, siis paljud võimalikud meetodid viirusest vabanemiseks tapaksid ka raku. Seetõttu on pingutused koondunud viirusele iseloomulikele protsessidele nagu viiruse seostumine rakuga, membraanide fusioon, viiruse RNA vabanemine kapsiidist, RNA pöördtranskriptaas, viiruse DNA sisestamine raku DNA hulka ja proteaasi tegevus[9]. Edukaimaks neist on kujunenud pöördtranskriptaasi ja proteaasi inhibeerimine[10, 11, 12, 13, 14].

2.2 HIV proteaas

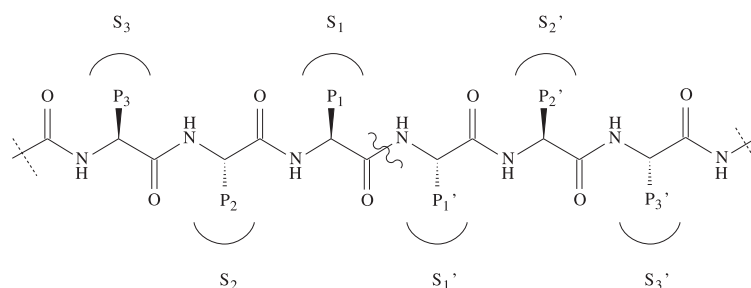
Proteaas üldiselt on iga ensüüm mis hüdrolüüsib aminohapetevahelisi peptiidsideid valguahelas. Nad esinevad looduslikult igas organismis ja etendavad tähtsat osa mitmesugustes füsioloogilistes reaktsioonides nagu toiduvalkude seedimine või organismile mittevajalike valkude lagundamine.

Joonisel 2.2 on kujutatud aktiivne HIV proteaas (HIV PR). See koosneb kahest identsest mittekovalentselt seotud monomeerist, mis kumbki koosneb 99-st aminohapest. Aktiivtsenter on keskse tühimiku põhjas, kus asub kummagi monomeeri *Asp-Thr-Gly* triaad. See järjestus on iseloomulik aspartüül-proteaasidele ja hilisemad uuringud ongi paigutanud HIV PR-i sellesse klassi. Tühimik on kaetud kahe väljaulatuva β -konformatsioonis klapi (*flap*). Klappid on väga paindlikud, substraadi aktiivtsentrisse sidumisel läbivad nad suured konformatsioonilised muutused[15, 16]. Kummagi klapis on isoleutsiini jäägid *Ile-50* ja *Ile-50'* aitavad siduda substraati NH rühmade abil. Ensüümi pinnal on leitud kaheksa sobivat sidumistsentrit või taskut substraadi sidumiseks (*binding pockets*), need on nimetatud traditsiooniliselt S_1, S_2, \dots ühele poole aktiivtsentrist ja S'_1, S'_2, \dots teises suunas. S_1 ja S'_1 taskud on väga hüdrofoobsed, S_2 ja S'_2 on peamiselt hüdrofoobsed välja arvatud *Asp-29, Asp-29', Asp-30* ja *Asp-30'*. S_3 ja S'_3 taskud on kõrvuti S_1 ja S'_1 -ga ning on samuti peamiselt hüdrofoobsed. Substraat seostub avatud konformatsioonis ja peab olema vähemalt 7, enamasti 8 aminohappe pikkune. HIV PR spetsiifika tuleneb



Joonis 2.2: HIV proteaas. *Asp*-rühmad aktiivtsentris ja *Ile*-rühmad klappides on kujutatud jämedamate pulkadena. PDB kood: 1bwb.

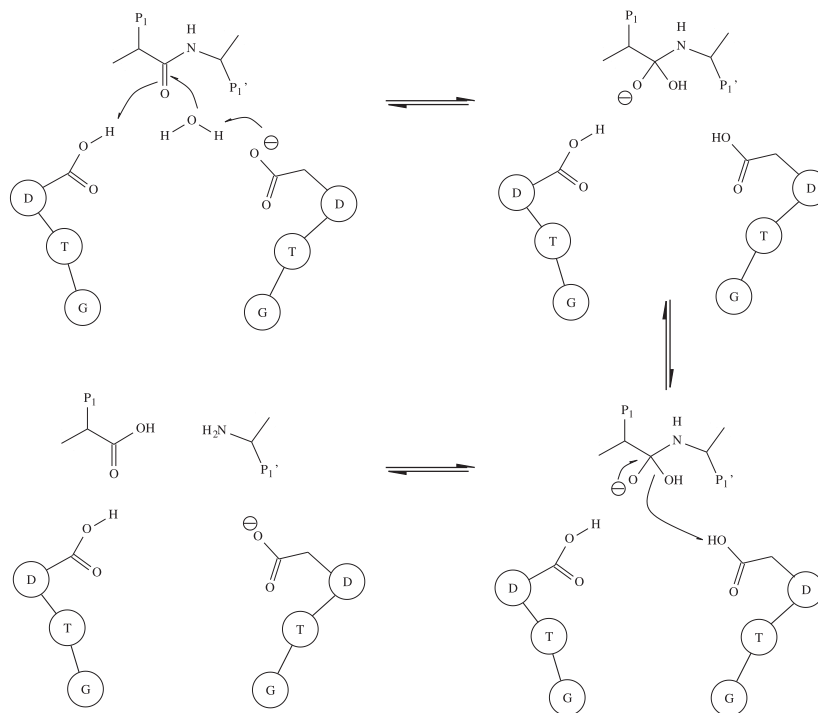
interaktsioonidest erinevate kõrvalahelatega ja ei ole eriti piiratud, juurdepääsu korral on proteaas võimeline lõikama praktiliselt iga peptiidsidet. Sobivamaid jäägid S_1 taskusse on *Tyr, Phe, Leu, Met* ja S_1' taskusse *Pro, Ala, Met, Phe, Leu* ja *Tyr*, kusjuures kergelt eelistatakse *Pro*-d. [17, 18]. Interaktsioonid substraadi ja ensüümi vahel on kujutatud joonisel 2.3.



Joonis 2.3: HIV PR interaktsioonid substraadiga. Katkev side on märgitud lainelise joonega.

Suguna *et al*[19] on välja pakkunud praeguseks enim aktsepteeritud aspartüülproteaaside katalüütilise mehhanismi, mis on kujutatud joonisel 2.4. Selle järgi aktiveerib negatiivse laenguga *Asp*-rühm veemolekuli, mis seejärel atakeerib substraadi hüdrolüüsitava sideme karbonüülrühma ning moodustab tetraheedrilise oksüaniooni vaheastme. Sellele järgnev katkeva sideme lämmastikuaatomi protoneerimine lõhub vaheastme ning viib hüdrolüüsiproductideni. Lisaks sellele hõlbustavad sideme katkemist ka interaktsioonid substraadi ja ensüümi vahel, peamiselt vesiniksidemed klappiga, mis asetavad peptiidsidemele pinget, pööravad seda tasandist välja ja vähendavad seega hüdrolüüsi

aktivatsioonibarjääri[20].

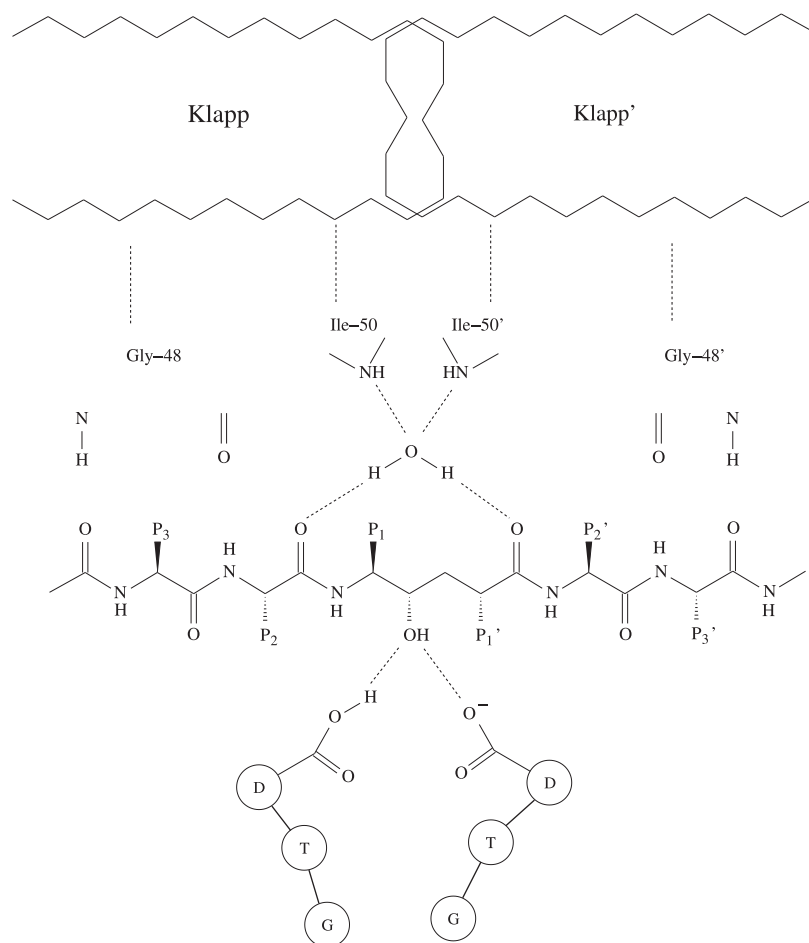


Joonis 2.4: Aspartüül-proteaaside lõikemehhanism

Erinevalt mitteviiruselistest proteaasidest ei moodustu HIV PR puhul vesiniksidemed klapi ja substraadiga mitte otse, vaid veemolekulist vaheastme kaudu. Nagu enamik röntgendifraktsioonil leitud proteaasi struktuure kinnitavad, moodustavad *Ile-50* and *Ile-50'* vesiniksidemed ühe ja sama veemolekuliga, nagu näidatud joonisel 2.5. Veemolekul moodustab veel kaks sidet substraadiga, mis väänab peptiidsidet ja teeb selle hüdrolüüsile vastuvõtlikumaks. Lõikemehhanismi on Jaskólski *et al*[21] kirjeldanud kui ühesammulist protsessi, kus nukleofiilne veemolekul ja elektrofiilne prooton atakeerivad peptiidsidet korraga. Mehhanismi detailid ei ole täpselt teada, ülevaate erinevatest väljapakutud meetoditest on koostanud Brik *et al*[13].

2.3 HIV-1 proteaasi inhibeerimine

HIV proteaasi bioloogiline funktsioon on lõigata *gag* ja *gag-pol* proteiini prekursorid bioloogiliselt aktiivseteks valkudeks[22]. Selle protsessi inhibeerimine takistab viiruse küpsemist ning tekkivad virionid ei ole võimelised nakatama uusi rakke[22, 23]. See avab võimaluse HIV nakkusega võitlemiseks. Proteaasi uuringute populaarsusele on kaasa aidanud ka tõik, et sellega töötamine ei ole eriti keeruline. Proteaasi on võimalik toota bakterite abil, seda saab puhastada denatureerivates tingimustes ja ta võldib end ilusti



Joonis 2.5: Veemolekul HIV PR aktiivtsentris

tagasi katalüütiliselt aktiivseks ensüümiks. Palju informatsiooni oli ka kogunenud teiste aspartüül-proteaaside, näiteks reniini uurimisest, ning sellele disainitud inhibiitoreid oli võimalik katsetada HIV proteaasil. Kõige selle tulemusena on HIV PR praeguseks üks paremini uuritud valke.

Ravimite disain ei ole lihtne protsess ja HIV proteaasi juhtum ei ole erand. Iga hea inhibiitor ei ole veel hea ravim, viimasele on vajalikud veel mitmed omadused. Ta peab küll tugevasti proteaasiga seonduma, kuid peale selle peab olema efektiivne ka viiruse küpsemise takistamisel. Need kaks omadust ei ole sugugi alati üksüheselt koos. Ta peab olema spetsiifiline, et mitte takistada teiste organismis leiduvate proteaaside tegevust. Suukaudne manustatavus on vägagi tervitatav omadus, nagu ka võime ületada hematoentsefaalset barjääri võitlemaks ajus oleva viirusega. Lisaks sellele, kuna HIV-i vastane ravi kestab elu lõpuni, siis peaks ravim olema ka piisavalt odav.

