TARTU ÜLIKOOL

Loodus- ja tehnoloogiateaduskond

Keemia instituut

Sofja Tšepelevitš

Arvutuslik mudel vedelik-vedelik ekstraktsiooni tulemuste ennustamiseks

Magistritöö

Juhendajad:

Prof. Ivo Leito

Karin Kipper, PhD

Joel M. Hawkins, PhD (Pfizer Inc.)

Koji Muteki, PhD (Pfizer Inc.)

Kaitsmisele lubatud .....

Juhendaja .....

allkiri, kuupäev

Tartu 2014

Sisukord

Kasutatud lühendid

COSMO-RS	–	*Conductor-Like Screening model for Realistic Solvation*
EM	–	*Expectation Maximization algoritm* – ooteväärtuste tõenäosuse maksimeerimise algoritm
LC	–	*Liquid Chromatography* – vedelikkromatograafia
LFER	–	*Linear Free Energy Relationship* – lineaarne vabaenergia sõltuvus
logD	–	jaotuskoefitsiendi logaritm (arvestades ionisatsiooni)
LOO	–	*Leave-One-Out test* – „jäta-üks-välja“ ristkontroll
PCA	–	*Principal Component Analysis* – peakomponentide analüüs
SPE	–	*Square Prediction Error* – ennustusvigade ruutude summa

Ekstraktsioonilahuste komponentide tähistused

Veefaasi lühendis tähistab iga sümbol ühte komponenti (v.a vesi, mida soola- ja happelahuste korral eraldi ei tähistata):

w	vesi
s	13% NaCl vesilahuses
x	19.7% $MgCl_2$ vesilahuses
a	6% etaanhappe lahus
c	18% sidrunhappe lahus
h	3.5% soolhappe lahus

(kontsentratsioonid on toodud massiprotsentides).

Näited:

as – vesilahus, milles on 6% etaanhapet ja 13% naatriumkloriidi;

hx – vesilahus, milles on 3.5% soolhapet ja 19.7% magneesiumkloriidi.

1. Sissejuhatus

Tänapäeva keemia- ja farmaatsiatööstuses on väga olulised protsesside efektiivsus ja tööstusmeetodite optimeerimise kiirus [1]. Nii efektiivsuse kui kiiruse tõstmisele aitab kaasa protsesside modelleerimine ning seejärel mudelil saadud tulemuste kasutamine protsesside optimeerimisel. Sobiva arvutusmetoodika olemasolul on katse tulemuse arvutuslik ennustamine kiirem, vähem kulukas ja nii keskkonnale kui töötajatele vähem kahjulik kui katse läbiviimine.

Väga tähtis protsesside rühm on faasiülekandeprotsessid, nende hulgas vedelik-vedelik ekstraktsioon. Seda kasutatakse nii huvipakkuvate komponentide (nt lõpp-produkt) isoleerimiseks kui ka segudest lisandite eemaldamiseks. Ekstraktsiooniprotseduur peab olema võimalikult selektiivne, st ideaalis peab sihtühendi eraldamine olema täielik ja muude komponentide kaasa-ekstraheerumine tühine.

On olemas mitu arvutusmetoodikat, mille abil on võimalik ekstraktsiooni tulemusi (st ühendite jaotuskoefitsiente) ennustada (nt Abrahami LFER võrrand ja COSMO-RS), kuid need metoodikad eeldavad segu komponentide keemiliste struktuuride teadmist. Kui aga tegemist on seni uurimata seguga (eksperimentaalne reaktsioonisegu, taimeekstrakt jms), ei pruugi ainete struktuurid teada olla. Praktikas on info segu koostise kohta tihti piiratud vedelikkromatograafilise analüüsi tulemustega, mis annavad ainult ligikaudse komponentide arvu ja umbmäärase hinnangu nende polaarsuse kohta.

