

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
Keemia instituut
Molekulaartehnoloogia õppetool

Tanel-Sigmar Sildoja
**Kvantitatiivne struktuur-omadus sõltuvus
raviainete lahustuvusele inimese soolevedelikes**

Magistritöö
Keemia eriala
30 EAP

Juhendaja Uko Maran, PhD

Tartu 2020

Infoleht

Kvantitatiivne struktuur-omadus sõltuvus raviainete lahustuvusele inimese soolevedelikes

Lahustuvust inimese soolevedelikus dikteerib raviaine võime absorbeeruda seedetraktis ja on seega esmane parameeter raviainete ja nende kandidaatide kirjeldamisel. Selle omaduse *in silico* modelleerimine säästab aega ja raha raviainete väljatöötamisel ravimiarenduse varajases faasis. Seetõttu koostati ja süstematiseeriti kirjanduse baasi andmeseeria raviainete lahustuvusest inimese soolevedelikus ning modelleeriti seeriat QSAR meetodil. Leiti hea korrelatsioonikordajaga ($R^2 \approx 0.8$) kuueparameetriline sõltuvus 2D deskriptoritega, mida selgitati mehhanistlikest seisukohtadest lähtuvalt ja mudel valideeriti välise andmeseeriaga.

Märksõnad: QSAR, lahustuvus soolevedelikus, raviained, ravimitaolised ained

CERCS koodid: P410 Teoreetiline ja kvantkeemia

Quantitative structure-activity relationship for the solubility of drug substances in human intestinal fluids

Solubility in human intestinal fluid is dictated by the ability of the drug substance to be absorbed in the gastrointestinal tract and is thus a primary parameter in describing drugs substances and their candidates. *In silico* modeling of this property saves time and money in the early stages of drug development. Therefore, a data series on the solubility of drug substances in human intestinal fluid was compiled and systematized from the literature and this series was modeled by the QSAR method. A six-parameter relationship with 2D descriptors with a good correlation coefficient ($R^2 \approx 0.8$) was found, which was explained from mechanistic points of view, and the model was validated with an test data set.

Keywords: QSAR, solubility in intestinal fluids, drug substances, drug-like compounds

CERCS codes: P410 Theoretical chemistry, quantum chemistry;

Sisukord

Infoleht	2
Lühendite nimekiri	4
Sissejuhatus	5
1. Kirjanduse ülevaade	6
1.1 Raviainete lahustuvus vees.....	6
1.2 Raviainete lahustumine soolevedelikes.....	7
1.3 Soolevedelikes lahustumist ja selle mõõtmisi mõjutavad faktorid	7
1.4 Kunstlikud soolevedelikud ja nende kasutamine	8
1.3 Soolevedliku proovide võtmine, töötlemine ja säilitamine	9
1.4 QSAR mudelid lahustuvuse ennustamiseks soolevedelikes	10
2. Andmekomplekt ja meetodid	12
2.1 Andmete kogumine ja korrastamine	12
2.2 Struktuuride leidmine ja töötlemine	14
2.3 Mudeli treenimise ja testimise valimi koostamine	14
2.4 Molekulaartunnuste arvutamine	15
2.5 Kvantitatiivsete struktuur-aktiivsuse sõltuvuste tuletamine.....	15
2.6 Mudeli rakenduspiirkonna hindamine.....	16
3. Tulemused ja arutelu	17
4. Kokkuvõte	22
5. Summary	23
Viited.....	24

Lühendite nimekiri

- ADMET absorptsioon, distributsioon, metabolism, eritumine ja toksilisus
(i.k. *absorption, distribution, metabolism, elimination/excretion, and toxicity*)
- BMLR parima multilineaarse regressiooni lähenemine
(i.k. *best multi-linear regression*)
- CAS *Chemical Abstract Service*
i.k. inglise keeles
l.k. ladina keeles
- LFER lineaarne vabaenergia sõltuvus
(i.k. *linear free-energy relationship*)
- $\log P$ logaritm oktaanol-vesi jaotuskoefitsiendist
- $\log D$ logaritm oktaanol-vesi jaotustegurist
- MLR multilineaarne regressioon
(i.k. *multi-linear regression*)
- PLS osalise vähimruutude
(i.k. *partial least squares*) regressioon
- QSAR kvantitatiivne struur-aktiivus sõltuvus
(i.k. *quantitative structure-activity relationship*)

Sissejuhatus

Raviainete avastamine ja arendamine on keerukas ja ressursimahukas, mis nõuab laialdasi teadmisi erinevates valdkondades meditsiinist keemia ja arvutuskeemiani [1, 2]. Ravimite arendamise varajases etapis on tänapäeval üheks oluliseks sammuks keemilise ühendi struktuurist lähtuvate omaduste alusel sobivate raviainete kandidaatide virtuaalne väljasõelumine [3]. See omaduste komplekt koosneb parameetritest nagu raviaine võime absorbeeruda seedetraktis, jaotuda, muunduda ja erituda organismis(t) ning mitte olla toksilise toimega organismi suhtes (nn ADMET omadused) [4]. ADMET omaduste alusel raviaine kandidaatide väljanappimine suurtest andmehulkadest on oluline samm varajase ravimiarenduse faasis ja sõltub suuresti ADMET omaduste ennustamiseks loodud arvutusmudelitest [5].

Esimene raviaine ADMET omadus on absorptsioon, mis sõltub kahest raviaine jaoks olulisest parameetrist, kõigepealt lahustuvusest soolevedelikes ja seejärel imendumisest seedetrakti epiteelkihis, et jõuda vereringesse [6]. See teeb lahustuvusest soolevedelikes olulise raviaine omaduse, mida korrektselt modelleerides on võimalik muuta täpsemaks raviainete kandidaatide virtuaalset väljasõelumist molekulide andmebaasidest. Samas on lahustumine soolevedelikes keeruline protsess võrreldes tavalise vesikeskkonnaga, kuna soolevedelik inimkehas on heterogeenne ja dünaamiline (st. pidevalt liikuv ja muutuv) keskkond, mis sisaldab endas palju varieeruvaid faktoreid, ja mis omakorda mõjutavad lahustuvust. Näiteks võivad sapp ja muud pindaktiivsed ühendid solubiliseerida ka vees lahustumatuid lipofiilsemaid aineid, või siis pH tase, mis mõjutab raviaine molekuli ionisatsiooni ja sealt tulenevalt lahustuvust, või toidu komponendid ja toidu olemasolust tingitud füsioloogilised muutused soolevedelike koostises. Eelpool kirjeldatud parameetrid mõjutada lahustuvust suurusjärgude võrra [7, 8].

Arvestades eelnevat on magistritöö eesmärgiks luua andmebaas raviainete lahustuvusest soolevedelikes ja leida kvantitatiivne struktuur-aktiivus sõltuvus (QSAR) raviainete ning nende kandidaatide lahustuvuse hindamiseks aspireeritud soolevedelikes.

1. Kirjanduse ülevaade

Kirjanduse ülevaade keskendub lahustuvusele kui ühele tähtsaimale ADMET omadusele ja vaatleb raviainete lahustuvust soolevedelikes. Kuna lahustuvus soolevedelikus on struktuuripõhiseks modelleerimiseks uus omadus, siis vaadeldakse soolevedeliku kui heterogeenset keskkonda ja ka paljusid erinevaid faktoreid, mis selles keskkonnas lahustuvust mõjutavad. Seejärel kirjeldatakse seni kirjanduses ilmunud ainukest QSAR modelleerimise tulemust ja diskuteeritakse ka lahustuvuse modelleerimist kunstlikes soolevedelikes, sh selle probleeme võrreldes reaalsete soolevedelikega.

1.1 Raviainete lahustuvus vees

Raviainete lahustuvus määrab nende biokättesaadavuse ja on seega oluline molekuli omadus määramaks raviaine jõudmist sihtmärgini. Kuna lahustuvus on üks raviainete ADMET omadustest, siis on ta kaudselt esindatud ka erinevates piirväärtusepõhistes ravimitaolisuse hindamise juhtnöörides, nagu seda on nn. Lipinski viie reegel [9]. Need on molekulide struktuuriomadused, mille järgi võib eeldada, et molekul on sobilik raviaine kandidaat. Nii on Lipinski viie reegli järgi keemiline ühend hea raviaine kandidaat, kui täidetud on kaks viiest tingimusest: molekulmass < 500 Da, $\log P < 5$, vesiniku doonor-rühmade (O-H või N-H rühmad) arv < 5 , vesiniku aktseptor-rühmade arv < 10 (O või N aatom, sealhulgas doonorrühmades esinevad) [10, 11]. Seda reeglite kogumit kasutatakse laialdaselt raviainete kandidaatide andmebaaside otsingutes, kuid selgelt on välja kujunenud tendents, kus raviained kõiki neid omadusi ei järgi [9]. Umbes pooled uurimise all olevatest ühenditest on vähelahustuvad (pöördvõrdeline seos $\log P$ -ga) ja eeldatakse, et lähitulevikus see ei muutu [12]; trend on pigem vastupidine, uued raviainete kandidaadid on aina lipofiilsemad [13].