Üks võimalus inhibiitori kujundamiseks on proteaasi naturaalse substraadi jäljendamine, kus hüdrolüüsiv peptiidside asendatakse mõne mittekatkeva sideme, näiteks hüd-

roksüetüülamiini fragmendiga. Sellised peptiidi matkivad inhibiitorid (*peptidomimetic*) olid esimesed mis jõudsid ravimitena turule ja on praeguseni enamlevinud kliiniliselt aktsepteeritud proteaasi inhibiitorite klass, üheksast saadaolevast proteaasi inhibiitorist kaheksa on just seda tüüpi, nimelt Saquinavir, Nelfinavir, Ritonavir, Lopinavir, Indinavir, Amprenavir, Fosamprenavir ja Atazanavir (joonis 2.6). Üheksas olemasolev ravim, Tipranavir, on mittepeptiidne. Huvi mittepeptiidsete inhibiitorite vastu põhjustas soov saada väiksemaid inhibiitoreid ja saavutada seega parem veeslahustuvus ja suukaudne manustatavus. Kuigi neid hetkel veel ravimitena rohkem kasutusel pole, on uuritud mitmeid perspektiivseid inhibiitorite seeriaid, nagu asendatud 4-hüdroksü-püraan-2-oonid (mille hulka kuulub ka Tipranavir)[24], tsüklisel ureal põhinevad inhibiitorid[25] ja kromooni derivaadid[26].

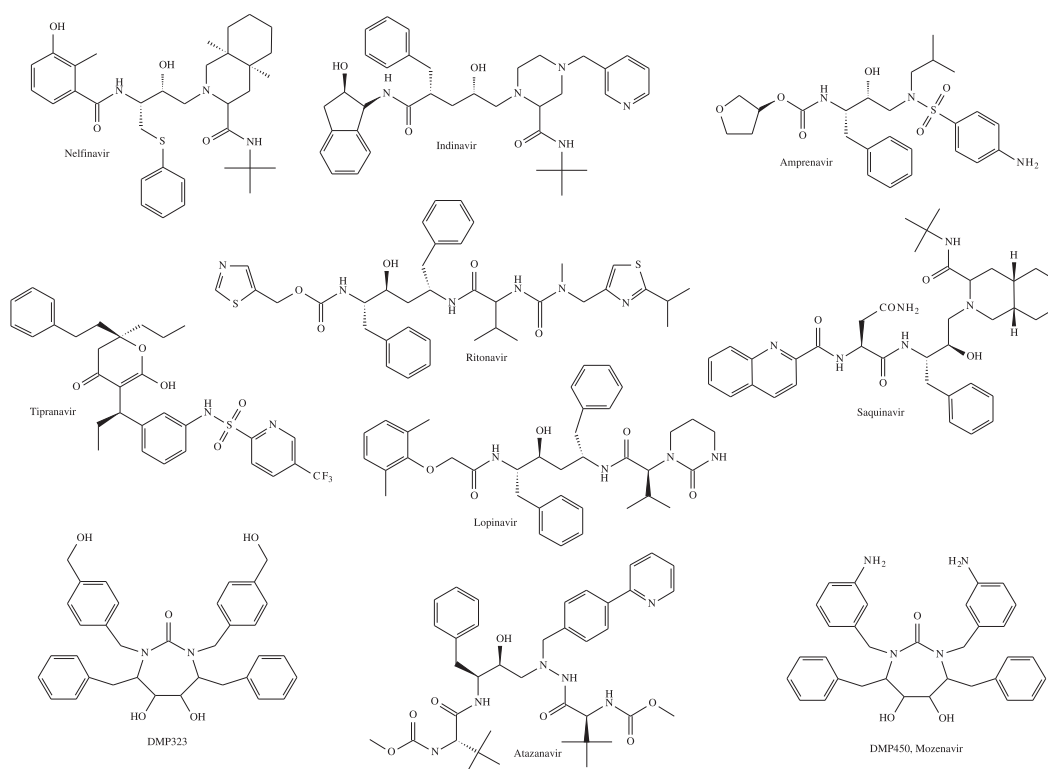
HIV-i kiire muteerumine seab ravimite disainile veel mõned raskused. Õieti ongi see üks põhiprobleeme HIV-i vastases võitluses - ka head ravimid kaotavad mõne aja jooksul oma efektiivsuse. Praeguseks on HIV proteaasis teada üle viiekümne mutatsiooni, millest 22 võivad põhjustada resistentsust inhibiitorite vastu[27]. Uute inhibiitorite disainis on tähtsale kohale tõusnud võime inhibeerida mitmeid muterunud tüvesid. Proteaasi inhibeerimisest on koostatud rohkelt ülevaateid[9, 11, 28, 29, 30].

DuPont Mercki poolt sünteesitud tsüklilisel ureal põhinevad inhibiitorid kuuluvad mittepeptiidsete inhibiitorite hulka[25]. Nad on väiksemad kui peptiidipõhised inhibiitorid ning reeglina proteaasi inhibeerimisel küllaltki head. Nende eripäraks on urea karbonüülne hapnik, mis asendab veemolekuli klappide vahel, ning tänu soodsale entroopilisele efektile annab paremad seostumisenergiad. Uureate struktuur lubab optimaalseid interaktsioone S_1 , S'_1 , S_2 ja S'_2 taskutega. Uureapõhised inhibiitorid on olnud ka kliinilistes katsetustes, näiteks DMP323 ja DMP450 (Mozenavir) kuid praeguseks on neist loobutud.

Inhibiitori efektiivsuse hindamiseks on kasutusel peamiselt kolm suurust, K_i , IC_X ja EC_X . Esimene neist, inhibeerimiskonstant, on oma olemuselt ensüümi ja inhibiitori vahelise interaktsiooni tasaakalukonstant. Linearsele skaalale viimiseks kasutatakse logaritmi ja tavaliselt antakse kirjanduses kujul $\log(1/K_i)$, $-\log K_i$ või pK_i . IC_X ja EC_X märgivad inhibiitori kontsentratsiooni, millel toimub $X\%$ -line inhibeerimine, vastavalt *in vitro* ja *in vivo*. Protsendiks on enamasti 50 või 90. Suurust EC_X on kasutatud mõnevõrra vähem, põhjuseks keeruline määramine.

2.4 Arvutusmeetodid

Tsüklilisel ureal põhinevad HIV proteaasi inhibiitorid sünteesiti esmakordselt 1994. aastal[25]. Sellest alates on neid sünteesitud seeriatega kaupa, saavutamaks paremaid inhibeerimis- ja farmakokineetilisi omadusi. Järjest suurem saadaolev eksperimentaalne



Joonis 2.6: HIV proteaasi inhibiitorid

andmehulk avas võimaluse kvantitatiivseks struktuur-omadus sõltuvuse (QSAR) analüüsiks. Selle käigus seotakse molekulide struktuur uuritava omadusega kujul

$$P = f(s) \quad (2.1)$$

kus struktuur on kirjeldatav mitmesuguste karakteristikute, deskriptorite abil. Deskriptoriks võib olla iga suurus mida on võimalik kvantitatiivselt antud struktuuridele kohandada. Nad jagunevad empiirilisteks ja teoreetilisteks, kus esimesed on eksperimentaalselt mõõdetud, teised teoreetiliselt arvatud. Teoreetilised omakorda on jagatavad konstitsoonilisteks, topoloogilisteks, geomeetrilisteks, elektrostaatilisteks, solvatatsioonilisteks, termodünaamilisteks ja kvantkeemilisteks deskriptoriteks[31]. Lihtsaimad neist on konstitutsioonilised, need kirjeldavad molekuli koostist ja on reeglina arvatavad molekuli brutovalemist. Topoloogilised deskriptorid vajavad arvutamiseks molekuli kahemõõtmelist struktuuri ning ülejäänud reeglina kolmemõõtmelist. Saadav QSAR-i mudel annab sõltuvalt kasutatavatest deskriptoritest rohkem või vähem täpsema võimaluse molekulide omaduste interpretatsiooniks ja ka uute ühendite omaduste arvutamiseks. Mudeli kvaliteeti saab kontrollida mitmete statistiliste parameetritega nagu Pearsoni korrelatsioonikoefitsient R või selle ruut R^2 , standardhälve s või selle ruut s^2 , Fisheri kriteerium F ja

ristvalideeritud korrelatsioonikoeffitsient R_{cv}^2 (q^2). Ristvalideerimise kaudu hinnatakse mudeli ennustusvõimet. Eelistatum võimalus ennustusvõime hindamiseks on siiski mudeli rakendamine treenimisel mittekasutatud ühendite omaduste arvutamiseks, mis on saavutatav näiteks andmekomplekti jagamisega treening- ja testkomplektiks. Kahjuks on sageli kasutatavate ühendite arv selleks liiga väike.

Üks lihtsamaid QSAR-i meetodeid on multilineaarne regressioon (MLR), kus molekuli aktiivsus seotakse deskriptoritega võrrandi 2.2 järgi,

$$P = b + \sum a_i x_i \quad (2.2)$$

kus b on vabaliige, a_i on regressioonikordaja ja x_i on deskriptor. Uureate aktiivsuse modelleerimisel on MLR meetodit palju rakendatud. Üks enimkasutatud parameetritest on hüdfoobsus, väljendatuna oktanool-vesi jaotuskoeffitsiendina $\log P$, tõenäoliselt tänu teadaolevatele hüdfoobsetele taskutele proteaasis. Kuna MLR meetodikaga arendatud mudelid on käesoleva töö peamine huviobjekt on põhjalikum ülevaade neist antud sektsioonis 2.5.

Alati ei ole sõltuvus deskriptorite ja uuritava omaduse vahel lineaarne, sellisel juhul annab parema tulemuse deskriptorite mittelineaarne transformatsioon või mõne mittelineaarse modelleerimismeetodi kasutamine. Uureate aktiivsuse hindamisel on kasutatud näiteks tehis-närvivõrku (*artificial neural network*, ANN). Milac *et al*[32] kasutasid seda 42-st seitsmeliikmelisel tsüklilisel uureal põhinevast proteaasi inhibiitorist koosneva andmekomplekti modelleerimiseks. Iga struktuuri kujutati vektorina mille elemendid kirjeldasid erinevaid struktuurseid omadusi. Mittelineaarsete efektide arvestamiseks valiti ülekandefunktsiooniks sigmoid. Parimad mudelid saadi kasutades koos $\log P$ -d, dipooli ja steerilist parameetrit, halvimal saadi aga $\log P$ asendamisel ruumalaga.

Fernández *et al*[33] kasutasid ANN meetodit modelleerimaks 55-st tsüklilisest uureast koosneva andmekomplekti K_i ja IC_{90} modelleerimiseks konstitutsiooniliste ja topoloogiliste deskriptorite abil. Parima närvivõrgu abil saadud mudeli statistilised parameetrid olid $R = 0.931$ ja $R_{cv}^2 = 0.703$, mis ületasid parimat lineaarset mudelit ($R = 0.862$; $R_{cv}^2 = 0.51$). IC_{90} jaoks lineaarsel meetodil ennustusvõimelist mudelit ei leitud, ANN suutis leida aga mudeli parameetritega $R = 0.891$ ja $R_{cv}^2 = 0.568$. Leitud mittelineaarsed mudelid osutasid lämmastiku aatomi olemasolu ja molekuli ruumala tähtsusele ühendi inhibeerimisvõimel.