Käesoleva töö eesmärgiks on luua arvutuslik metoodika, mis suudaks ennustada teadmata struktuuridega ainete segu ekstraktsiooni tulemusi mitmekümnes solvendipaaris kasutades sisendandmetena sama segu eksperimentaalseid ekstraktsiooni uurimise tulemusi väikeses arvus (6-10) solvendipaarides. Selline metodoloogia hõlbustaks optimaalsete ekstraktsioonitingimuste leidmist ja võimaldaks kiirendada tööstuslike protsesside optimeerimist.

Käesolev töö on teostatud firma Pfizer Inc tellimusel ja koostöös Pfizer Inc töötajatega.

2. Kirjanduse ülevaade

2.1 Solventekstraktsioon: meetodi põhimõte ja rakendused

Solventekstraktsioon, ehk vedelik-vedelik ekstraktsioon, on meetod ainete eraldamiseks nende suhtelise lahustuvuse põhjal. Lahus viiakse kontakti sellega mitteseguneva solvendiga, mille käigus toimub lahuste komponentide (nii solventide kui soluutide) jaotumine kahe faasi vahel. Seejärel lahuseid eraldatakse.

Solventekstraktsiooni kasutatakse laialdaselt nii analüütilises keemias kui tööstuses [2,3,4]. See on lihtne, odav, sobiv termolabiilsete komponentide ekstraheerimiseks ja seda on lihtne tööstuslikule mahule üle viia.

Traditsioonilisel solventekstraktsiooni protseduuril on kaks tähtsamat puudust: suurte orgaaniliste solventide koguste (paljud neist toksilised ja/või tuleohtlikud) kasutamine ja vähesed võimalused ekstraktsiooni automatiseerimiseks [5]. Tänapäeval on kasutuses mitu solventekstraktsioonil põhinevat metoodikat, kus eelmainitud puudustest on üle saadud: solventmikroekstraktsioon [6], dispersiivne solventmikroekstraktsioon [7], läbivooluekstraktsioon jt [8].

Enamik looduslike proovide analüüsimetoodikaid sisaldavad proovi puhastamise etappi, kuna looduslikud maatriksid on väga keerulised ja analüüdi kontsentratsioon neis võib olla väga madal [5,9,10]. Vaatamata mõnedele puudustele, on solventekstraktsioon prooviettevalmistuses väga laialdaselt kasutatav [5,9,10]. Õige solvendi valik võimaldab samaaegselt kontsentreerida analüüti ning eemaldada segava maatriksi mõju.

2.2 Vedelik-vedelik jaotustasakaalude modelleerimine

2.2.1 Jaotust määravad tegurid

Keemiliste ühendite jaotus erinevate keskkondade (faaside) vahel on määratud ühendite keemiliste potentsiaalide vahega kahes (või enamas) faasis. Solventekstraktsioonide korral on tegemist enamasti kahefaasiliste süsteemidega. Gibbs'i vabaenergia muut soluudi üleminekul faaside vahel on määratud soluudi keemiliste potentsiaalide vahega kahes faasis (võrrand 1).

Tasakaaluolekus on ühendi keemilised potentsiaalid mõlemas faasis võrdsed (võrrand 2). Tingimusel, et lahused on piisavalt lahjad, (st kummaski faasis ei esine soluudi molekulide vahelisi vastasmõjusid), on soluudi kontsentratsioonide suhe kahes faasis konstantne (võrrand 3).

$$\Delta G^{S}_{A\rightarrow B} = \mu^{S}_{B} - \mu^{S}_{A} \quad (1)$$

$$\mu_A = \mu^{\infty}_{A} + RT\ln a_A = \mu_B = \mu^{\infty}_{B} + RT\ln a_B \quad (2)$$

$$K_{A\rightarrow B} = \frac{[S]_B}{[S]_A} = \exp\left(\frac{\mu^{\infty}_{B} - \mu^{\infty}_{A}}{RT}\right) \quad (3)$$

Kus

$\Delta G_{A\rightarrow B}$ – soluudi S vabaenergia muut üleminekul faasist A faasi B.

μ^{S}_{A}, μ^{S}_{B} – soluudi S keemilised potentsiaalid vastavalt faasides A ja B.