Tavaliselt määratakse keemiliste ühendite lahustuvust vees. Madala lahustuvusega ühendite puhul, nagu seda on raviained, on lahustuvust vees aga raske täpselt mõõta ja sellest lähtuvalt võib lahustuvuse määramine olla ka aeganõudev [14]. Teisest küljest on vees lahustuvuse arvandmeid palju, mistõttu on erineva koostisega andmeseeriadele tuletatud palju ennustusmudeleid [15]. Aastatepikkune ennustusmudelite arendustöö on aga näidanud, et raviainete ja nende kandidaatide vees lahustuvuse ennustamine on suur väljakutse, mis kajastub arvutusmudelite ennustusvõime võistlustes ja nende tulemustes [16, 17, 18] ning seda just raviainete ja nende kandidaatide lahustuvuse ennustamisel. Raskused raviainete lahustuvuse hindamisel tulenevad sellest, et raviainete keemiline struktuuriruum on võrreldes tavaliste tööstuslike kemikaalidega eriline, kuna nende hulgas on üle 90% ühendeid, mis

inimese organismis ioniseerivad: alused, happed ja amfolüüdid [19, 20]. See on tekitanud vajaduse määrata ja ennustada lahustuvust otse soolevedelikes, kus raviaine peab sihtmärgini jõudmiseks lahustuma.

1.2 Raviainete lahustumine soolevedelikes

Suukaudselt manustatud raviainete imendumine (absorbtsioon) toimub põhiliselt peensooles [21, 22]. Ravimi kontsentratsioon peensoole seina lähedal on üks põhilisi parameetreid, mis iseloomustab biokättesaadavust suukaudsel manustamisel, teine on membraani läbitavuse koefitsient [7]. Soolevedelike keerulise koostise tõttu ei ole vees lahustuvus heaks parameetriks iseloomustamiseks aine biokättesaadavust, kuna see katab vaid osa molekulidevahelisi interaktsioonimehhanisme [7, 8]. Oluliselt komplitseeritum on olulord vähelahustuvate, kuid hea sooleseina läbitavusega ainete puhul, kus soolevedelikes lahustumine ja sellest tulenev sooleseina läbitavus on limiteerivaks faktoriks [21].

Raviainete lahustuvusest soolevedelikes täpsema ettekujutuse saamiseks on kindlasti vaja lahustuvust mõõta reaalsetes soolevedelikes. Otse, *in vivo* mõõtmine on aga väga keerukas ja kulukas ning eetiliste probleemidega seoses inimestele raviaine manustamisega ilma oodatava terapeutilise efektita, mistõttu kogutakse aspiraate, e. soolevedeliku proove, ning lahustuvuse mõõtmised sooritatakse väljaspool keha [8]. Nii ongi aspireeritud inimese soolevedelikud muutunud kuldstandardiks ravimite lahustuvuse ja imendumise mõõtmistel [8, 23]. Samas on selliseid mõõtmisi sooritatud vähe ja puudub standardne meetodika soolevedelikes lahustuvuse mõõtmiseks [6].

1.3 Soolevedelikes lahustumist ja selle mõõtmisi mõjutavad faktorid

Lahustuvust soolevedelikes mõjutavad mitmed aspektid, mis teevad puhtalt teoreetilistest lähtekohtadest (nn füüsikalised mudelid) lahustuvuse ennustamise või hindamise võimatuks [8]. Lahustuvusele soolevedelikes avaldavad põhiliselt mõju: lai pH gradient seedetraktis, mis maost tühisooleni (i.k. *jejunum*) Fagerberg *et al.* Järgi inimese toitmata olekus on umbkaudu 2.5 – 6.9 [23]; sapp, mis on põhiline lipofiilsete ainete solubilisatsiooni allikas [24], mistõttu lipofiilsed ained, mille $\log P > 3$, lahustuvad soolevedelikes tunduvalt paremini kui vees [7].

Soolevedeliku koostises esineb erinevate inimeste vahel palju variatsioone [21], mis tulenevad mitmetest faktoritest nagu vanus, sugu, päritolu, jne. Näiteks inimkatsete faasi jõudnud ravimikandidaate manustatakse kõigepealt tavaliselt noortele, tervetele inimestele, kelle soolevedeliku koostis võib märgatavalt erineda vanema ja/või kehvema tervisega isiku

omast [8]: erinevad näiteks pH, puhvermahtuvus, fosfolipiidide ja sapphapete kontsentratsioonid. Neid variatsioone saab vähendada segades mitme inimese soolevedelikud kokku ning eelistatud on koguda soolevedelikke sarnases füüsilis-tervislikus seisundis inimestelt. Ka saadakse nii suuremas koguses ühtlasema koostisega soolevedelikku, andes võimaluse testida samas keskkonnas paljusid aineid [8, 23]. Mõõtes aga mitmelt erinevalt inimeselt võetud soolevedelikes mingi kindla aine lahustuvust eraldi, on võimalik kindlaks teha tähtsaimad lahustuvust mõjutavad faktorid soolevedeliku koostises võrreldes komponentide hulka soolevedelikus ja korreleerides neid lahustuvusega [8].

Variatsioone soolevedelike koostises tekib ka erinevatest proovide võtmise viisidest. Üks kasutatavamaid on Loc-I-Gut kateedersond, millega saab täispumbatavate balloonide abil eraldada proovi võtmiseks 10 cm pikkuse soole osa [21]. Mõjutab ka ajaaken, mille vältel lahustunud aine viibib soolestikus ja vedeliku hulk soolestikus ning toitumus. Toidu olemasolu tõstab pH väärtust maos kuni 7-ni, [23] suurendades proovide variatsiooni. Korralikult ja reprodutseeritavalt pH mõõtmine on eriti tähtis, kuna pH avaldab väga suurt mõju happelis-aluseliste omadustega molekulide lahustumisele ioonisatsiooni tõttu [21]. Näiteks kõrgema keskkonna pH-ga peensooles lahustuvad ja seega ka imenduvad paremini nõrgalt happelised ained, samas aluselised ühendid hakkavad kaotama laengut ning võivad sealt hoopis välja sadenema hakata [7]. Mõõta võib nii aspireeritud vedelikku pH elektroodiga kui ka otse seedetraktis spetsiaalse raadiosaatjaga varustatud seadmega, mis neelatakse alla (nn. Heidelberg'i kapsel). Puhvermahtuvus määratakse enamasti kas ainult happe või aluse suhtes tiitrimise teel, mõlemat väga tihti ei määrata proovide vähese hulga tõttu [21]. Kõiki neid asjaolusid silmas pidades on selge, et raviainete ja nende kandidaatide puhul tuleks eelistada lahustuvuse määramist reaalses soolevedelikus ja ennustamist nendele andmetele tuletatud mudelitega, mitte ainult vees ning vees lahustuvuse mudelitega.

1.4 Kunstlikud soolevedelikud ja nende kasutamine

Soolevedeliku kättesaadavus on piiratud ja vedelik ise ei ole keemiliselt stabiilne: tema pH muutub süsihappegaasi eraldumisel, kuna toitmata olekus on soolevedelike vesinikkarbonaatpuhver toatemperatuuril ebastabiilne [21, 25]. Seetõttu tuleb pH ja puhvermahtuvus koheselt määrata ning vedelik külmutada, ensüümid deaktiveerida, et säiliks aspireerimise hetkel olnud koostis [8]. Need asjaolud muudavad uuringud võrdlemisi kalliks ja tülikaks. Üheks alternatiivseks võimaluseks on kasutada kunstlikke soolevedelikke, mis töötati välja Dressman'i ja kaastöölise poolt 1998. aastal [26]. Kuid ka kunstlike soolevedelike rakendamisel on probleemseid küsimusi – sobiva biorelevantse alternatiivi osas

ei ole üksmeelt, tekitades segadusi mitme kunstliku soolevedeliku kompositsiooni olemasolu tõttu [27] – olemas on nii lihtsaid puhvreid lisatud sünteetiliste amfiifilidega kui ka keerulisi segusid, mis sisaldavad fosfolipiide, sapphappeid ja nende soolasid, rasvhappeid, monoglütseriide [7]. Lisaks ei võimalda kaubanduslikult saadavad kindla koostisega kunstlikud soolevedelikud hinnata ravimite lahustuvuse tundlikkust soolevedelike koostise *in vivo* märgatava varieerumise suhtes [28, 29]. Seetõttu tuleks eelistada eksperimentaalsetel andmetel põhinevaid arvutuslikke ennustavaid meetodeid reaalsele soolevedelikele, nagu näiteks QSAR mudelid, mis on kiired ja kergesti rakendatavad suurtele ühendite andmebaasidele [23].