Rohked röntgendifraktsioonimeetodil läbi viidud uuringud proteaasi kristallstruktuuri kohta võimaldasid rakendada eksperimentaalsetele kolmemõõtmelistele geomeetriatele toetuva kvantitatiivse struktuur-omadus sõltuvuse analüüsi (3D-QSAR) meetodeid. Ülevaate nende meetodite rakendamisest proteaasi inhibiitoritel on koostanud Debnath[34]. Võrdlev molekulaarvälja analüüs (Comparative Molecular Field analysis, CoMFA) on olnud üks populaarsemaid meetodeid uureate aktiivsuse modelleerimiseks. Debnath[35]

uuris selle meetodi abil komplekti seitsmelülilisi sümmeetrilisi uureaid ja leidis heade statistiliste näitajatega mudeli ($R^2 = 0.971$, $q^2 = 0.724$), mis suutis ka testkomplektis olevate ühendite aktiivsust hästi ennustada. Analüüs näitas, et steerilised/hüdrofoobsed ja elektrostaatilised interaktsioonid on uureate proteaasivastase aktiivsuse interpreteerimisel peaaegu sama kaaluga.

Nair *et al*[36] uurisid samuti uureate aktiivsust kasutades CoMFA meetodit. Kasutatud andmekomplekt koosnes 60-st ühendist, mis oli jagatud treening- ja testkomplekti vahel suhtes 49/11. Huvitav aspekt töö juures oli, et lisaks CoMFA parameetritele kasutati mudelis ka mitmesuguseid lisaparaameetreid nagu arvutatud ensüümi-inhibiitori kompleksatsioonenergia, inhibiitori ja ensüümi kontaktpindalad ja inhibiitori solvatatsioonenergia. Meetodika, mida nimetasid nad ise „laiendatud” CoMFA-ks, tõi kaasa R_{cv}^2 märkimisväärse kasvu, eriti kui mudelisse kaasati inhibiitori solvatatsioonenergia (~0.6-lt ~0.8-ni).

Eespool juba mainitud Debnath teostas veel ühe CoMFA töö tsükliliste uureate peal[37]. Kasutatud andmekomplekt koosnes seekord 118-st ühendist, mis on üks suuremaid kasutatud uureate andmekomplekte üldse. Treening- ja testkomplektiks jagamine toimus suhtes 93/25, kusjuures jagati ja arendati mudeleid mitu korda. Tulemuseks saadi mitmeid heade statistiliste näitajatega mudeleid kus $R^2 = 0.965...0.973$, $q^2 = 0.699...0.727$ ja testkomplekti $R^2 = 0.556...0.639$.

Peaaegu sama suurt andmekomplekti kasutati Tervo *et al*[38] poolt. Võrreldi mudelite kvaliteeti mis saadud automaatse ja käsitsi sooritatud molekulide joondamisel, nii CoMFA kui CoMSIA (*Comparative Molecular Similarity Indices Analysis*, võrdlev molekulide sarnasusindeksite analüüs) meetodil. Käsitsi joondamisel saadud mudelid olid paremate statistiliste näitajatega, kuid treeningkomplekti ühendite aktiivsusi ennustasid huvitaval kombel paremini automaatse joondamise abil saadud mudelid.

Sama andmekomplekti mida eelnevalt kasutas Nair *et al*[36], kasutas nüüd Senese *et al*[39]. Kasutatav metodoloogia kandis nime retseptorist sõltumatu 4D-QSAR. 4D-QSAR-i erinevus 3D-QSAR-st seisneb selles, et iga struktuuri jaoks ei kasutata ainult ühte geometriat, vaid keskmistatakse üle termodünaamiliselt sobivate struktuuride ansambli, kus siis mitmed geometriad moodustavad n.ö. neljanda mõõtme. Iga struktuuri jaoks postuleeritakse üks „aktiivne” konformatsioon, mida on võimalik kasutada edasises, näiteks 3D-QSAR analüüsis. 24 statistiliselt olulise ($q^2 > 0.8$) mudeli hulgast valiti klasterdamistehnika abil viis, mis oleksid ennustamises üksteisest sõltumatud. Autorid järeldasid, et kui inhibiitori ja ensüümi vastasmõjus on võimalikud mitu sidumismehhanismi, ei ole üksik geomeetria piisav ja arvestama peaks mitmeid võimalikke konformatsioone.

2.5 Lineaarsed mudelid

Järgneb ülevaade varasemates töodes avaldatud MLR meetodikat kasutades saadud mudelitest kus on tsükliliste uureate inhibeerimiskonstandi modelleerimiseks kasutatud teoreetilisi molekulaardeskriptoreid.

Wilkerson *et al*[40] sünteesisid seeria sümmeetrilisi tsüklilisi uureaid ja teostasid multilineaarse regressioonianalüüsi, kasutades deskriptoritena arvatud $\log P$ -d ($ClogP$), molekulaarmassi ($MW/100$), ionisatsioonipotentsiaali (IP) ja indikaatorparameetrit I väärtusega 1 ühenditele lämmastikuga kindlas positsioonis. Tulemuseks oli võrrand 2.3,

$$\begin{aligned} pK_i &= -1.273(\pm 0.512)IP + 0.076(\pm 0.021)ClogP^2 & (2.3) \\ &- 0.866(\pm 0.27)ClogP - 1.006(\pm 0.33)(MW/100) \\ &+ 0.99(\pm 0.184)I + 20.77(\pm 5.267) \\ n &= 29 \quad R = 0.864 \end{aligned}$$

kus n tähistab kasutatud ühendite arvu. Hilisemas töös[41] sünteesisid samad autorid 16 asümmeetriliselt asendatud ureat. MLR analüüsi tulemustena saadud mudelites (võrrandid 2.4 ja 2.5) olid kaks uut parameetrit, arvatud molaarne refraktsioon (CMR) ja amiidse prootoni keemiline nihe (ϑ_{NH}).

$$\begin{aligned} pK_i &= 0.221(\pm 0.076)ClogP^2 - 1.963(\pm 0.789)ClogP & (2.4) \\ &- 2.344(\pm 0.377)CMR - 0.592(\pm 0.118)\vartheta_{NH} \\ &+ 56.052(\pm 8.195) \\ n &= 15 \quad R = 0.941 \end{aligned}$$

$$\begin{aligned} pK_i &= 0.190(\pm 0.123)ClogP^2 - 2.377(\pm 1.346)ClogP & (2.5) \\ &- 7.398(\pm 2.308)(MW/100)^2 + 96.294(\pm 30.473)(MW/100) \\ &- 304.541(\pm 99.365) \\ n &= 15 \quad R = 0.852 \end{aligned}$$

Kuigi võrrandi 2.4 korrelatsioonikoefitsient oli kõrgem kui võrrandi 2.5 oma, eelistati viimase muutujate kõrge interkorrelatsiooni tõttu võrrandit 2.5. Sarnaselt käsitleti mõlemas töös ka pIC_{90} -ne modelleerimist.

Hüdrofoobsust uurisid hiljem põhjalikult Garg ja Bhatarai[42]. Koostati neli seitsmelülilistest uureatest koosnevat andmekomplekti, igauks erinevast kirjandusallikast[43,

44, 45, 46], ning esialgu modelleeriti kõiki eraldi. Andmekomplekte koos modelleerides saadi tulemuseks võrrand 2.6.

$$pK_i = -0.34(\pm 0.07)ClogP + 12.82(\pm 0.53) \quad (2.6)$$

$$n = 28 \quad R^2 = 0.801$$

Arvestades ühendite hulka ja deskriptorite arvu on tulemus päris hea ning see lubas autoritel järeldada, et hüdrofoobsus on inhibiitorite bioloogilise aktiivsuse juures üks olulisemaid kriteeriume. Garg ja Patel jätkasid uurimust kuuelüliliste tsükliliste ureatega[47], mis olid sünteesitud ja mõõdetud DeLucca *et al* poolt[48]. Lisaparameetritena on saadud võrrandis 2.7 kasutusel *CMR*. Võrrand 2.8 kirjeldab isomeeriaefekte, *I* on väärtusega 1 RRR- ja 0 SSS-isomeeride jaoks.

$$pK_i = -0.738(\pm 0.416)ClogP - 1.997(\pm 1.114)CMR \quad (2.7)$$

$$+ 14.029(\pm 2.633)$$

$$n = 12 \quad R^2 = 0.83$$

$$pK_i = 3.726(\pm 1.229)I + 5.653(\pm 0.869) \quad (2.8)$$

$$n = 8 \quad R^2 = 0.859$$

Gayathri *et al*[49] uurisid teist DeLucca *et al* sünteesitud kuuelülilistest sümmeetrilistest ja asümmeetrilistest ureatest koosnevat andmekomplekti[50]. Kasutati ühte topoloogilist indeksit, esimest järku valentsühenduvuse indeksit (*valence connectivity index of order 1*, ${}^1\chi^v$), mis kirjeldab aatomite omavahelist ühenduvust, molekuli hargnevust ja küllastumatust.

$$pK_i = 0.451(\pm 0.378){}^1\chi^v - 1.939(\pm 0.7)I_1 \quad (2.9)$$

$$+ 1.02(\pm 0.59)I_2 + 8.293(\pm 1.266)$$

$$n = 23 \quad R = 0.895$$

$$pK_i = 0.862(\pm 0.232){}^1\chi^v(P_2) + 0.438(\pm 0.153){}^1\chi^v(P'_2) \quad (2.10)$$

$$- 1.398(\pm 0.6)I_1 + 5.84(\pm 0.76)$$

$$n = 25 \quad R = 0.906$$

Mudel 2.9 on treenitud sümmeetriliste ja mudel 2.10 asümmeetriliste ureate peal. ${}^1\chi^v$ on

arvutatud P_2/P'_2 asendajatele eraldi, seepärast on mudelis 2.10 kaks indeksit. I_1 on indikaatorparameeter väärtusega 1 ühenditele, mille P_2/P'_2 asendaja sisaldab benseenituumas CN rühma. I_2 on teine indikaator, väärtusega 1 ühenditele mis sisaldavad benseenituumas $C(NH_2)=N-OH$ rühma.

Suuremat andmekomplekti modelleerisid Gupta *et al*[51], saades tulemuseks mudeli 2.11.

$$\begin{aligned}
 pK_i &= 2.751(\pm 0.771)\pi - 0.606(\pm 0.166)\pi^2 & (2.11) \\
 &+ 1.766(\pm 0.486)I_H - 1.257(\pm 0.603)I_o \\
 &- 1.68(\pm 0.436)I_e + 5.4 \\
 n &= 48 \quad R = 0.93 \quad s = 0.46
 \end{aligned}$$

Kasutatud parameetrid on hüdrofoobsuskonstant (π) ja indikaatorparameetrid I_H , I_e ja I_o väärtusega 1 kui P_2/P'_2 asendaja sisaldab vastavalt OH/NH₂-rühma, C-O-C rühma või on aroomaatne. Gupta *et al* teine sama aasta töö[52] sisaldas huvitavat katset modelleerida koos tsüklilisi ureaid ja tsüklilisi tsüaanoguanidiine. Tulemuseks oli mudel 2.12.

$$\begin{aligned}
 pK_i &= 2.942(\pm 0.878)\pi - 0.646(\pm 0.171)\pi^2 & (2.12) \\
 &+ 1.392(\pm 0.491)I_H - 0.625(\pm 0.332)I_g \\
 &+ 0.5208 \\
 n &= 46 \quad R = 0.847 \quad s = 0.53
 \end{aligned}$$

Parameeter I_g on väärtusega 1 tsüaanoguanidiinidele ja 0 ureatele. Oma P_1/P'_1 ja P_2/P'_2 asendajate poolest olid mõlemad aineklassid sarnased. Mudel 2.12 näitab, et proteaasi inhibeerimises on uread paremad. Seda on Jadhav *et al*[53] põhjendanud nii, et tsüaanoguanidiinrühm on võrreldes karbonüülrühmaga suurem ja sunnib ühte proteaasi klappi võtma ebasoodsamat asendit, mis annab tulemuseks halvemad seostumisenergiad.