μ^{∞}_{A} , μ^{∞}_{B} – soluudi s keemilised potentsiaalid lõpmata lahjas lahuses (standardpotentsiaalid), vastavalt faasides A ja B.

a_A, a_B – soluudi S aktiivsus vastavalt faasides A ja B.

$[S]_A$, $[S]_B$ – soluudi S tasakaaluline kontsentratsioon vastavalt faasides A ja B.

$K_{A\rightarrow B}$ – faasidevahelise üleminekuprotsessi tasakaalukonstant (jaotuskoefitsient).

Tegurid, mis määravad ühendite lahustumise ning jaotumise mitme solvendi vahel on suuresti samad [2]. Neutraalsete molekulide korral on need nii molekuli enda kui ka lahusti komponentide polaarsus (üksiku molekuli dipoolmomendi tähenduses), polariseeritavus, vesiniksideme donoorne ja aktseptoorne võime; ioonsete osakeste käitumist mõjutavad lisaks tugevasti lahuste ioonjõud ja dielektriline läbitavus [11,12].

Molekulide jaotuse ennustamiseks eri keskkondade vahel kasutatakse enamasti andmeid sarnaste molekulide jaotumise kohta samades või sarnastes solventides [2].

2.2.2 Abrahami LFER mudel

Michael Abraham ja tema kaastöölised on mitmes töös näidanud, et neutraalsete molekulide lahustuvus erinevates lahustites ja nende jaotumine kahe vedeliku või orgaanilise solvendi ja gaasi vahel on kirjeldatav järgmise lineaarse vabaenergia sõltuvuse võrrandiga [13,14]:

$$SP = c + eE + sS + aA + bB + vV \quad (4)$$

Kus

SP – soluudi omadus, antud juhul jaotuskoefitsiendi logaritm.

E – *excess molar refraction*, molekuli polariseeritavuse iseloomustaja: sisuliselt molekuli polariseeritavuse ja võrdse molekuli ruumalaga alkaani polariseeritavuse vahe [14].

S – soluudi polaarsust (molekuli dipoolmomenti) ja polariseeritavust kirjeldav parameeter.

A – soluudi üldine vesiniksideme donoorsus.

B – soluudi üldine vesiniksideme aktseptoorsus.

V – soluudi McGowani molekulruumala.

a, b, c, e, ja v on solvendile omistatud koefitsiendid (määratud multilineaarse regressiooni meetodil).

Abrahami koefitsientide süsteem on koostatud eksperimentaalselt määratud gaasi- ja vedelikkromatograafilise retentsiooni kirjeldavate parameetrite ja ühendite jaotuskoefitsientide (vee ja mõne orgaanilise solvendi vahel) alusel [14].

Võrrand 4 võimaldab kirjeldada ja ennustada ühendite jaotumist vee ja mitme orgaanilise solvendi ning ioonvedeliku vahel [15]. Selle täiendamisel kahe ionisatsioonikonstantidest tuletatud liikmega (üks katioonide ja üks anioonide jaoks) on võimalik kirjeldada ka ioonide jaotumist, kuigi saadud tulemused on mõnevõrra madalama kvaliteediga kui neutraalsete molekulide korral [15]. Võrrandit 4 on muuhulgas kasutatud ka rakumembraanide läbimisega seotud biokeemiliste protsesside kirjeldamiseks [16,17].

Eelmainitu tõestab, et ühendite jaotus eri faaside vahel on edukalt kirjeldatav suhteliselt väikese arvu (5-7) deskriptorite abil. Samas on näidatud, millised molekuli omadused peavad olema arvesse võetud ekstraktsiooni edukaks modelleerimiseks.