1.3 Soolevedeliku proovide võtmine, töötlemine ja säilitamine

Soolevedelikes lahustuvuse määramiseks proovide võtmise üleilmselt standardiseeritud protokollid antud hetkel puuduvad, olemasolevad on kasutatavad ainult kindlate tööühemate poolt (näiteks [30]) ja valikut on palju. Olemasolevatel eeskirjadel on ühine, et proovid võetakse tervetelt, puhanud ja toitmata või kindla eeskirja järgi toidetud inimestelt, kuid leidub ka uuringuid, kus soolevedeliku proovid on võetud vanematelt inimestelt [7, 21]. Kõige enam võetakse soolevedeliku proove tühisooles algosast, kuna sondi paigutamine sinna on kiire, proovikogused saab võrdlemisi suured ning sondi asetus on korratav [7]. Proove on võetud ka kaksteistsõrmiksoolest (l.k. *duodenum*) ning ka soolestiku alg-, kesk- kui ka lõpuosast. Leidub ka artikleid, kus kahe sondiga on võetud üks proov nii kaksteistsõrmsoolest kui ka tühisooles [21]. Katseisikule võib anda umbes 180-250 ml vett, seda nii sondiga kui ka lihtsalt lasta ära juua. Vee olemasolu on biokättesaadavuse seisukohalt tegelikule olukorrale lähedasem. Märkida tuleb aga, et vee manustamine muudab pH-d ja soolevedelike komponentide kontsentratsioone [8, 21].

Koostise muutumise vältimiseks säilitamise ajal on vaja ensüümid deaktiveerida. Trüpsiini saab deaktiveerida näiteks fenüülmetüülsulfonüülfluoriidiga, lipaasi tetrahüdrolistatiiniga [8, 21]. Inhibiitorite lõplik hulk peaks olema vähem kui 2% mahu järgi proovist. Mõnikord lisatakse ka mikroobide kasvu pärssivaid aineid. Samas võivad need lahustuvusele mõju avaldada. Proovid säilitatakse temperatuuril -20°C või madalamal kuni kasutamiseni [8, 21]. Proovi pikaajalisel säilitamisel külmas (-20 - -80°C) ning mitmekordsel sulatamisel-külmutamisel suuri pH muutusi täheldatud ei ole [21].

1.4 QSAR mudelid lahustuvuse ennustamiseks soolevedelikes

QSAR seob sõltuva muutuja (modelleeritav omadus) sõltumatute muutujatega (molekulaartunnused), mis iseloomustavad andmeseeria molekulide struktuurseid omadusi. Sellise lähenemise eesmärgiks on tuletada eksperimentaalsetel andmetel põhinev matemaatiline mudel, mille abil saaks kirjeldada ja ennustada uute molekulide käitumist ja omadusi ilma, et peaks sooritama aeganõudvaid, kulukaid ja mõnikord eetilistelt küsitavaid eksperimente, sh. ka molekule kasutult sünteesima [31, 32].

Tuletatav omaduse ja struktuuri vaheline seos peab olema loogiline, ideaaljuhul ka mehhanistliku seletusega, st. molekulaartunnused ehk molekulide struktuurseid omadusi kirjeldavad parameetrid peavad olema molekuli struktuuri ja käitumise vaatenurgast lihtsasti ja loogiliselt seletatavad ning mudel valideeritud ehk tõestatud ta sobivus ka teiste andmeseeriatega korral [33]. Nagu ka eelnevates peatükkideski mainitud, säästaks selline lähenemine aega ja raha ning väldiks eetilisi probleeme, mis tekiks inimestelt proovide võtmisega [34].

Tihti peale väidetakse, et mudel ei saa olla täpsem, kui mudeli tuletamiseks kasutatud lähteandmed [35]. Samas on Palmer ja Mitchell oma töös leidnud, et suuremad vead tekivad hoopis algoritmide mittetäiuslikkusest ja sobimatute molekulaartunnuste kasutamisest [36], mis viitab vajadusele kriitiliselt hinnata molekulaartunnuste sobivust eksperimentaalse omaduse kirjeldamiseks.

Leidub ainult üks QSAR mudel raviainete lahustuvusele inimese soolevedelikus, mille on struktuurselt varieeruvatele raviainetele tuletanud Fagerberg *et al.* [23]. Mudeli saamiseks kasutati osaliste vähimruutude (PLS) regressiooni meetodit, treening- ja valideerimisvalimi koostamiseks joondati ühendid eksperimentaalse väärtuse alusel ja valiti kindla sammuga ühendeid valideerimise andmekomplekti. Lisameetodina kasutati põhikomponentide analüüsi, mille abil viidi treeningvalimist liialt kõrvalekalduvaid ühendeid valideerimisvalimisse, et vältida nende liialt suurt mõju mudeli treenimisele. Saadi sõltuvus, mis kirjeldas uuritavat omadust üheksa molekulaartunnusega ($R^2 = 0.84$), mis iseloomustasid molekuli suurust, kompleksust ja vesiniksidemeid. Mõned neist valdavalt topoloogilistest molekulaartunnustest olid kaalutud elektronstruktuuri kirjeldavate parameetritega. Seda mudelit parandati täiendavalt viies mudelisse parameetri, mis kirjeldas eksperimentaalselt mõõdetud lahustuvust pH=6.5 puhvis; mudel mõnevõrra paranes $R^2 = 0.86$. Mudeli ennustusvõime parandamist üritati lisaks saavutada ka sulamistemperatuuri ja $\log D$ sisseviimisega mudelisse, kuid see ei andnud soovitud tulemust.

Lisaks eelnevale tööle leidub kirjanduses üksikuid uuringuid, kus on modelleeritud raviainete lahustuvust kunstlikes soolevedelikes [37, 38]. Fagerberg *et al.* töös [37] leiti seitsme molekulaartunnusega QSAR struktuurselt varieeruvatele raviainetele, $R^2 = 0.69$. Selles sõltuvuses olid kõik molekulaartunnused seotud molekuli suurusega ja olid negatiivselt korreleeritud lahustuvusega, e. suurem – lahustub vähem. Autorid kasutasid analoogiat eelnevas lõigus kirjeldatud uuringuga ja püüdsid mudelit parandada täiendavate molekulaartunnustega, milleks kõigepealt oli eksperimentaalne lahustuvus ning mudel ka paranes ($R^2 = 0.76$). Samas ka selle andmeseeria ja omaduse juures sulamistemperatuur ja $\log D$ lisamine mudelit ei parandanud.

Niederquell ja Kuentz tuletasid mudeli biorelevantsetes meediumis 40'le kirjandusest kogutud ühendile, mis pH=6.5 juures on neutraalsed [38]. Selleks kasutati nn lineaarset vabaenergia sõltuvust (LFER) lahustuvusele ($R^2 = 0.810$), mis koosneb viiest Abraham'i molekulaartunnusest: vesiniksideme happelisus ja aluselisisus, polariseeritavus, molaarse refraktsiooni liig, ruumala. Need molekulaartunnused koos kirjeldavad lahustunud aine üleminekut ühest faasist teise (lahusesse). Keskmise viga oli 0.299, kuid see on eeldatav lahustunud aine üliväikese kontsentratsiooni tõttu. Huvitaval kombel oli ruumalal tugev positiivne korrelatsioon, mis vees lahustumist peaks takistama. Seda seletati lipofiilsema aine mitselleerumisvõimega soolevedelike koostisosades. Mudelit valideeriti statistiliselt tehes deskriptoritele PLS uuringu, et leida nende olulisus.

2. Andmekomplekt ja meetodid

2.1 Andmete kogumine ja korrastamine

Töös kasutatud andmeseeria saamiseks liideti ja korrastati kaks eelnevat andmekogumit, mis olid teineteisest sõltumatult koostatud Augustijns *et al.* [13] ning Fagerberg'i ja Bergström'i [7] poolt. Modelleeritavaks suuruseks oli lahustuvus inimeste soolevedelikes (S_{sv}), ühikuks $\mu\text{g/ml}$. See omakorda viidi üle logaritmi kujule ($\log S_{sv}$). Andmeid võib eksperimentaalse külje pealt iseloomustada järgnevalt:

- kasutatud on inimese toitmata olekus aspireeritud soolevedelike mõõtmistulemusi;
- enamikel ühenditel on mitmeid ajas ja ruumis ning ka meetodika suhtes erinevaid mõõtmisi, mõnikord erinevad tulemused lausa suurusjärgude võrra;
- proovid on võetud nii üksikindiviididelt, kui ka mitmelt inimeselt kokku segatud;
- erinevad sood ja vanused;
- pH vahemik 5.4 – 7.5;
- proovid pärinevad kas kaksteistsõrmiksoolest või tühisoolest;
- kõigil juhtudel ei olnud täpset asukohta välja toodud.