Katritzky *et al*[54] kasutasid sama andmekomplekti nagu Gayathri *et al*[49], kuid erinevalt viimastest ei käsitlenud nad sümmeetrilisi ja asümmeetrilisi ühendeid eraldi.

$$\begin{aligned}
pK_i &= 2.168(\pm 0.535) + 0.108(\pm 0.01)^2 SIC & (2.13) \\
&+ 12.75(\pm 1.848) HDFPSA^{(2)} \\
R^2 &= 0.816 \quad R_{cv}^2 = 0.792 \quad F = 107
\end{aligned}$$

$$\begin{aligned}
pK_i &= 42.585(\pm 8.455) + 15.601(\pm 1.776) HDFPSA^{(2)} & (2.14) \\
&- 141.775(\pm 34.923) I^A \\
&- 6.204(\pm 0.732) J - 0.234(\pm 0.073) E_{e-n}^{min}(O-H) \\
R^2 &= 0.873 \quad R_{cv}^2 = 0.847 \quad F = 79
\end{aligned}$$

Kokku avaldati töös 4 mudelit, millest siin on ära toodud 2, mudelid 2.13 ja 2.14. 2SIC on teist järku strukture informatsiooni indeks (*structural information content of order 2*), $HDFPSA^{(2)}$ on vesinikudoonorite positiivselt laetud osapindala (*fractional positive surface area of hydrogen donors*), I^A on esimene inertsimoment (*first moment of inertia*), J on Balabani indeks ja $E_{e-n}^{min}(O-H)$ on minimaalne elektroni-tuuma tõmbejõud O-H sidemes (*minimum electron-nuclear attraction energy for O-H bond*).

Maran *et al*[55] kasutasid OpenMolGRID[56] süsteemi ureate inhibeerimisaktiivsuse modellemiseks. Andmekomplekt koosnes 80-st kuue- ja seitsmelülilise baasstruktuuriga ühendist mis jagati põhikomponentanalüüsi (PCA) abil treening- ja testkomplektiks, kusjuures treeningkomplekt koosnes 26-st ühendist ja testkomplekt ülejäänud 54-st. Tulemuseks oli mudel 2.15.

$$\begin{aligned}
pK_i &= -254.83(\pm 60.1) + 0.196(\pm 0.02) HA HDSA^{(2)} & (2.15) \\
&+ 2.858(\pm 0.589) E_C^{min} - 1.894(\pm 0.452) E_{tot}^{min}(C-O) \\
n &= 26 \quad R^2 = 0.81 \quad F = 44.88
\end{aligned}$$

$HA HDSA^{(2)}$ on pinnaga kaalutud laeng vesiniksideme doonorite pinnal (*area-weighted surface charge of hydrogen bonding donor atoms*), E_C^{min} on süsiniku minimaalne aatomiseisundi energia (*minimum atomic state energy for atom C*) ja $E_{tot}^{min}(C-O)$ on C-O sideme minimaalne vastasmõjuenergia (*minimum total interaction energy for bond C-O*). Mudelil on küllaltki hea ennustusvõime, testkomplekti ennustatud väärtuste $R^2 = 0.61$.

Nagu ilmneb, on seni kasutatud andmekomplektid olnud võrdlemisi väikesed ja ühendid nendes küllaltki sarnased. Suurimat ja mitmekesisemait komplekti kasutasid Maran *et al*[55], peale selle töö ei ole kuue- ja seitsmelülilisi ureaid koos modelleeritud. Käesolev

uurimus ongi selle töö edasiarendus ja täiendus. Võeti eesmärgiks koostada veelgi suurem ja mitmekesisem andmekomplekt ja uurida kuidas see allub analüüsile ning, kui võimalik, arendada statistiliselt ja ennustusvõimelt hea mudel.

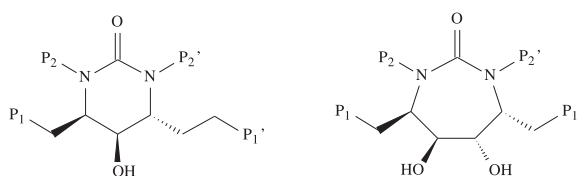
3

Andmekomplekt ja metoodika

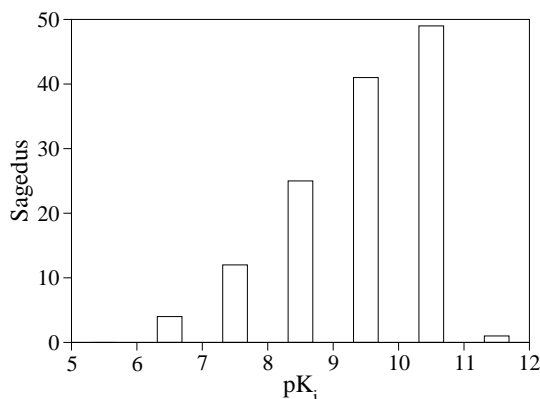
3.1 Andmekomplekt

Kasutatud andmekomplekt koosnes 135-st HIV-1 proteaasi inhibiitorist. Struktuurid ja aktiivsused koguti kirjandusest[43, 44, 45, 46, 50, 57, 58, 59]. Kõik ühendid põhinesid kahel baasstruktuuril, kuue- ja seitsmelülilisel tsüklilisel uureal, mis on kujutatud joonisel 3.1. Molekulide suurus varieerus 45-st aatomist 134-ni ja pK_i 6.01-st 11.01. Nagu joonisel 3.2 kujutatud histogramm näitab olid andmekomplektis aktiivsused ebahühtlaselt jaotunud, suurema pK_i -ga ühendid olid ülekaalus. Kõik struktuurid ja nende aktiivsused on toodud lisas A.

Mudelite valideerimiseks jagati andmekomplekt treening- ja testkomplektiks. Treeningkomplekt koostati eesmärgiga saavutamaks võimalikult suurt struktuuride variatsiooni, sinna paigutati ühendeid igast allikast ja olemasolevate võimalike asendajatega. Ühtlasi püüti tagada ka treeningseti aktiivsuste ühtlast jaotust. Treeningseti koostati käsitsi, selle suuruseks kujunes 87 ühendit, umbes 2/3 kogu komplektist. Mudelite arendamisel ilmnenud hälbijad eemaldati treeningsetist ja sektsioonis 4.2 on neid ka põhjalikumalt analüüsitud. Hälbijate eemaldamine, kui tegemist on ülejäänud komplektist tugevalt erineva struktuuriga või ka näiteks mõõteveaga, parandab saadava mudeli kvaliteeti.



Joonis 3.1: Baasstruktuurid



Joonis 3.2: Histogramm

3.2 Metoodika

Töös kasutatud struktuuride 3D geomeetriad leiti konformatsiooniotsingust, mis sooritati MacroModel-ga[60] (versioon 8.5), kasutades selle Monte Carlo Multiple Minimum (MCM) meetodit ja MMFFs jõuvälja. Iga struktuuri jaoks valiti madalaima tekkeentalpiaga konformeer. Struktuurid optimeeriti täiendavalt MOPAC-ga[61] (versioon 7.05), kasutades AM1 parametrisatsiooni[62]. Pärast optimeerimist teostati kvantkeemilised arvutused kasutades võtmesõnu VECTORS BONDS PI POLAR ENPART. Järgnes molekulaardeskriptorite arvutamine CODESSA PRO-ga[63]. Kokku arvutati üle 700 konstitutsioonilise, topoloogilise, geomeetrilise, poolempiirilise ja laengujaotust kirjeldava deskriptori. Peale nende, tuginedes selle varasemale edukale rakendamisele[40, 41, 42, 47, 49], arvutati programmiga KowWin[64] *logP*.

Arvutatud deskriptorid ei pääsenud kõik edasi mudelite arendamisesse. Paljudele ei olnud ühe või mitme struktuuri jaoks väärtust arvutatud ning nad eemaldati deskriptorite komplektist. Samamoodi eemaldati väga väikese variatsiooniga deskriptorid. Kui selliste deskriptorite eemaldamise teostab CODESSA PRO automaatselt, siis edasine deskriptorikomplekti piiramine teostati käsitsi. Eemaldati sellised kus deskriptori-omaduse graafik peegeldas väga ebahühtlast jaotust, näiteks kui käputäis ühendeid olid teistest suurel kaugusel. Lisaks eemaldati enamus konstitutsioonilisi ja miinimum-maksimum energia tüüpi deskriptoreid, põhjuseks vähene informatiivsus. Viimased kirjeldavad energiat mingi sideme, aatomi või vastasmõju kohta kuid mitme sobiva kandidaadi korral ei ole võimalik öelda millist neist deskriptor kirjeldab. Kõikidest arvutatud deskriptoritest eemaldati enne mudeli arendamiseni jõudmist rohkem kui pooled.

Deskriptor-omadus graafikud tõid välja asjaolu, et seos deskriptori ja omaduse vahel on sageli mittelineaarne. Kuna eesmärgiks oli kasutada lineaarseid meetodeid, rakendati deskriptoritele mittelineaarsed matemaatilised transformatsioonid pöördväärtus ja naturaallogaritm. Deskriptoreid eelnevalt ei skaleeritud, sellest tulenevalt jäid siin välja sel-

lised kus deskriptori väärtus mõne struktuuri jaoks oli null. Logaritmi puhul jäid välja lisaks veel need, kus oli nii positiivseid kui negatiivseid väärtusi, kui kõik väärtused olid negatiivsed, võeti logaritmi absoluutväärtusest.

Mudeleid arendati CODESSA PRO-ga kasutades selle parima multilineaarse regressiooni (Best MultiLinear Regression, BMLR) meetodit. Selle käigus arvutatakse kõigepealt kõik võimalikud kaheparameetriselised korrelatsioonid, seejärel hakatakse neljasajale parimale neist lisama uusi deskriptoreid. Otsing kestab, kuni järjekordse deskriptori lisamisega mudeli statistilised näitajad paranevad vähem kui kasutaja poolt määratud väärtused. Algoritmi on täpsemalt kirjeldatud CODESSA PRO manuaalis[65].

4

Tulemused ja analüüs

4.1 Mudel

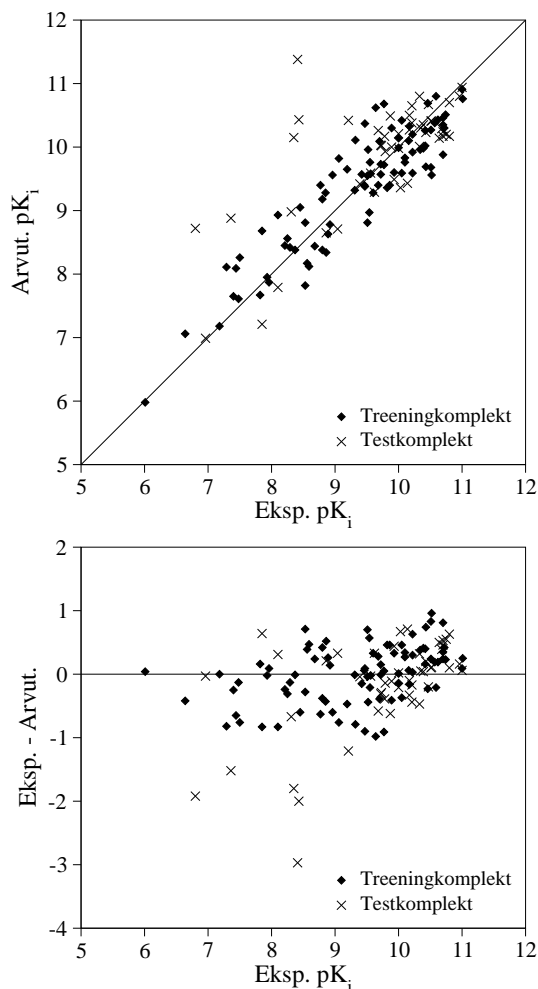
Statistilised parameetrid mida arvestati parima mudeli valimisel olid korrelatsioonikoefitsient (R^2), ristvalideeritud korrelatsioonikoefitsient (R_{cv}^2), testkomplekti korrelatsioonikoefitsient (R_{test}^2), Fisheri kriteerium F ja standardhälve s . Jälgiti ka ennustusvigade jaotust kuna nende ebahütlane jaotus võib viidata süstemaatilisele veale. Nende näitajate põhjal valiti välja tabelis 4.1 lahtikirjutatud mudel, kus esimene veerg B on regressioonikordaja, teine veerg t on t-testi väärtus, kolmandas on toodud tekstis kasutatav lühend deskriptori nimest, mis täispikkuses on viimases veerus.