2.2.3 COSMO-RS

COSMO-RS (*Conductor-Like Screening Model for Realistic Solvation*) [18,19] on praegusel ajal üks levinumaid meetodeid vedelate segude termodünaamiliste omaduste ennustamiseks. Meetodis kasutatakse väga efektiivselt kvantkeemia ja statistilise termodünaamika kombinatsiooni. Esimese etapina optimeeritakse kvantkeemia meetoditega molekuli või iooni (COSMO-RS meetodis käsitletakse ioone samal viisil kui neutraalseid molekule) geomeetria ja arvutatakse laengujaotus selle pinnal, väljendatuna sigma-profiilina (osalaeng vs vastav pinnaosa). Teise etapina arvutatakse segu komponentide sigma-profiilide abil nende vastasmõjude tugevused. Molekulidevahelised vastasmõjud taanduvad selles lähenemises laetud pinnasegmentide paaride elektrostaatilistele vastasmõjudele. Vastasmõju energia komponendid (elektrostaatiline, vesiniksidemed jne) esitatakse funktsioonina kahe pinnasegmendi laengutest ning aatomile ja vastasmõju tüübile vastavatest lisaparameetritest. Statistilise termodünaamika abil saab leida segu kõikide komponentide keemilised potentsiaalid, millest omakorda võib arvutada mitmesuguseid termodünaamilisi parameetreid, sealhulgas tasakaalukonstante. Molekulide vaheliste vastasmõjude arvestamine COSMO-RS meetodis on arvutusmahu seisukohast väga efektiivne, kuid kätkeb endas ka puudusi: vastasmõjude arvestamine on lihtsustatud, osaliselt jäävad arvestamata steerilised efektid ja täielikult jäävad arvestamata kaug-vastasmõjud.

COSMO-RS meetod on poolempiiriline, kuid ei ole parametriseeritud ühegi konkreetse solvendi ega aineklassi jaoks. Empiirilised parameetrid seotakse mitte konkreetsete ainete või funktsionaalrühmade, vaid keemiliste elementide, vastasmõju tüüpide (nt vesiniksideme tugevuse sõltuvus temperatuurist) ja molekulide suuruse ja kujuga. Seetõttu saab seda arvutusmeetodit kasutada täiesti uute soluutide ja seniuurimata solventide omaduste ennustamiseks. COSMO-RS teiseks suureks eeliseks enamiku muude tänapäeval kasutatavate lahuste modelleerimise meetodite ees on võimalus uurida suvalise koostisega segusid. Komponentide arv, kontsentratsioonid ja osakeste laengud sisuliselt ei ole piiratud, kuigi ioonidega (iseäranis lokaliseeritud laengutega väikeste ioonide puhul) annab COSMO-RS

madalama kvaliteediga tulemusi kui neutraalidega [19,20]. Selle üheks põhjuseks on nn „väljaulatuvast“ laengutihedusest põhjustatud viga (*outlying charge error, OCE*) [19].

COSMO-RS kasutatakse edukalt vedelik-aur faasidiagrammide koostamiseks, solventekstraktsiooni modelleerimiseks, lahustuvuse ja muude termodünaamiliste parameetrite arvutamiseks [19,21,22,23,24,25]. Eksperimentaalsete andmete kvantitatiivseks reprodutseerimiseks võib olla vajalik kasutada empiirilisi parandusi, kuid kvalitatiivseid sõltuvusi annab meetod reeglina hästi edasi. Selle abil on õnnestunud ka väljasoolamisefekte kvalitatiivselt ennustada [22,26].

2.3 Ülevaade kasutatud andmetöötlusmeetoditest

2.3.1 Peakomponentide analüüs

Peakomponentide analüüs (PCA, P*rincipal Component Analysis*) [27,28] on statistiline meetod, mille abil andmemaatriksi esialgsed muutujad (enamasti omavahel korreleeruvad) muundatakse uuteks ortogonaalseteks muutujateks (peakomponentideks). Peakomponendid kujutavad endast lineaarseid kombinatsioone esialgsetest muutujatest. Leitud peakomponendid moodustavad uue koordinaadistiku (teljestiku). Koefitsiente, mis kirjeldavad primaarsete muutujate panuseid peakomponentidesse, nimetatakse laadungiteks (*loadings*). Algsete objektide koordinaate uues teljestikus nimetatakse skoorideks (*scores*). Eukleidese kaugused objektide vahel koordinaatsüsteemi teisendamise käigus ei muutu.