Andmeseeria korrastamise käigus:

- eemaldati kahe komplekti korduvad tulemused ning parandati mõningad vead kontrollides algallikaid;
- andmeseeriast jäeti välja anomaalselt kõrge $\log P / \log D$ väärtusega ühendid, mille ennustatud väärtuse korrektsuses kaheldi.

Lõpuks jäi modelleerimiseks kasutatavasse kogumikku alles 71 raviainet (Tabel 1). Tegemist on struktuurselt väga varieeruvate ühenditega. Korduvate mõõtmistulemuse korral on kasutatud kõigi mõõtmistulemuste mediaani.

Tabel 1. Modelleeritav andmeseeria: Raviaine, CAS number, seeriad, ennustatud väärtus ($\log S_{sv}$).

jrkk	CAS NR	Raviaine (i.k.)	Seeriad	$\log S_{sv}$ (ennust.)
2	1951-25-3	<i>Amiodarone</i>	Treening	2.404
5	170729-80-3	<i>Aprepitant</i>	Treening	1.329
6	68844-77-9	<i>Astemizole</i>	Treening	1.964
7	198904-31-3	<i>Atazanavir</i>	Treening	0.749
8	29122-68-7	<i>Atenolol</i>	Treening	5.451
9	94015-53-9	<i>Atovaquone</i>	Treening	0.359
10	248919-64-4	<i>AZD0865</i>	Treening	1.502
11	1009298-09-2	<i>AZD8055</i>	Treening	1.388
12	5534-09-8	<i>Beclomethasone dipropionate</i>	Treening	-0.134
13	298-46-4	<i>Carbamazepine</i>	Treening	2.411
14	72956-09-3	<i>Carvedilol</i>	Treening	1.880
15	298-57-7	<i>Cinnarizine</i>	Treening	1.335
17	59865-13-3	<i>Cyclosporine</i>	Treening	0.981
18	17230-88-5	<i>Danazol</i>	Treening	1.675
19	206361-99-1	<i>Darunavir</i>	Treening	1.803
20	439-14-5	<i>Diazepam</i>	Treening	2.089
21	15307-86-5	<i>Diclofenac</i>	Treening	2.228
22	56-53-1	<i>Diethylstilbestrol</i>	Treening	1.220
23	58-32-2	<i>Dipyridamole</i>	Treening	1.154
26	49562-28-9	<i>Fenofibrate</i>	Treening	1.861
27	83799-24-0	<i>Fexofenadine</i>	Treening	2.182
28	530-78-9	<i>Flufenamic acid</i>	Treening	2.594
29	54-31-9	<i>Furosemide</i>	Treening	2.876
30	184475-35-2	<i>Gefitinib</i>	Treening	1.962
31	10238-21-8	<i>Glibenclamide</i>	Treening	1.534
32	29094-61-9	<i>Glipizide</i>	Treening	2.319
34	58-93-5	<i>Hydrochlorothiazide</i>	Treening	2.769
37	150378-17-9	<i>Indinavir</i>	Treening	2.204
40	84625-61-6	<i>Itraconazole</i>	Treening	0.975
41	65277-42-1	<i>Ketoconazole</i>	Treening	1.647
42	22071-15-4	<i>Ketoprofen</i>	Treening	3.136
43	248919-64-4	<i>Linaprazan</i>	Treening	1.502
44	192725-17-0	<i>Lopinavir</i>	Treening	0.742
46	89-57-6	<i>Mesalazine</i>	Treening	3.372
47	51384-51-1	<i>Metoprolol</i>	Treening	4.468
48	26159-31-9	<i>Naproxen</i>	Treening	2.722
50	21829-25-4	<i>Nifedipine</i>	Treening	0.760
51	51803-78-2	<i>Nimesulide</i>	Treening	2.306
52	39562-70-4	<i>Nitrendipine</i>	Treening	1.076
54	36322-90-4	<i>Piroxicam</i>	Treening	2.645
55	171228-49-2	<i>Posaconazole</i>	Treening	1.524
56	19216-56-9	<i>Prazosin</i>	Treening	2.071
57	50-24-8	<i>Prednisolone</i>	Treening	2.586
58	57-66-9	<i>Probenecid</i>	Treening	3.108
59	57-83-0	<i>Progesterone</i>	Treening	1.220
60	318-98-9	<i>Propranolol</i>	Treening	4.341
61	56-54-2	<i>Quinidine</i>	Treening	3.934
62	13292-46-1	<i>Rifampicin</i>	Treening	3.079
63	168273-06-1	<i>Rimonabant</i>	Treening	0.332
64	155213-67-5	<i>Ritonavir</i>	Treening	1.124

jrk	CAS NR	Raviaine (i.k.)	Seeriad	$\log S_{sv}$ (ennust.)
65	127779-20-8	<i>Saquinavir</i>	Treening	1.983
66	52-01-7	<i>Spironolactone</i>	Treening	0.918
67	599-79-1	<i>Sulfasalazine</i>	Treening	2.712
68	10540-29-1	<i>Tamoxifen</i>	Treening	2.117
69	174484-41-4	<i>Tipranavir</i>	Treening	1.557
70	738-70-5	<i>Trimethoprim</i>	Treening	3.300
71	81-81-2	<i>Warfarin</i>	Treening	2.541
1	54965-21-8	<i>Albendazole</i>	Test	2.102
3	161814-49-9	<i>Amprenavir</i>	Test	0.270
4	60-80-0	<i>Antipyrine</i>	Test	-0.024
16	23593-75-1	<i>Clotrimazole</i>	Test	0.675
24	269055-15-4	<i>Etravirine</i>	Test	2.421
25	119945-59-4	<i>Felodipine</i>	Test	2.061
33	126-07-8	<i>Griseofulvin</i>	Test	1.750
35	50-23-7	<i>Hydrocortisone</i>	Test	2.385
36	15687-27-1	<i>Ibuprofen</i>	Test	2.780
38	53-86-1	<i>Indomethacin</i>	Test	0.417
39	138402-11-6	<i>Irbesartan</i>	Test	1.374
45	147362-57-0	<i>Loviride</i>	Test	0.682
49	159989-64-7	<i>Nelfinavir</i>	Test	1.476
53	73590-58-6	<i>Omeprazole</i>	Test	3.837

2.2 Struktuuride leidmine ja töötlemine

Andmeseerias välja toodud ühendite triviaal- või kommertsiaalnimetuste järgi otsiti andmebaasist *PubChem* [39] välja ainete struktuurid ning laeti alla *Z*-maatriksit sisaldav fail, mille korrektsust kontrolliti visuaalselt tarkvaraga *Molden* (versioon 5.8.2) [40] ning vajadusel parandati ühendite struktuure manuaalselt. Sooritati algne konformatsiooniotsing tarkvarapaketi *Schrödinger Suite 2019* alamprogrammiga *Macromodel* versioon 11.3 [41] molekulide energiat minimiseerides, et oleks parem lähtegeomeetria poolempiirilisteks kvantkeemilisteks arvutusteks tarkvaraga *MOPAC* versioon 7.0 [42]. Eelnevad sammud andsid vajalikud sisendtoimikud programmi *CODESSA PRO* [43] jaoks, mida kasutati mudeli tuletamiseks.

2.3 Mudeli treenimise ja testimise valimi koostamine

Mudeli treenimise ja testimise valimi koostamise kasutati intuitiivset lähenemist [44], mis tagaks andmepunktide võimalikult ühtlase jaotuse eksperimentaalse väärtuste skaalal. Selleks reastati andmeseeria 71 ühendit eksperimentaalse väärtuse kasvavas järjekorras ja loeti viieks. Seal alates kolmandast ühendist valiti iga viies ühend valideerimisvalimisse. Treeningvalimisse jäi 57 ja valideerimiseks 14 ühendit. Valimisel lähtuti selles, et minimaalset ja maksimaalset omaduse väärtust omav ühend ei tohiks sattuda

valideerimisvalimisse, st. valideerimisvalim peab mahtuma treeningvalimi sisse ning kajastama treeningvalimi andmejaotust võimalikult suurel määral [32].