Tabel 4.1: Mudel

B	t	Lühend	Deskriptor
21.158	29.716	-	Vabaliige
-372.369	-12.65	$1/{}^2SIC$	1/Structural Information Content (order 2)
-0.0829	-12.395	$1/{}^HPCSA$	1/Partial Charged Surface Area for atom H
-0.00691	-8.081	$DPSA1$	DPSA1 Difference in CPSAs (PPSA1-PNSA1) (Zefirov PC)
64.177	5.095	$1/DPSA3$	1/DPSA3 Difference in CPSAs (PPSA3-PNSA3) (Zefirov PC)

Mudeli statistilised parameetrid on järgmised: $R^2 = 0.825$, $R_{cv}^2 = 0.806$, $R_{test}^2 = 0.49$ (R_{test}^2 kust hälbijad on eemaldatud = 0.786), $F = 96.87$, ja $s = 0.465$. Mudel on graafiliselt kujutatud joonisel 4.1 ja kasutatud deskriptorite korrelatsioonid pK_i -ga joonisel 4.3.

Esimene ja t-testi järgi suurima kaaluga deskriptor mudelis on $1/{}^2SIC$. Topoloogilisi deskriptoreid on varemgi tsükliliste uureate modelleerimisel kasutatud[42, 49, 54], nad põhinevad graafiteoorial ja annavad informatsiooni molekuli suuruse, kuju ja hargnevuse kohta. Praegusel juhul on 2SIC tugevasti seotud molekuli suurust kirjeldavate deskriptoritega, näiteks molekuli ruumala ($R^2 = 0.9$) ja aatomite arv molekulis ($R^2 = 0.93$). Nii tugev seos tuleb tõenäoliselt kasutatud ühendite suhteliselt sarnasusest, kui baasstruktuur



Joonis 4.1: Mudel

püsib muutumatuna siis molekuli topoloogiat kirjeldav osa deskriptoris omandab väiksema osatähtsuse. Seda kinnitab ka tõik, et kuigi topoloogilised deskriptorid eristavad molekule nende struktuuri põhjal, siis ei 2SIC ega mõni teine deskriptor ei jaganud struktuure kuue- ja seitsmelülilisteks ureateks. Keskse ahela suurus ei ole nii suure tähtsusega kui muutused asendajates. 2SIC -i ilmumine on heas kooskõlas varasemate mudelitega, mis näitavad inhibiitori suuruse ja kuju tähtsust. Tegelikult on isegi sama deskriptor varasemates mudelites esinenud[54] (seksioonis 2.5 mudelid 2.13 ja 2.14), küll oma originaalmitte pöördfunktsioonina. Põhjus miks pöördväärtus praegu mudelisse paremini sobib on tõenäoliselt selles, et praegu kasutatav suurem andmekomplekt avaldab 2SIC -i mittelineaarse olemuse - deskriptori väärtused kasvavad kiiremini kui pK_i väärtused ning seetõttu on pöördväärtuse kasutamine lineaarses mudelis otstarbekam.

Ülejäänud kolm deskriptorit mudelis on laetud osapindalasisid (CPSA) kirjeldavad deskriptorid[66, 67]. Esimene neist on $1/HPCSA$. $HPCSA$ on kahe teise deskriptori, $HCSA$ (vesinikuaatomite laenguga pindala, *charged surface area for atom H*) ja $TMSA$ (moleku-

li kogupindala, *total molecular surface area*) jagatis, mis võimaldab $1/H$ PCSA kirjutada kui $TMSA/H$ CSA. Nagu ka $1/2$ SIC, on $1/H$ PCSA uuritava omadusega negatiivselt korreleeritud ($R = -0.61$). Selle deskriptori eemaldamine mudelist tõi asemele $\ln(H$ PCSA), näidates nii vesinike pindala olulisust, kui ka mittelineaarsete seoste osakaalu.

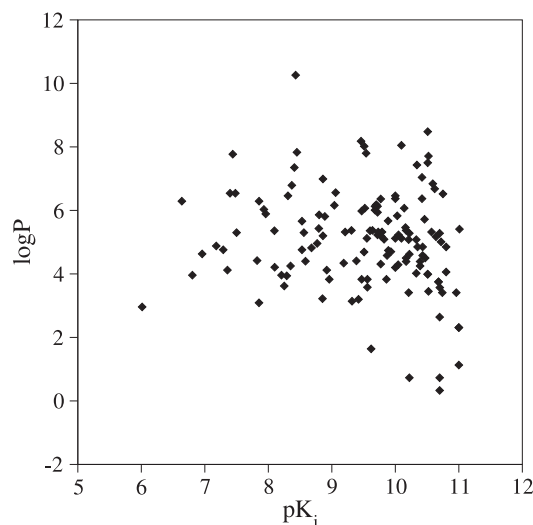
Järelejäänud kaks deskriptorit mudelis on $DPSA1$ ja selle osalaengutega kaalutud analoog $1/DPSA3$. Kumbki neist ei ole otseselt seotud uuritava omadusega ($DPSA1$ ja $1/DPSA3$ korrelatsioonikoeffitsiendid omadusega vastavalt $R^2 = 0.1$ ja $R^2 = 0.2$) mis teeb nende interpreteerimise raskeks. Mõlemad on kaudselt seotud molekuli võimega luua vesiniksidet, seda kinnitavad nende korrelatsioonid teiste seda tüüpi deskriptoritega. $DPSA1$ on seotud vesiniksideme doonorite arvuga (*count of H-donor sites*) ($R^2 = 0.82$) ja $DPSA3$ vesiniksideme aktseptoraatomite pindalaga kaalutud osalaenguga (*area-weighted surface charge of hydrogen bonding acceptor atoms*) $HACA - 2$ ($R^2 = 0.96$). On teada, et inhibiitor seotakse ensüümi aktiivtsentrisse vesiniksideme abil, kuid on huvitav, et $DPSA$ -d töötavad paremini kui otseselt vesiniksideme tekkevõimet kirjeldavad deskriptorid. See käib mõnevõrra ka $1/H$ PCSA kohta sest mitte kõik vesinikud molekulis ei oma kindlat funktsiooni inhibiitoriga seostumises. Mudelite arendamise käigus oli sama iseärasust näha ka teiste potentsiaalsete mudelite puhul - spetsiifiliste vesiniksidet kirjeldavate deskriptorite asemel ilmusid mudelitesse üldisemad, laiemalt laengujaotust kirjeldavad deskriptorid.

Hoolimata $\log P$ varasemast edukast kasutamisest uureate aktiivsuse analüüsis[40, 41, 49, 42, 47] ei omanud see käesolevas töös erilist tähtsust. Korrelatsiooni omadusega põhimõtteliselt ei ole ($R^2 = 0.025$). Samas annab $\log P$ häid korrelatsioone vaadeldes väiksemaid ja sarnaseid alamkomplekte. Näiteks 30 ühendit, 88-118, kus enamasti varieeritakse lihtsaid süsivesinikasendajaid, on korrelatsioon $\log P$ ja pK_i vahel märkimisväärne, $R^2 = 0.64$.

4.2 Mudeli rakenduspiirid

Mudelite arendamise käigus leiti kokku sadu mudeleid ja selgus et enamasti on neil raskesti teatud ühendite aktiivsuse hindamisel (ühendid 38, 39, 59, 124 ja 131 lisas A). Nende pK_i väärtused on alati ülehinnatud, nii et joonisel 4.1 eristuvad nad küllalt selgesti. Kõikidel neil ühenditel on omad eripärad mis teevad nad andmekomplektis mõnevõrra unikaalseks. Suhtelise harulduse tõttu on nende mõju mudeli arendamisele väike ning see ei luba nende inhibitsioonikonstante täpselt hinnata.

Ühenditel 38 ja 39 in väikseim võimalik P_2 asendaja - H. Samas on nad nõrgad inhibiitorid, pK_i väärtused vastavalt 7.356 ja 6.804. Kummalgi on andmekomplektis regioisomeer, ühend kus P_2 ja P'_2 asendajad on vahetuses ja vesinik on seega P'_2 asendis. Isomeerid ühenditele 38 ja 39 on vastavalt 37 ja 32, mis on märksa paremad inhibiitorid, pK_i -d vas-



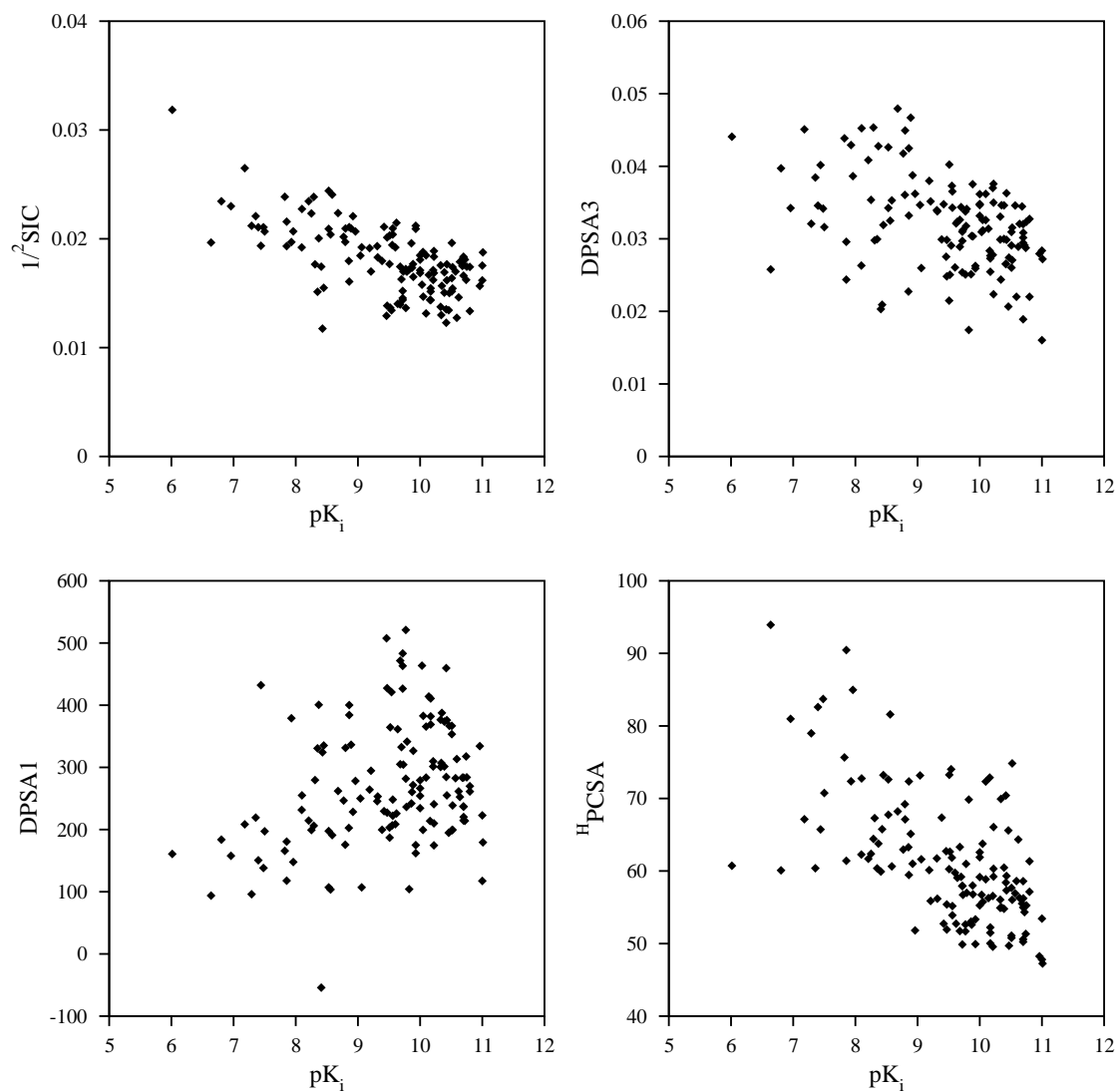
Joonis 4.2: logP vs. pK_i

tavalt 8.92 ja 8.207. Deskriptorite analüüs näitas, et isomeeridele arvatud deskriptorite väärtused on sarnased, mis tähendab, et deskriptorid ei suuda selliseid struktuure piisavalt eristada ja tulemuseks on sarnased ennustatud inhibeerimiskonstandid.

Järgmine hälbija on ühend number 59. Kõige sarnasem sellele on ühend nr. 56, nende kahe erinevuseks on et 59-s on kaks CH_3 rühma asendatud CF_3 -ga. Kuna fluori aatom on suurem on vastavalt suurem ka molekul, molekuli maht kasvab 6.7% $585.85\text{-lt } 625.13\text{\AA}^3$ -ni ja on oletatud, et inhibeerimisvõime langus on tingitud steerilistest takistustest[59]. Kuid kuna anmekomplektis on suuremaid molekule, mis on samas paremad inhibiitorid, tunduvad muutused laengujaotuses tõenäolisematena. Tugevalt elektronegatiivsed rühmad vähendavad kõrval asuvate vesiniksideme aktseptorite vabade elektronpaaride kättesaadavust. Seda kinnitavad ka MOPAC-i arvutused, kust ilmneb elektrontiheduse vähenemine CF_3 -le lähedastel aatomitel. Vastavalt $DPSA1$ -le on 59 hälbija, ainus ühend millel negatiivselt laetud pindala on ülekaalus.

Ühendid 124 ja 131 on silmapaistvad vastavalt oma väikseimate ja suurimate P_1/P'_1 asendajate poolest, 124-l on selleks etüül ja 131-l 4-(fenüülmetoksü)bensüül. Mõlema struktuuri aktiivsused on ülehinnatud, 124 puhul 10.154 vs. 8.35 ja 131-l 10.433 vs. 8.43. Selgitust nende halvemale inhibeerimisvõimele ei ole kerge leida, seda ei pakkunud välja ka ainete sünteesijad[58]. Samas, kuna P_1/P'_1 asendajate hulgas on variatsioon väike, ei ole andmekomplekt nende suhtes eriti esinduslik.

Hälbijatega toimetulekuks on üks võimalus andmekomplekti laiendamine. Kõik viis kirjeldatud ühendit on oma asendajate või funktsionaalrühmade poolest andmekomplek-



Joonis 4.3: Deskriptorid vs. pK_i

tis unikaalsed ning seega nende mõju mudeli trenimisel on väike. Niikuinii on uureate arendamisel peamiselt keskendunud P_2/P'_2 rühmade muutmisele, seega on ka praeguses andmekomplektis P_1/P'_1 rühmade variatsioon väike, suur enamus asendajatest nendes positsioonides on bensüül. Samas võib laiendamisel ilmned hoopis, et variatsioonid korraga nii P_1/P'_1 kui ka P_2/P'_2 rühmades ei olegi koos modelleeritavad. Seda peab selgitama analüüs veelgi mitmekesisema andmekomplektiga.

5

Kokkuvõte

Käesolevas töös modelleeriti tsüklilisel ureal põhinevate HIV-1 proteaasi inhibiitorite bioloogilisi aktiivsusi QSAR-i meetodiga kasutades teoreetilisi molekulaardeskriptoreid. Töötati suurema ja mitmekesisema andmekomplektiga kui varemalt ja sellest tulenevalt muutus mõnevõrra probleemi rõhuasetus. Tugevamini tõusid esile mittelineaarsed seosed omaduse ja deskriptorite vahel ning nende mõju hakkas segama lineaarsete regressiooni-meetodite rakendamist. Nende ulatust märgib asjaolu, et valitud mudelis kolm deskriptorit neljast on mittelineaarselt transformeeritud. Selgemini eristusid ka teatud struktuurid mille P-asendajad või üksikud funktsionaalrühmad olid andmekomplektis haruldased ning vähene esindatus ei lubanud nende kohta teha täpseid ennustusi. Väga väikese andmekomplekti korral on ühe struktuuri suhteline mõju suurem, ning ka üksik hälbija võib mudelit tugevalt kallutada.

Viis hälbijat välja arvatud, allus andmekomplekt analüüsile üldiselt hästi. Leitud mudel arvutab testkomplekti ühendite aktiivsusi sarnase kvaliteediga nagu treeningkomplekti omi, R^2 testkomplektile 0.768 ja kogu andmekomplektile 0.824. Mudel seob ureate inhibeerimisomadused küllaltki lihtsate struktuursete karakteristikutega, nagu varasematest töödeski tuntud molekuli suurus ja kuju, ning molekuli pinnaomadused ja vesinike osapindala.

6

Conclusion

A QSAR analysis using theoretical molecular descriptors was used for modelling the activities of cyclic urea-based HIV-1 protease inhibitors. Large and multifarious data set was used and that revealed some significant new aspects governing biological activity prediction. Nonlinear descriptor-property relationships became overwhelming and that reduced the applicability of linear regression methods. The extent of nonlinear relationships in the data set is illustrated by the fact, that nonlinear transformations have been applied to three out of four descriptors in the selected model. Certain strong outliers also emerged that had rare P-substituents or functional groups, and due to their uniqueness their antiviral activities were inaccurately predicted.

Despite the five identified outliers, the data set subjected well to the analysis. The developed model is capable of calculating activities for the compounds in testing set with the quality comparable to the compounds in training set. Without the outliers, R^2_{test} becomes 0.768 and the R^2 for the whole set 0.824. The model associates the antiviral activity of the cyclic ureas with rather simple structural characteristics, such as the size and the shape of the molecule, charged surface area of the hydrogen atoms, and overall charge distribution.

Kirjandus

- [1] Marx, J. L., New disease baffles medical community., *Science*, **1982**, *217*, 618–21.
- [2] Gao, F.; Bailes, E.; Robertson, D. L.; Chen, Y.; Rodenburg, C. M.; Michael, S. F.; Cummins, L. B.; Arthur, L. Ö.; Peeters, M.; Shaw, G. M.; Sharp, P. M.; Hahn, B. H., Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*., *Nature*, **1999**, *397*, 436–441.
- [3] Keele, B. F.; Heuverswyn, F. V.; Li, Y.; Bailes, E.; Takehisa, J.; Santiago, M. L.; Bibollet-Ruche, F.; Chen, Y.; Wain, L. V.; Liegeois, F.; Loul, S.; Ngole, E. M.; Bienvenue, Y.; Delaporte, E.; Brookfield, J. F. Y.; Sharp, P. M.; Shaw, G. M.; Peeters, M.; Hahn, B. H., Chimpanzee reservoirs of pandemic and nonpandemic HIV-1., *Science*, **2006**, *313*, 523–6.
- [4] Robertson, D.; Anderson, J.; Bradac, J.; Carr, J.; Foley, B.; Funkhouser, R.; Gao, F.; Hahn, B.; Kalish, M.; Kuiken, C.; Learn, G.; Leitner, T.; McCutchan, F.; Osmanov, S.; Peeters, M.; Pieniazek, D.; Salminen, M.; Sharp, P.; Wolinsky, S.; Korber, B., A Reference Guide to HIV-1 Classification.
- [5] Reeves, J. D.; Doms, R. W., Human immunodeficiency virus type 2., *J. Gen. Virol.*, **2002**, *83*, 1253–65.
- [6] UNAIDS, 2006 report on the global AIDS epidemic: Executive summary, **2006**.
- [7] Preston, B. J. L. L. A., B. D. and Poiesz, Fidelity of HIV-1 reverse transcriptase, *Science*, **1988**, *242*, 1168–1171.
- [8] Sierra, S.; Kupfer, B.; Kaiser, R., Basics of the virology of HIV-1 and its replication, *Journal of Clinical Virology*, **2005**, *34*, 233–244.
- [9] Pereira, C. F.; Paridaen, J. T., Anti-HIV Drug Development - An Overview, *Curr. Pharm. Des.*, **2004**, *10*, 4005–4037(33).
- [10] De Clercq, E., Perspectives of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection, *Il Farmaco*, **1999**, *54*, 26–45.

- [11] Eron, J. J., HIV-1 protease inhibitors., *Clin. Infect. Dis.*, **2000**, *30 Suppl 2*, S160–S170.
- [12] De Clercq, E., New developments in anti-HIV chemotherapy, *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, **2002**, *1587*, 258–275.
- [13] Brik, A.; Wong, C.-H., HIV-1 protease: mechanism and drug discovery., *Org. Biomol. Chem.*, **2003**, *1*, 5–14.
- [14] Burlet, S.; Pietrancosta, N.; Laras, Y.; Garino, C.; Quelever, G.; Kraus, J.-L., Prospects for the Resistance to HIV Protease Inhibitors: Current Drug Design Approaches and Perspectives., *Curr. Pharm. Des.*, **2005**, *11*, 3077–3090.
- [15] Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B.; Wlodawer, A., Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution, *Science*, **1989**, *246*, 1149–1152.
- [16] Gustchina, A.; Weber, I. T., Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap, *FEBS Lett.*, **1990**, *269*, 269–272.
- [17] Tomasselli, A. G.; Hui, J.Ö.; Sawyer, T. K.; Staples, D. J.; Bannow, C.; Reardon, I. M.; Howe, W. J.; DeCamp, D. L.; Craik, C.Š.; Heinrikson, R. L., Specificity and inhibition of proteases from human immunodeficiency viruses 1 and 2, *J. Biol. Chem.*, **1990**, *265*, 14675–14683.
- [18] Tomasselli, A. G.; Sarcich, J. L.; Barrett, L. J.; Reardon, I. M.; Howe, W. J.; Evans, D. B.; Sharma, S. K.; Heinrikson, R. L., Human immunodeficiency virus type-1 reverse transcriptase and ribonuclease H as substrates of the viral protease, *Protein Sci.*, **1993**, *2*, 2167–2176.
- [19] Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; Davies, D. R., Binding of a Reduced Peptide Inhibitor to the Aspartic Proteinase from *Rhizopus chinensis*: Implications for a Mechanism of Action, *Proc. Natl. Acad. Sci.*, **1987**, *84*, 7009–7013.
- [20] Pearl, L. H., The catalytic mechanism of aspartic proteinases, *FEBS Lett.*, **1987**, *214*, 8–12.
- [21] Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, S. B. H., Jensa and Kent; Wlodawer, A., Structure at 2.5-Å resolution of chemically synthesized Human Immunodeficiency Virus Type 1 protease complexed with a hydroxyethylene-based inhibitor, *Biochemistry*, **1991**, *30*, 1600–1609.

- [22] Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I.Š., Active Human Immunodeficiency Virus Protease is Required for Viral Infectivity, *Proc. Natl. Acad. Sci.*, **1988**, *85*, 4686–4690.
- [23] Seelmeier, S.; Schmidt, H.; Turk, V.; Helm, K. V., Human Immunodeficiency Virus Has an Aspartic-Type Protease that Can Be Inhibited by Pepstatin A, *Proc. Natl. Acad. Sci.*, **1988**, *85*, 6612–6616.
- [24] Thaisrivongs, S.; Strohbach, J. W., Structure-based discovery of tipranavir disodium (PNU-140690E): A potent, orally bioavailable, nonpeptidic HIV protease inhibitor, *Peptide Science*, **1999**, *51*, 51–58.
- [25] Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bachelier, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N., Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors, *Science*, **1994**, *263*, 380–384.
- [26] Ungwitayatorn, J.; Samee, W.; Pimthon, J., 3D-QSAR studies on chromone derivatives as HIV-1 protease inhibitors, *J. Mol. Struct.*, **2004**, *689*, 99–106.
- [27] Svicher, V.; Ceccherini-Silberstein, F.; Erba, F.; Santoro, M.; Gori, C.; Bellocchi, M. C.; Giannella, S.; Trotta, M. P.; Monforte, A. d.; Antinori, A.; Perno, C. F., Novel Human Immunodeficiency Virus Type 1 Protease Mutations Potentially Involved in Resistance to Protease Inhibitors, *Antimicrob. Agents Chemother.*, **2005**, *49*, 2015–2025.
- [28] Turner, S. R., HIV Protease Inhibitors - The Next Generation, *Current Medicinal Chemistry -Anti-Infective Agents*, **2002**, *1*, 141–162(22).
- [29] Rusconi, S.; Vigano, O., New HIV protease inhibitors for drug-resistant viruses, *Therapy*, **2006**, *3*, 79–88.
- [30] Eder, J.; Hommel, U.; Cumin, F.; Martoglio, B.; Gerhartz, B., Aspartic Proteases in Drug Discovery, *Curr. Pharm. Des.*, **2007**, *13*, 271–285.
- [31] Karelson, M., *Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons, **2000**.
- [32] Milac, A.-L.; Avram, S.; Petrescu, A.-J., Evaluation of a neural networks QSAR method based on ligand representation using substituent descriptors: Application to HIV-1 protease inhibitors, *J. Mol. Graphics Modell.*, **2006**, *25*, 37–45.