2.4 Molekulaartunnuste arvutamine

Ühe- ja kahedimensionaalsed molekulaartunnused arvutati vabavaralise tarkvaraga *PaDEL-descriptor* versioon 2.21 [45]. Neid oli kokku 1130. Töö üldise kättesaadavuse ja reprodutseerimise hõlbustamiseks on deskriptorite arvutamiseks kasutatud vabavaralist tarkvara ning 2D deskriptoreid, kuna need on lihtsa matemaatilise tagapõhjaga ning ei nõua keerukaid ja aeganõudvaid arvutusi [46] ja võimaldavad mudelid hiljem lihtsamalt kättesaadavaks ja reprodutseeritavaks teha.

Lisaks arvutati molekulaartunnus *log D*. Kui *log P* kajastab vaid neutraalse vormi jaotumist, võtab *log D* arvesse molekuli tegeliku vormi kindlatel tingimustel (pH, ioontugevus). *Log D* arvutati välja tarkvaraga *Marvin* [47]. *Log D* jaoks lähtestati parameetriteks pH = 6.5 ja ioontugevuseks 0.15 mM, mis kajastavad kõige paremini inimese seedetrakti füsioloogilisi tingimusi ning meid huvitavat pH vahemikku kaksteistsõrmiksooles ja tühisooles. Võrdluseks, Lindahl *et al.* [48] on arvutanud ioontugevuseks toitmata inimese soolevedelikes 0.139 ± 0.014 mM, Fuchs ja Dressman on toonud välja võrdlusi soolestiku eri osade pH osas, mis kattuvad käesolevas töös kasutatuga [21].

Uurmistöö algaasis arvutati hulk keerulisemaid, 3D ja elektroonstruktuuri kajastavaid molekulaartunnuseid ka programmi *CODESSA PRO* abiga [43], kuid neid kasutati vaid töö algaasis ja jäeti kõrvale, kuna nad ei andud piisavalt häid tulemusi võrreldes *PaDEL-descriptor* molekulaartunnustega.

2.5 Kvantitatiivsete struktuur-aktiivsuse sõltuvuste tuletamine

Tarkvaraga *CODESSA PRO* tuletati treeningvalimile mudelid, kusjuures kasutati vaid vabavaralisest tarkvarast *PaDEL-descriptor* pärit 1D ja 2D deskriptoreid ning *Marvin*'iga arvutatud *log D* väärtust. Mudelid tuletati BMLR [49] meetodil ning üheparameetrine mudel *CODESSA PRO*'le omasel HMPRO [43] meetodil. BMLR korral moodustab *CODESSA PRO* korrelatsioonid mittekorreleeruvate deskriptorite vahel, mispeale hakatakse parimale kaheparameetrilisele võrrandile lisama järjest parimaid statistiliste parameetrite alusel sobivaid molekulaartunnuseid, kuni võrrand enam piisavalt ei parane. Arvestati ka seda, et deskriptoritel oleks uuritava omadusega seletatav seos, nad sobiksid loogiliselt ühendite kogumiga lähtudes nende struktuurist ehk [31] et nad oleksid mehhanistlikult

interpreteeritavad [33]. Tuletatud mudelite hindamiseks kasutati statistilisi parameetreid nagu korrelatsioonikordaja ruut (R^2), ristvalideeritud korrelatsioonikordaja ruut (R^2_{cv}), Fisher'i kriteerium (F) ja standardhälbe ruut (s^2) ning kontrolliti sobivust välise valideerimise abil jagades andmehulga treening- ja valideerimisvalimiks (vt. pt. 2.3).

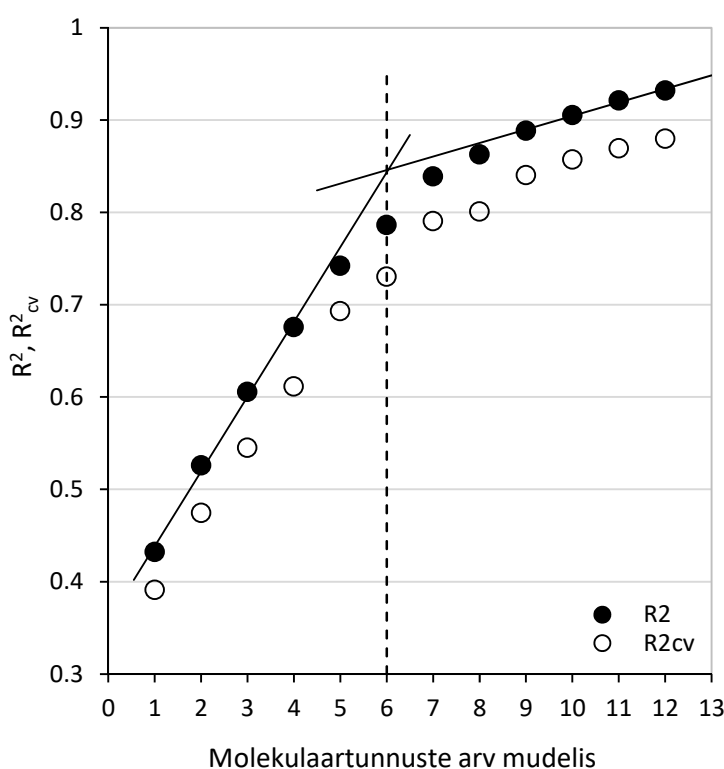
Nagu eelnevas lõigus selgub, on üheks vastuolulisemaks küsimuseks QSAR-te tuletamisel mudelisse valitav sobiv molekulaartunnuste arv. Üheks võimaluseks on leida olukord, kus R^2 enam märgatavalt ei parane, kui lisada molekulaartunnuseid mudelisse, teiseks on nn. käänupunkti meetod [50]. Selleks vaadeldakse mudelisse deskriptorite lisamisel R^2 muutust ja otsitakse käänupunkti R^2 kasvu ja platoole jõudmise vahel, käesolevas töös kombineeriti need meetodid ning vaadeldi mõlemat.

2.6 Mudeli rakenduspiirkonna hindamine

QSAR'i rakenduspiirkond on antud uuringus põhimõtteliselt määratud raviainete struktuuriruumiga, mis varieerub laias vahemikus. Täiendava ülevaate saamiseks QSAR mudeli rakenduspiirkonnast analüüsiti ja vaadeldi tuletatud mudeli kvaliteeti nn Williams'i graafikute [51, 52] abiga. Williams'i graafikul esitatakse standardiseeritud jäägi (standardhälbega normaliseeritud ennustuse viga) ning nn omapära [51, 52]. Iga ühendi omapära arvutatakse mudeli molekulaarsetest deskriptoritest: $h = x^T(X^T X)^{-1}x$ [51, 52], kus x on prognoositava ühendi deskriptori reavektor ja X on treeningvalimi ühendite deskriptorite maatriks. Suurema kui kriitilise omapäraga (defineeritud kui $3p/n$, kus p on ühe võrra suurem kui deskriptorite arv ja n on treeningvalimi ühendite arv) [52] ühendid on struktuurselt erilised ja sellistel treeningvalimi ühenditel võib olla ebaproportsionaalselt suur mõju regressioonisirge kujunemisele ja see mõjutab mudeli usaldusväärsust [52], mistõttu seda tuleb mudeli kasutamisel analoogsete ühendite puhul silmas pidada.

3. Tulemused ja arutelu

QSAR mudelis sobiv molekulaartunnuste arv tuvastati kasutades nn. käänupunkti meetodit (Joonis 1). Selleks vaadeldi R^2 muutumist vastavalt BMLR algoritmi poolt lisatavate molekulaartunnuste arvule. Saadud kõvera tõus ja platoo joonepikenduste lõikumise koht on kuuenda molekulaartunnuse kohal, ehk nn käänupunkt (vt Joonis 1, vertikaalne punktiirjoon). Sellest saab järeldada, et antud treenimise andmekomplekt on optimaalselt modelleeritav kuue molekulaartunnusega.