- [33] Fernandez, M.; Caballero, J., Modeling of activity of cyclic urea HIV-1 protease inhibitors using regularized-artificial neural networks, *Bioorg. Med. Chem.*, **2006**, *14*, 280–294.
- [34] Debnath, A. K., Application of 3D-QSAR Techniques in Anti-HIV-1 Drug Design - An Overview., *Curr. Pharm. Des.*, **2005**, *11*, 3091–3110.
- [35] Debnath, A. K., Comparative Molecular Field Analysis (CoMFA) of a Series of Symmetrical Bis-Benzamide Cyclic Urea Derivatives as HIV-1 Protease Inhibitors, *J. Chem. Inf. Model.*, **1998**, *38*, 761–767.
- [36] Nair, A. C.; Jayatilleke, P.; Wang, X.; Miertus, S.; Welsh, W. J., Computational Studies on Tetrahydropyrimidine-2-one HIV-1 Protease Inhibitors: Improving Three-Dimensional Quantitative Structure-Activity Relationship Comparative Molecular Field Analysis Models by Inclusion of Calculated Inhibitor- and Receptor-Based Properties, *J. Med. Chem.*, **2002**, *45*, 973–983.
- [37] Debnath, A. K., Three-Dimensional Quantitative Structure-Activity Relationship Study on Cyclic Urea Derivatives as HIV-1 Protease Inhibitors: Application of Comparative Molecular Field Analysis, *J. Med. Chem.*, **1999**, *42*, 249–259.
- [38] Tervo, A. J.; Nyronen, T. H.; Ronkko, T.; Poso, A., Comparing the Quality and Predictiveness between 3D QSAR Models Obtained from Manual and Automated Alignment, *J. Chem. Inf. Model.*, **2004**, *44*, 807–816.
- [39] Senese, C. L.; Hopfinger, A. J., A Simple Clustering Technique To Improve QSAR Model Selection and Predictivity: Application to a Receptor Independent 4D-QSAR Analysis of Cyclic Urea Derived Inhibitors of HIV-1 Protease, *J. Chem. Inf. Model.*, **2003**, *43*, 2180–2193.
- [40] Wilkerson, W. W.; Akamike, E.; Cheatham, W. W.; Hollis, A. Y.; Collins, R. D.; DeLuca, I.; Lam, P. Y.Š.; Ru, Y., HIV Protease Inhibitory Bis-benzamide Cyclic Ureas: A Quantitative Structure-Activity Relationship Analysis, *J. Med. Chem.*, **1996**, *39*, 4299–4312.
- [41] Wilkerson, W. W.; Dax, S.; Cheatham, W. W., Nonsymmetrically Substituted Cyclic Urea HIV Protease Inhibitors, *J. Med. Chem.*, **1997**, *40*, 4079–4088.
- [42] Garg, R.; Bhatarai, B., A mechanistic study of 3-aminoindazole cyclic urea HIV-1 protease inhibitors using comparative QSAR, *Bioorg. Med. Chem.*, **2004**, *12*, 5819–5831.

- [43] Rodgers, J. D.; Johnson, B. L.; Haisheng, W.; Erickson-Viitanen, S.; Klabe, R. M.; Bacheler, L.; Cordova, B. C.; Chang, C.-H., Potent cyclic urea HIV protease inhibitors with 3-aminoindazole P2/P2' groups, *Bioorg. Med. Chem. Lett.*, **1998**, *8*, 715–720.
- [44] Kaltenbach, R. F.; Klabe, R. M.; Cordova, B. C.; Seitz, S. P., Increased antiviral activity of cyclic urea HIV protease inhibitors by modifying the P1/P1' substituents, *Bioorg. Med. Chem. Lett.*, **1999**, *9*, 2259–2262.
- [45] Rodgers, J. D.; Lam, P. Y.Š.; Johnson, B. L.; Wang, H.; Ko, S.Š.; Seitz, S. P.; Trainor, G. L.; Anderson, P.Š.; Klabe, R. M.; Bacheler, L. T.; Cordova, B.; Garber, S.; Reid, C.; Wright, M. R.; Chang, C.-H.; Erickson-Viitanen, S., Design and selection of DMP 850 and DMP 851: the next generation of cyclic urea HIV protease inhibitors, *Chemistry & Biology*, **1998**, *5*, 597–608.
- [46] Kaltenbach, R. F.; Patel, M.; Waltermire, R. E.; Harris, G. D.; Stone, B. R. P.; Klabe, R. M.; Garber, S.; Bacheler, L. T.; Cordova, B. C.; Logue, K.; Wright, M. R.; Erickson-Viitanen, S.; Trainor, G. L., Synthesis, antiviral activity and pharmacokinetics of P1/P1' substituted 3-aminoindazole cyclic urea HIV protease inhibitors, *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 605–608.
- [47] Garg, R.; Patel, D., Hydrophobicity in the design of P2/P2' tetrahydropyrimidinone HIV protease inhibitors, *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 3767–3770.
- [48] De Lucca, G. V.; Liang, J.; Aldrich, P. E.; Calabrese, J.; Cordova, B.; Klabe, R. M.; Rayner, M. M.; Chang, C. H., Design, Synthesis, and Evaluation of Tetrahydropyrimidinones as an Example of a General Approach to Nonpeptide HIV Protease Inhibitors, *J. Med. Chem.*, **1997**, *40*, 1707–1719.
- [49] Gayathri, P.; Pande, V.; Sivakumar, R.; Gupta, S. P., A quantitative structure-activity relationship study on some HIV-1 protease inhibitors using molecular connectivity index, *Bioorg. Med. Chem.*, **2001**, *9*, 3059–3063.
- [50] De Lucca, G. V.; Liang, J.; De Lucca, I., Stereospecific Synthesis, Structure-Activity Relationship, and Oral Bioavailability of Tetrahydropyrimidin-2-one HIV Protease Inhibitors, *J. Med. Chem.*, **1999**, *42*, 135–152.
- [51] Gupta, S. P.; Babu, M.Š.; Garg, R.; Sowmya, S., Quantitative structure-activity relationship studies on cyclic urea-based HIV protease inhibitors, *J. Enzym. Inhib.*, **1998**, *13*, 399–407.

- [52] Gupta, S. P.; Babu, M.Š.; Sowmya, S., A quantitative structure-Activity relationship study on some sulfolanones and arylthiomethanes acting as HIV-1 protease inhibitors, *Bioorg. Med. Chem.*, **1998**, *6*, 2185–2192.
- [53] Jadhav, P. K.; Woerner, F. J.; Lam, P. Y.Š.; Hodge, C. N.; Eyermann, C. J.; Man, H. W.; Daneker, W. F.; Bacheler, L. T.; Rayner, M. M.; Meek, J. L.; Erickson-Viitanen, S.; Jackson, D. A.; Calabrese, J. C.; Schadt, M.; Chang, C. H., Nonpeptide Cyclic Cyanoguanidines as HIV-1 Protease Inhibitors: Synthesis, Structure-Activity Relationships, and X-ray Crystal Structure Studies, *J. Med. Chem.*, **1998**, *41*, 1446–1455.
- [54] Katritzky, A. R.; Oliferenko, A.; Lomaka, A.; Karelson, M., Six-membered cyclic ureas as HIV-1 protease inhibitors: A QSAR study based on CODESSA PRO approach, *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 3453–3457.
- [55] Maran, U.; Sild, S.; Kahn, I.; Takkis, K., Mining of the chemical information in GRID environment, *Future Generation Computer Systems*, **2007**, *23*, 76–83.
- [56] Sild, S.; Maran, U.; Lomaka, A.; Karelson, M., Open Computing Grid for Molecular Science and Engineering, *J. Chem. Inf. Model.*, **2006**, *46*, 953–959.
- [57] Lam, P. Y.Š.; Ru, Y.; Jadhav, P. K.; Aldrich, P. E.; DeLucca, G. V.; Eyermann, C. J.; Chang, C. H.; Emmett, G.; Holler, E. R.; Daneker, W. F.; Li, L.; Confalone, P. N.; McHugh, R. J.; Han, Q.; Li, R.; Markwalder, J. A.; Seitz, S. P.; Sharpe, T. R.; Bacheler, L. T.; Rayner, M. M.; Klabe, R. M.; Shum, L.; Winslow, D. L.; Kornhauser, D. M.; Jackson, D. A.; Erickson-Viitanen, S.; Hodge, C. N., Cyclic HIV Protease Inhibitors: Synthesis, Conformational Analysis, P2/P2' Structure-Activity Relationship, and Molecular Recognition of Cyclic Ureas, *J. Med. Chem.*, **1996**, *39*, 3514–3525.
- [58] Patel, M.; Kaltenbach, R. F.; Nugiel, D. A.; McHugh, R. J.; Jadhav, P. K.; Bacheler, L. T.; Cordova, B. C.; Klabe, R. M.; Erickson-Viitanen, S.; Garber, S.; Reid, C.; Seitz, S. P., The synthesis of symmetrical and unsymmetrical P1/P1' cyclic ureas as HIV protease inhibitors, *Bioorg. Med. Chem. Lett.*, **1998**, *8*, 1077–1082.
- [59] Han, Q.; Chang, C. H.; Li, R.; Ru, Y.; Jadhav, P. K.; Lam, P. Y.Š., Cyclic HIV Protease Inhibitors: Design and Synthesis of Orally Bioavailable, Pyrazole P2/P2' Cyclic Ureas with Improved Potency, *J. Med. Chem.*, **1998**, *41*, 2019–2028.
- [60] Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C., Macromodel-an integrated software

system for modeling organic and bioorganic molecules using molecular mechanics, *J. Comput. Chem.*, **1990**, *11*, 440–467.

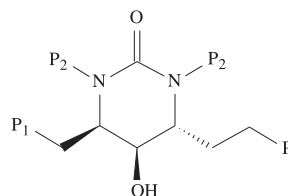
- [61] Stewart, J. J. P., MOPAC Program Package, **1989**.
- [62] Dewar, M. J.Š.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, *J. Am. Chem. Soc.*, **1985**, *107*, 3902–3909.
- [63] <http://www.codessa-pro.com>.
- [64] Meylan, W. M.; Howard, P. H., Atom/fragment contribution method for estimating octanol-water partition coefficients., *J. Pharm. Sci.*, **1995**, *84*, 83–92.
- [65] University of Florida, *CODESSA PRO User's Manual*, **2005**.
- [66] Stanton, D. T.; Jurs, P. C., Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies, *Anal. Chem.*, **1990**, *62*, 2323–2329.
- [67] Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G., Computer-assisted prediction of normal boiling points of pyrans and pyrroles, *J. Chem. Inf. Model.*, **1992**, *32*, 306–316.