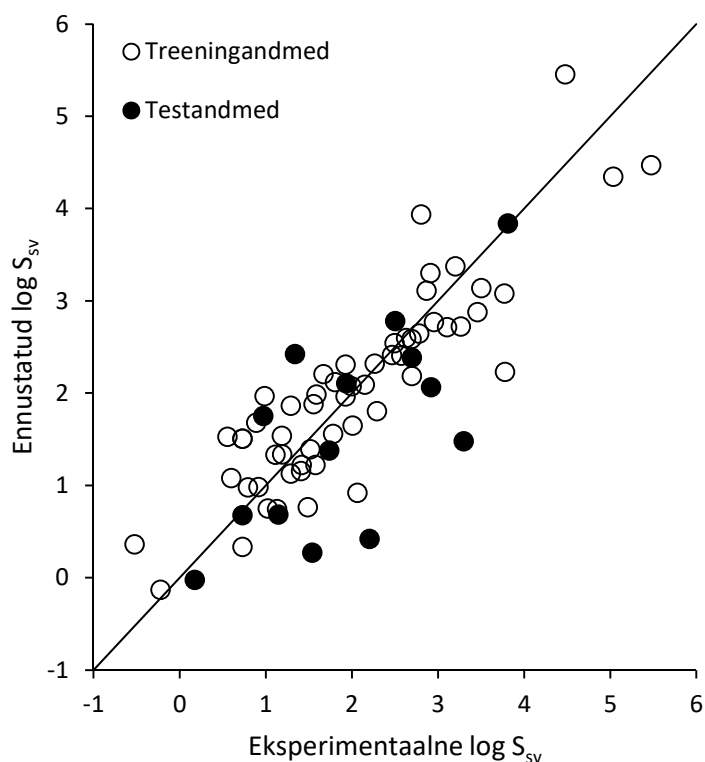


Joonis 1. Korrelatsioonikordaja ruudu ja ristvalideeritud korrelatsioonikordaja ruudu sõltuvus mudelisse lisatavate molekulaartunnuste arvust.

BMLR algoritmi poolt 57'le tuntud raviainele saadud QSAR'i (vt Tabel 2) korrelatsioonikordaja ruut ja vastav ristvalideeritud korrelatsioonikordaja ruut on üksteisele lähedal, $R^2 = 0.7864$ ja $R^2_{cv} = 0.7302$, vastavalt. Nende kahe parameetri mitte väga suur erinevus näitab tuletatud mudeli võimet hinnata piisava täpsusega antud andmekogumiga sarnase keemilise struktuuriga raviained ja nende kandidaate. Saadud sõltuvuse visuaalne esitus (Joonis 2) näitab andmepunktide head jaotumist andeseeria ulatuses nii treening- kui ka testandmete korral.

Tabel 2. Kuue molekulaartunnusega QSAR (parameetrid on järjestatud t -testi järgi): a -koefitsient, s -koefitsiendi hälve, t näitab t -testi väärtusi. $R^2 = 0.7864$, $R^2_{CV} = 0.7302$, $F = 30.69$, $s = 0.5855$, $N = 6$, $n = 57$

Nr	a	s	t	Tähis	Nimetus
0	91.0478	16.055	5.67098	B	Vabaliige
1	-0.52825	0.0452035	-11.686	$\log D$	Oktanool-Vesi jaotuskoefitsient pH-l 6.5 ioontugevusel 0.15 M
2	-0.635037	0.0945757	-6.71459	minssssC	Minimaalne aatomitüübi elektrotoloogiline olek: $>C<$ (sp ³ süsinik)
3	-1.78645	0.279509	-6.39138	SpMax4_Bhs	Burdeni modifitseeritud maatriksi suurim absoluutne omaväärtus - n 4 / kaalutud suhtelise I-olekuga
4	-6.86837	1.35793	-5.05799	BCUTw-11	Kõrgeim madalamatest aatom kaalutud BCUT väärtustest
5	0.505813	0.130033	3.88988	L.	Kõrvalekallete arv Lipinski seadustest
6	71.2289	22.0935	3.22397	JGI9	Üheksandat järku keskmine topoloogiline laenguindeks



Joonis 2. Soolevedelikes vaadeldud ja ennustatud lahustuvuse vaheline seos

Saadud QSAR'is on kuus molekulaartunnust, mis kirjeldavad erinevaid raviainete keemilise struktuuri omadusi. Olulisim molekulaartunnus t -testi järgi on ootuspäraselt $\log D$, mis võtab arvesse raviainete ionisatsiooni, täpsemini ioniseerinud ja mitteioniseerunud vorme antud pH-l, milleks käesoleva töö raames on 6.5. $\log D$ arvutamise tingimused on

seadistatud bioloogilistele tingimustele võimalikult lähedale ning kuna vees lahustuvus on kogu soolevedelikes lahustumise mehhanismi aluseks, on ka oodatav, et see molekulaartunnus saavutab kõrgeima positsiooni. Negatiivne korrelatsioon näitab ka ära, et lipofiilsem aine lahustub soolevedlikus halvemini. $\log D'$ oli ka üksinda parim korrelatsioonikordaja ruut 0.4367, mis on liiga väike, ja ei võimalda üheparameetrilist mudelit iseseisvalt kasutada.

Teine molekulaartunnus QSARis on minsssc (originaaltähtsuses, i.k. „*Minimum atom-type E-State: >C<*“). Tegemist on aatomitüübi elektrotoloogilise oleku molekulaartunnusega [53, 54, 55], mis omab väärtust, kui ühendis on sp^3 süsinik, ja märgib ära tema minimaalse väärtuse kui seda tüüpi süsinikke on rohkem kui üks. Molekulaartunnus omab negatiivset korrelatsiooni, mis tähendab, et deskriptori väärtuse kasvades lahustuvus soolevedelikes kahaneb.

Oluliselt kolmas molekulaartunnus SpMax4_Bhs (*Largest absolute eigenvalue of Burden modified matrix - n 4 / weighted by relative I-state*) on seotud molekuli suuruse ja kompleksusega, millesse panustab iga aatom ja side [56]. Negatiivne korrelatsioon näitab suurema ja kompleksema molekuli raskemat lahustuvust. Siinkohal on andmeseeriast näha, et suurematel molekulidel on palju alifaatseid süsinikahelaid, mis raskendavad lahustumist.

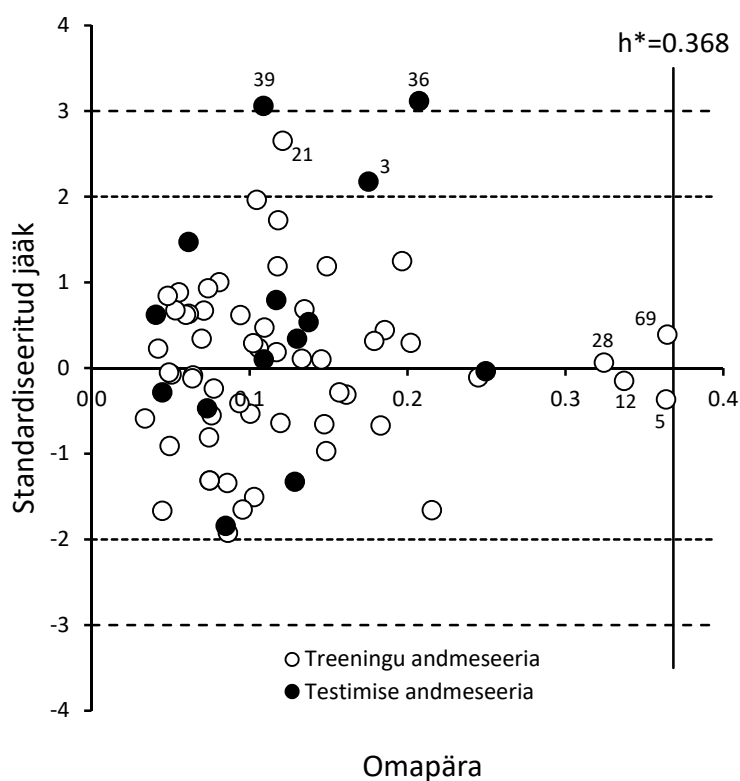
Järgmine molekulaartunnus, BCUTw-11 (*nhigh lowest atom weighted BCUTS*) [57, 58, 59], hõlmab endas molekuli suurust, osalaengut ja polariseeritavust [56]. See tähendab, et analoogselt eelnevaga polariseeritavam ja suurem molekul lahustub raskemini ning korrelatsioon on negatiivne.

Viies molekulaartunnus näitab ära kõrvalekallete arvu Lipinski viie reeglist. Selle huvitav külg on positiivne koefitsient. Võimalik, et need paar kõrvalekallet, mis pole ka eriti suured, ei oma nii märkimisväärset tähtsust, ning muud solubilisatsiooniefektid (mitsellid, sapphapped) peale vees lahustumise kaaluvad sel juhul raskemat lahustuvust vees piisavalt üles.

Mudelis kuues molekulaartunnus, JGI9 (*Mean topological charge index of order 9*) iseloomustab laenguülekandeid molekuli aatomite vahel [56]. Suur võimalus laenguülekanneteks annab suurema osalaenguga molekuli, mistõttu on korrelatsioon tugevalt positiivne- vesi on polaarne lahusti ja polaarsed molekulid lahustuvad paremini.

Esitatud QSAR mudeli sisemine valideerimine ($R^2_{cv} = 0.7302$) on sarnane mudeli R^2 väärtusele, mis näitab head kooskõla treeningandmeseeria ühendite vahel. Samas väline testimise andmeseeria annab valideerimisel suure lahknevuse ($R^2 = 0.4604$). Lähimal vaatlusel ilmneb, et testandmeseerias on 3 hälbivat andmepunkti. Kui testvalimist eemaldada kõrvalekaldujad, siis paraneb korrelatsioonikordaja ruut oluliselt ($R^2 = 0.756$). Mudeli

rakenduspiirkonna analüüs (Joonis 3) näitab, et treeningandmeseeria raviained on kõik ilusti mudeli rakenduspiirkonnas, va. üks, mis on nn tagasihoidlik kõrvalekalduja standardiseeritud jäägi järgi. Samas on testandmeseerias kolm standardiseeritud jäägi järgi hälbijat, milledest üks on 2σ joone lähedal ja ei tohiks mudeli ennustusvõimet palju mõjutada. Teised kaks paiknevad 3σ joonel, mis lubab oletada, et and tugevate hälbijatena ei mõjuta mudeli ennustusvõimet väga palju; kui nad oleks 3σ kaugemal, siis oleks muretsemiseks rohkem põhjust.



Joonis 3. QSAR mudeli rakenduspiirkond esitatuna mudeli standardiseeritud jäägi ja omapära vahelise seosena.

Treeningvalimi nn keskmise kõrvalekalduja standardiseeritud jäägi järgi on ühend 21, i.k. *Diclofenac*. Tegemist on valu- ja põletikuvastase ravimiga, üks populaarsemaid raviaineid üldse [60]. Erilisus seisneb arvatavasti kloreeritud benseenituuma olemasolus.

Valideerimisvalimis tuvastati standardiseeritud jäägi alusel 3 kõrvalekalduvat ühendit. Üks nendest on väga mõõdukas kõrvalekalduja, ühend 3, i.k. *Clotrimazole*. Selle puhul on näha 4 aromaatsset tuuma, mis muudavad selle molekuli suuremaks võrreldes teistega. Napilt üle 3σ jääb ühend 39 (i.k. *Irbesartan*), millel on samuti 2 aromaatsset tuuma ning lisaks annab struktuurile eripära pikad alifaatsed osad. Sellest vaid veidi rohkem kõrvalekalduv on ühend 36, i.k. *Ibuprofen*. Selle puhul on tegemist jällegi väga tavalise raviainega, siinkohal võib

kõrvalekaldumist põhjustada just tema võrdlemisi väikesed mõõtmed ning kindel kuuluvus happeliste ainete klassi.

Kui vaadelda omapära, siis võime Williams'i graafikul põhimõtteliselt eristada kahte klastrit treeningseerias. Omapäraga üle 0.3 jääb neli ühendit, kuid nad ei ületa kriitilist omapära ($h^*=0.368$) olles selle piirjoone lähedal, samas peavad eristuma andmeseerias millegi poolest. Lähimal vaatlusel selgub et neil on ühine joon – nende nelja puhul on tegemist võrdlemisi suurte ühenditega, millel esineb mitmeid aromaatsid tuumi ning palju halogeeniaatomeid, eriti fluori.

4. Kokkuvõte

Raviainete ADMET omadustest on esmasena oluline lahustuvus inimese soolevedlikes, mis määrab raviaine bio-kättesaadavuse ja loob eeldused raviainete absorbeerumiseks (imendumiseks) vereringesse. Võime modelleerida lahustuvust inimese soolevedlikes on vajalik juba ravimiarenduse varajases faasis ja loob eeldused potentsiaalsete raviainete kandidaatide korrektseks valikuks, mis aitab parandada varajase ravimiarenduse tulemuslikkust.

Käesoleva uurimuse käigus analüüsiti kirjandust ja koostati andmeseeria 71'le mitmekülgse struktuuriga raviainele. Seda andmeseeriat modelleeriti kasutades QSAR lähenemist. Saadi hea ($R^2 \approx 0.8$) kuue molekulaartunnusega sõltuvus, mis aitab selgitada lahustuvuse ja raviainete struktuuri vahelisi seoseid. Kasutatud kahe-dimensionaalseid molekulaartunnused on lihtsasti arvutatavad. Esitatud sõltuvuses omasid kõige olulisemat kaalu molekulide suurust ja elektroonstruktuuri kirjeldavad molekulaartunnused. Saadud mudel valideeriti nii sisemiselt kui väliselt ja üksikute kõrvalekaldujate analüüs näitas, et nende mõju mudeli ennustusvõimele ei ole oluline.

5. Summary

Of the ADMET properties, the solubility in human intestinal fluids is of primary importance, which determines the bioavailability of the drug substance and creates the preconditions for the absorption of the drug substance into the bloodstream. The ability to model solubility in human intestinal fluids is needed at an early stage of drug development and creates the preconditions for the correct selection of potential drug substance candidates, which helps to improve the efficacy of early drug development.

In the present study, the literature was analyzed and a data series was compiled for 71 drug substances with variable molecular structure. This data series was modeled using the QSAR approach. A good ($R^2 \approx 0.8$) relationship with six molecular descriptors was obtained, which helped to elucidate the relationships between solubility and drug structure. The most significant descriptors take into account the size and the electron structure of the molecules. The two-dimensional molecular descriptors used are easily calculated. The resulting model was validated both internally and externally, and analysis of individual biases showed that their effect on the predictive power of the model was not significant.

Viited

1. Anderson, A. C. The Process of Structure-Based Drug Design. *Chemistry & Biology* **2003**, *10* (9), 787-797.
2. Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432*, 855-861.
3. Cumming, J., Davis, A., Muresan, S. *et al.* Chemical predictive modelling to improve compound quality. *Nat Rev Drug Discov* **12**, 948–962 (2013). <https://doi.org/10.1038/nrd4128>
4. van de Waterbeemd, H., Gifford, E. ADMET *in silico* modelling: towards prediction paradise?. *Nat Rev Drug Discov* **2**, 192–204 (2003). <https://doi.org/10.1038/nrd1032>
5. Kerns, E. H.; Di, L. *Drug-like Properties: Concepts, Structure Design and Methods*, 1st ed.; Elsevier Academic Press: USA, 2008, pp. 3, 62-67.
6. Norinder, U.; Bergström, C. A. S. Prediction of ADMET Properties. *ChemMedChem*. **2006**, *1*, 920-937.
7. Fagerberg, J. H.; Bergström, C. A. S. Intestinal solubility and absorption of poorly water soluble compounds: predictions, challenges and solutions. *Ther. Deliv.* **2015**, *6* (8), 935-959.
8. Dressman, J. B.; Vertzoni, M.; Goumas, G.; Reppas, C. Estimating drug solubility in the gastrointestinal tract. *Adv. Drug Delivery Rev.* **2007**, *59* (7), 591-602.
9. Lipinski, C.A. Drug-like properties and the causes of poor solubility and permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235-249.
10. Lipinski, C.A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337-341.
11. Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. Evaluation of Human Intestinal Absorption Data and Subsequent Derivation of a Quantitative Structure-Activity Relationship (QSAR) with the Abraham Descriptors. *J. Pharm. Sci.* **2001**, *90* (6), 749-784.
12. Elder, D.; Holm, R. Aqueous solubility: Simple predictive methods (in silico, in vitro and bio-relevant approaches). *Int. J. Pharm.* **2013**, *453*, 3-11.
13. Augustijns, P.; Wuyts, B.; Hens, B.; Annaert, P.; Butler, J.; Brouwers, J. A review of drug solubility in human intestinal fluids: Implications for the prediction of oral absorption. *Eur. J. Pharm. Sci.* **2014**, *57*, 322-332.