Lisa A

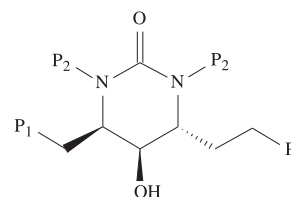
Andmekomplekt

Eksperimentaalsed ja arvutatud pK_i -d on märgitud vastavalt Eksp. ja Arvut. Kuulumine treeningkomplekti on märgitud t-ga.

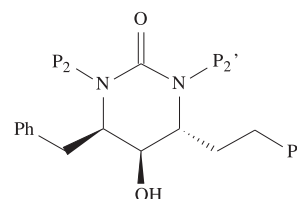
	P_1/P_2		Eksp.	Arvut.
1	fenüül/3-tsüanobensüül	t	7.959	7.869
2	4-fluorofenüül/3-tsüanobensüül		7.854	7.214
3	fenüül/3-tsüano-4-fluorobensüül	t	6.638	7.056
4	fenüül/3-atsetüülbensüül		9.04	8.713
5	fenüül/3-hüdroksümetüülbensüül	t	9.309	9.321
6	fenüül/3-karboksübensüül	t	9.06	9.817
7	fenüül/3-(karboksamido) bensüül	t	10.045	9.587
8	4-fluorofenüül/ 3-(karboksamido) bensüül	t	9.508	9.550
9	3,4-difluorofenüül/ 3-(karboksamido) bensüül	t	9.823	9.363
10	fenüül/ 3-(karboksamido-4-fluorobensüül	t	8.853	9.281
11	fenüül/ 3-(karboksamido oksiiim)bensüül	t	10.698	10.458
12	4-fluorofenüül/ 3-(karboksamido oksiiim)bensüül	t	10.698	10.349
13	3,4-difluorofenüül/ 3-(karboksamido oksiiim)bensüül	t	11	10.915
14	fenüül/3-(karboksamido oksiim)-4-fluorobensüül	t	10.221	10.195
15	fenüül/3-aminobensüül	t	8.769	9.396
16	fenüül/3-amino-4-fluorobensüül	t	9.602	9.277



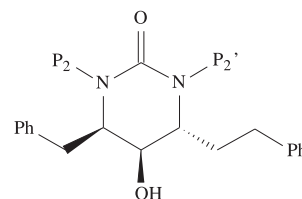
P ₁ /P ₂		Eksp.	Arvut.
17	fenüül/4-amino-3-fluorobensüül	t	8.096 8.927
18	fenüül/3-(N-metüülamino)- 4-fluorobensüül		8.309 8.980
19	fenüül/3-(pürasool-3-üül)bensüül	t	10 10.141
20	fenüül/indasool-5-üül-metüül	t	10.698 9.883
21	fenüül/indasool-6-üül-metüül	t	10.221 9.594
22	fenüül/ (3-metüülindasool-5-üül)-metüül	t	10 9.988
23	fenüül/ 3-aminoindasool-5-üül-metüül		10.522 10.408
24	fenüül/3-aminobensisokasool- 5-üül-metüül		9.387 9.425
25	fenüül/3-(5-metüül-2-püridiil karboksamido)-bensüül	t	10.096 9.829
26	fenüül/3-(N-2-tiasolüül karboksamido)-bensüül	t	10.522 9.562
27	fenüül/H	t	6.013 5.977



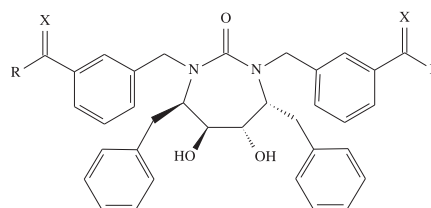
P ₂ /P ₂ '		Eksp.	Arvut.
28	bensüül/H	t	7.18 7.179
29	3-tsüanobensüül/H	t	7.823 7.666
30	3-tsüano-4-fluorobensüül/H		6.958 6.987
31	3-hüdroksüülbensüül/H	t	8.585 8.116
32	3-aminobensüül/H	t	8.207 8.448
33	3-(karboksamido) bensüül/H	t	8.251 8.561
34	3-(karboksamido)- 4-fluorobensüül/H	t	7.853 8.680
35	3-(karboksamido oksiim)bensüül/H		9.619 9.289
36	3-aminoindasool-5-üül-metüül/H	t	9.42 9.573
37	indasool-5-üül-metüül/H	t	8.92 8.784
38	H/indasool-5-üül-metüül		7.356 8.880
39	H/3-aminobensüül		6.804 8.719
40	bensüül/3-tsüano-4-fluorobensüül	t	7.481 7.612
41	3-tsüano-4-fluorobensüül/bensüül	t	7.397 7.646



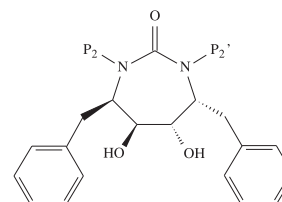
P_2/P_2'		Eksp.	Arvut.
42	3-(karboksamido oksii)bensüül/ tsüklopropüülmetüül	t	9.318 10.105
43	bensüül/ 3-aminindasool-5-üül-metüül	t	9.187 9.654
44	3-aminindasool-5-üül-metüül/ bensüül	t	9.552 9.760
45	indasool-5-üül-metüül/ 3-aminobensüül	t	10.096 9.760
46	3-aminobensüül/ indasool-5-üül-metüül		10 9.976
47	3-aminobensüül/ (3-metüülindasool-5-üül)-metüül	t	9.886 10.297
48	3-aminobensüül/3-tsüanobensüül	t	8.795 9.176
49	3-aminobensüül/ 3-aminindasool-5-üül-metüül		10 10.214
50	3-aminobensüül/ 3-(karboksamido) bensüül	t	10.221 9.920
51	3-aminobensüül/ 3-(karboksamido oksii)bensüül	t	10.698 10.171



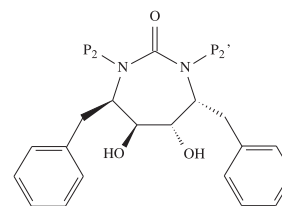
R/X		Eksp.	Arvut.
52	etüül/O	t	9.68 9.399
53	n-propüül/O		8.86 8.651
54	t-butüül/O	t	8.45 9.047
55	H/N(OH)	t	11.01 10.764
56	metüül/N(OH)		10.75 10.196
57	etüül/N(OH)	t	10.51 10.268
58	n-propüül/N(OH)		10.51 10.407
59	CF ₃ /N(OH)		8.41 11.376



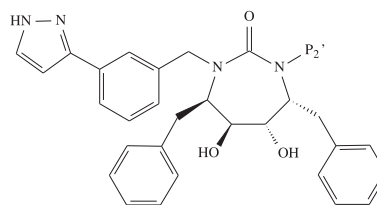
P_2/P_2'		Eksp.	Arvut.
60	allüül	t	8.29 8.420
61	n-propüül		8.1 7.794
62	n-butüül	t	8.86 8.339
63	3,3-dimetüülallüül	t	8.8 8.382
64	3-metüülallüül	t	7.93 7.954



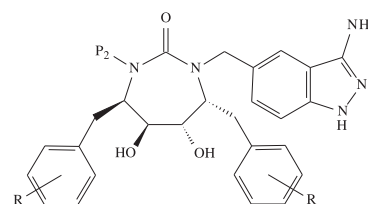
P_2/P'_2		Eksp.	Arvut.
65	tsüklopropüülmetüül	t	8.68
66	tsüklobütüülmetüül	t	8.89
67	tsüklopentüülmetüül	t	8.37
68	tsükloheksüülmetüül	t	7.44
69	bensüül	t	8.53
70	3-nitrobensüül	t	8.56
71	4-nitrobensüül	t	7.5
72	3-aminobensüül	t	9.56
73	4-aminobensüül	t	8.96
74	3-tsüanobensüül	t	8.53
75	4-tsüanobensüül	t	7.29
76	3-hüdroksübensüül		9.93
77	4-hüdroksübensüül	t	9.93
78	3-(hüdroksümetüül)bensüül	t	9.86
79	4-(hüdroksümetüül)bensüül	t	9.47
80	2-naftüülmetüül	t	9.51



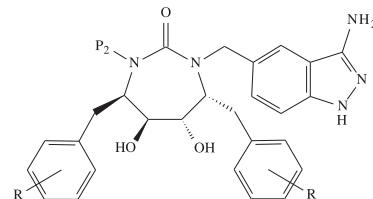
P'_2		Eksp.	Arvut.
81	H		9.56
82	m-MeO ₂ CPhCH ₂		9.78
83	m-HOH ₂ CPhCH ₂		9.87
84	m-H ₂ NPhCH ₂		10.43
85	m-püridiinCH ₂	t	9.77
86	p-HOPhCH ₂	t	10.72
87	m-(morfoliin etüül-NHCO)PhCH ₂		10.33



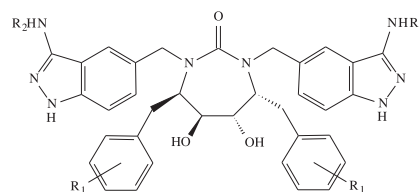
R/P_2		Eksp.	Arvut.
88	H/bensüül	t	10.51
89	4-CH ₃ /bensüül		10.33
90	3-CH ₃ /bensüül		10.21
91	4-C ₂ H ₅ /bensüül		10.14
92	3,5-di-CH ₃ /bensüül	t	9.7
93	H/bütüül	t	10.68
94	4-CH ₃ /bütüül		10.43
95	3-CH ₃ /bütüül		10.35



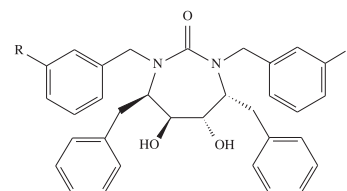
R/P ₂	Eksp.	Arvut.
96 4-C ₂ H ₅ /butüül	t	10.03 9.726
97 3,5-di-CH ₃ /butüül	9.72	10.236
98 H/tsüklopropüülmetüül	10.7	10.169
99 H/tsüklobutüülmetüül	10.8	10.144
100 H/CH ₂ -2-naftüül	10.64	10.287
101 H/C ₅ H ₁₁	t	10.39 9.980
102 H/C ₆ H ₁₃	9.89	9.995



R ₁ /R ₂	Eksp.	Arvut.
103 H/H	11	10.940
104 4-CH ₃ /H	10.21	10.652
105 3-CH ₃ /H	10.96	10.803
106 3,5-di-CH ₃ /H	10.17	10.501
107 2,5-di-CH ₃ /H	10.47	10.670
108 4-C ₂ H ₅ /H	10.17	10.271
109 4-C ₃ H ₇ /H	t	10.17 10.330
110 4-CH(CH ₃) ₂ /H	9.72	10.033
111 4-C(CH ₃) ₃ /H	t	9.72 9.566
112 3-C(CH ₃) ₃ /H	9.68	10.263
113 4-C ₄ H ₉ /H	t	9.77 9.720
114 H/CH ₃	t	10.74 10.506
115 H/C ₂ H ₅	10.39	10.352
116 H/CH(CH ₃) ₂	t	10.05 10.418
117 H/tsüklopropüülmetüül	t	9.47 10.366
118 H/C ₃ H ₇	t	9.64 10.616



R	Eksp.	Arvut.
119 3-pürasool	t	10.57 10.376
120 4-pürasool	9.21	10.416
121 2-imidasool	9.79	9.932
122 4-imidasool	9.73	10.024



	P_1/P'_1		Eksp.	Arvut.
123	bensüül	t	10.62	10.428
124	etüül		8.35	10.154
125	isobutüül	t	9.52	9.961
126	heksüül	t	9.46	9.400
127	tsükloheksüül	t	9.54	8.974
128	p-H ₂ N-bensüül		10.8	10.702
129	p-Me ₂ N-bensüül	t	10.42	10.263
130	p-OH-bensüül	t	10.46	10.693
131	p-OCH ₂ Ph-bensüül		8.43	10.433
132	p-OCH ₃ -bensüül	t	10.59	10.799
133	etüül/bensüül	t	10.16	10.099
134	isobutüül/bensüül	t	10.42	10.020
135	heksüül/bensüül	t	10.34	9.959