14. Yalkowski, S. H.; Banerjee, S. Aqueous Solubility: Methods of Estimation for Organic Compounds. *Marcel Dekker, Inc.*, New York (1992).
15. Deardem, J. C. *In silico* prediction of aqueous solubility, *Expert Opin. Drug Discovery*, **2006**, 1 (1)1, 31-52, DOI: 10.1517/17460441.1.1.31
16. Llinas, A.; Glen, R. C.; Goodman, J. M. Solubility challenge: Can you predict solubilities of 32 molecules using a database of 100 reliable measurements?. *J. Chem. Inf. Model.* **2008**, 48, 1289– 1303, DOI: 10.1021/ci800058v
17. Hopfinger, A. J.; Esposito, E. X.; Llinas, A.; Glen, R. C.; Goodman, J. M. Findings of the challenge to predict aqueous solubility. *J. Chem. Inf. Model.* **2009**, 49, 1– 5, DOI: 10.1021/ci800436c
18. Llinas, A.; Avdeef, A. Solubility Challenge Revisited after Ten Years, with Multilab Shake-Flask Data, Using Tight (SD ~ 0.17 log) and Loose (SD ~ 0.62 log) Test Sets. *J. Chem. Inf. Model.* **2019**, 59 (6), 3036-3040. <https://doi.org/10.1021/acs.jcim.9b00345>
19. Manallack D. T. The acid–base profile of a contemporary set of drugs: implications for drug discovery SAR QSAR *Environ. Res.*, **2009**, 20, 611-655, 10.1080/10629360903438313
20. Oja, M.; Maran, U., pH-permeability profiles for drug substances: Experimental detection, comparison with human intestinal absorption and modelling, *Eur. J. Pharm. Sci.*, **2018**, 123, 429-440. <https://doi.org/10.1016/j.ejps.2018.07.014>
21. Fuchs, A.; Dressman, J. B. Composition and Physicochemical Properties of Fasted-State Human Duodenal and Jejunal Fluid: A Critical Evaluation of the Available Data. *J. Pharm. Sci.* **2014**, 103, 3398-3411.
22. Abeele, J. v.d.; Rubbens, J.; Brouwers, J.; Augustijns, P. The dynamic gastric environment and its impact on drug and formulation behaviour. *Eur. J. Pharm. Sci.* **2017** 96, 207-231.
23. Fagerberg, J. H.; Karlsson, E.; Ulander, J.; Hanisch, G.; Bergström, C. A. S. Computational Prediction of Drug Solubility in Fasted Simulated and Aspirated Human Intestinal Fluid. *Pharm. Res.* **2015** 32, 578-589.
24. Holm, R.; Müllertz, A.; Mu, H. Bile salts and their importance for drug absorption. *Int. J. Pharm.* **2013**, 453 (1), 44-55.
25. Bergström, C. A. S.; Holm, R.; Jørgensen, S. A.; Andersson, S. B. E.; Artursson, P.; Beato, S.; Borde, A.; Box, K.; Brewster, M.; Dressman, J.; Feng, K. I.; Halbert, G.; Kostewicz, E.; McAllister, M.; Muenster, U.; Thinnes, J.; Taylor, R.; Mullertz, A. Early

- pharmaceutical profiling to predict oral drug absorption: Current status and unmet needs. *Eur. J. Pharm. Sci.* **2014**, *57*, 173-199.
26. Galia, E.; Nicolaidis, E.; Hörter, D.; Löbenberg, R.; Reppas, C.; Dressman, J. B. Evaluation of various dissolution media for predicting in vivo performance of class i and ii drugs. *Pharm. Res.* **1998**, *15*, 698–705.
27. Niederquell, A.; Kuentz, M. Biorelevant Drug Solubility Enhancement Modeled by a Linear Solvation Energy Relationship. *J. Pharm. Sci.* **2018**, *107*, 503-506.
28. Riethorst, D.; Mols, R.; Duchateau, G.; Tack, J.; Brouwers, J.; Augustijns, P. Characterization of human duodenal fluids in fasted and fed state conditions. *J. Pharm. Sci.* **2016**, *105* (2), 673-681. doi:10.1002/jps.24603
29. Riethorst, D.; Mitra, A.; Kesisoglou, F. Human intestinal fluid layer separation: The effect on colloidal structures & solubility of lipophilic compounds. *Eur. J. Pharm. Biopharm.* **2018**, *129*, 104-110. doi:10.1016/j.ejpb.2018.05.026
30. Riethorst, D.; Mols, R.; Duchateau, G.; Tack, J.; Brouwers, J.; Augustijns, P. Characterization of human duodenal fluids in fasted and fed state conditions. *J. Pharm. Sci.* **2016**, *105* (2), 673-681. doi:10.1002/jps.24603
31. Karelson, M. *Molecular Descriptors in QSAR/QSPR*, 2000.
32. Roy, K.; Kar, S.; Das, R. N. *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. Elsevier Academic Press: Amsterdam, The Netherlands. 2015; pp. 28-31.
33. OECD Principles for the Validation, for Regulatory Purposes. of (Quantitative) Structure-Activity Relationship Models. 2004. <http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf>.
34. *In Silico Toxicology: Principles and Applications*; Cronin, M. T. D.; Madden, J. C., Ed.; RSC Publishing: Liverpool, 2010; pp. 1-10.
35. Bergström, C.A.S.; Larsson, P. Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting. *Int. J. Pharm.* **2018**, *540*, 185-193.
36. Palmer, D. S.; Mitchell, J. B. Is experimental data quality the limiting factor in predicting the aqueous solubility of druglike molecules? *Mol. Pharm.* **2014**, *11*, 2962-2972.

37. Fagerberg, J. H.; Karlsson, E.; Ulander, J.; Hanisch, G.; Bergström, C. A. S. Computational prediction of drug solubility in fasted simulated and aspirated human intestinal fluid. *Pharm. Res.* **2015**, *32* (2), 578-589. doi:10.1007/s11095-014-1487-z
38. Niederquell, A.; Kuentz, M. Biorelevant drug solubility enhancement modeled by a linear solvation energy relationship. *J. Pharm. Sci.* **2018**, *107* (1), 503-506. doi:10.1016/j.xphs.2017.08.017
39. The PubChem Project; <https://pubchem.ncbi.nlm.nih.gov/> viimati alla laetud 23.01.2020.
40. Schaftenaar, G. *MOLDEN a visualization program of molecular and electronic structure*; <http://www.cmbi.ru.nl/molden/> viimati alla laetud 23.01.2020.
41. Macromodel, Schrödinger, LLC: New York, NY, USA, 2018.
42. Stewart, J. J. MOPAC: a semiempirical molecular orbital program. *J. Comput. Aided Mol. Des.* **1990**, *4*, 1-103.
43. Katritzky, A. R.; Petrukhin, R.; Petrukhina, I.; Lomaka, A.; Tatham, D. B.; Karelson, M. *CODESSA PRO, user's manual*; University of Florida, 2002.
44. Viira, B.; García-Sosa, A.-T.; Maran, U. Chemical structure and correlation analysis of HIV-1 NNRT and NRT inhibitors and database-curated, published inhibition constants with chemical structure in diverse datasets, *J. Mol. Graphics Modell.*, **2017**, *76*, 205-223. <https://doi.org/10.1016/j.jmglm.2017.06.019>.
45. Yap, C. W. PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32* (7), 1466-1474.
46. Roy, K.; Das, N. R. A Review on Principles, Theory and Practices of 2D-QSAR. *Curr. Drug Metab.* **2014**, *15* (4), 346-379.
47. Marvin, 2019, ChemAxon (<http://www.chemaxon.com>)
48. Lindahl, A.; Ungell, A. L.; Knutson, L.; Lennernas, H. Characterization of fluids from the stomach and proximal jejunum in men and women. *Pharm. Res.* **1997**, *14* (4), 497-502.
49. Liu, P.; Long, W. Current Mathematical Methods Used in QSAR/QSPR Studies. *Int. J. Mol. Sci.* **2009**, *10*, 1978-1998.
50. Katritzky, A. R.; Dobchev, D. A.; Tulp, I.; Karelson, M.; Carlson, D. A. QSAR study of mosquito repellents using Codessa Pro. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2306-2311.
51. J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review, *ATLA*. *33* (2005) 445-459.

52. T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J.J.M. van de Sandt, W. Tong, G. Veith, C. Yang, Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure–Activity Relationships, *ATLA*. 33 (2005) 155–173.
53. Kier, L. B.; Hall, L. H. An Electrotopological State Index for Atoms in Molecules. *Pharm. Res.* **1990**, 7, 801-807.
54. Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 1039-1045.
55. Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press: London, 1999.
56. Todeschini, R.; Consonni, V. *Molecular descriptors for chemoinformatics*. Wiley-VHC: Weinheim, Germany. 2009; pp. 714-726, 811.
57. Stanton, D. T. Evaluation and Use of BCUT Descriptors in QSAR and QSPR Studies *J. Chem. Inf. Comput. Sci.* **1999**, 39 (1), 11-20. DOI: 10.1021/ci980102x
58. Pearlman, R. S. and Smith, K. M. Novel Software Tools for Chemical Diversity. *Perspect. Drug Discovery Des.*, 1998, 9, 339-353. https://doi.org/10.1007/0-306-46857-3_18
59. Pearlman, R. S.; Smith, K. M. Metric Validation and the Receptor-Relevant Subspace Concept. *J. Chem. Inf. Comput. Sci.* **1999**, 39 (1), 28-35. DOI: 10.1021/ci980137x
60. ClinCalc. <https://clincalc.com/DrugStats/Drugs/Diclofenac>. Viimati alla laetud 23 aprill 2020.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Tanel-Sigmar Sildoja,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
„Kvantitatiivne struktuur-omadus sõltuvus raviainete lahustuvusele inimese soolevedelikes“,

mille juhendaja on Uko Maran,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, alates **01.07.2023** kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Tanel-Sigmar Sildoja
25.05.2020