Peakomponentide analüüsi põhimõtet kirjeldavad võrrandid 5-8:

$$X \rightarrow T \cdot P^T \quad (5)$$

$$\hat{X} = T_{1:k} \cdot P_{1:k}^T \quad (6)$$

$$\hat{X} - X = E \quad (7)$$

$$X_{ij} = \sum_{a=1}^{k} T_{ia} \cdot P_{aj} + E_{ij} \quad (8)$$

Kus on kasutatud järgmisi tähistusi:

X – esialgne andmemaatriks;

T – skooride maatriks;

P – laadungite maatriks (P^T – transponeeritud laadungite maatriks);

$\widehat{X}$ – rekonstrueeritud andmestik;

k – rekonstrueerimiseks kasutatud peakomponentide arv;

E – jääkliikmete maatriks.

PCA põhiomaduseks on uute telgede suurem informatiivsus võrreldes esialgsetega. Seejuures peakomponendi infotihedus väheneb peakomponendi järjekorranumbri kasvuga. Tihti piisab vaid mõnest esimesest peakomponendist, et hästi kirjeldada andmestiku struktuuri ja väärtuste varieeruvust [27]. Tänu sellele kasutatakse PCA meetodit laialdaselt mahukate paljumõõtmeliste andmestike andmemahu vähendamiseks, visualiseerimiseks, interpreteerimiseks ja ennustuste tegemiseks.

Esialgset andmemaatriksit on võimalik rekonstrueerida skooride ja laadungite põhjal (võrrandid 6 ja 7). Üksiku elemendi rekonstrueerimist kirjeldab võrrand 8. Andmete rekonstrueerimine on seda täpsem, mida rohkem peakomponente on selleks kasutatud. Kõikide peakomponentide kaasamisel saadakse esialgne andmemaatriks.

2.3.2 EM algoritm

Nii teadustöös kui tööstuslike protsesside jälgimisel esineb sageli olukordi, kus suuremahulise uurimisobjekti kõiki vajalikke parameetreid ei ole võimalik eksperimentaalselt määrata [29,30] ja seetõttu jääb uuringu käigus kogutud andmestik mittetäielikuks. Andmestikust puuduvad väärtused takistavad mitmete statistiliste meetodite (peakomponentide analüüs, multilineaarne regressioon, vähimruutude meetod jt) rakendamist. Nende asendamine nullide või andmekogumi keskmiste väärtustega ei ole enamasti vastuvõetav, kuna võib tulemust märgatavalt mõjutada. Ooteväärtuste tõenäosuse maksimeerimise algoritm (*Expectation Maximization algoritm,* edaspidi EM) [29,30,31] on üks meetoditest, mida kasutatakse statistiliste mudelite loomisel mittetäielike andmestike alusel. EM algoritm võimaldab puuduvad väärtused asendada maksimaalse tõenäosusega ooteväärtustega.

EM algoritm on iteratiivne ning koosneb kahest vahelduvast sammust. E-sammus (*expectation*) arvutatakse andmestiku (teadaolevad väärtused + puuduvate andmepunktide hinnangud) parameetritest puuduvate elementide tõenäosusfunktsioonid. Tõenäosusfunktsiooni abil leitakse puuduvate väärtuste uued hinnangud. M-sammus (*maximization*) arvutatakse eelmises sammus uuendatud andmestiku parameetrid, mida omakorda kasutatakse järgmises E-sammus. Algoritmi korratakse kuni mudeli koondumiseni.

Nii E- kui M-sammu teostamiseks võib sõltuvalt konkreetsest ülesandest kasutada erinevaid matemaatilisi meetodeid, nt PCA, Monte-Carlo [32] ja maksimaalse entroopia [33] meetodid. Kemomeetrias kasutatakse EM algoritmis enamasti peakomponentide analüüsi. EM-PCA algoritmi üldskeem on järgmine [29,31]:

1. Puuduvad väärtused asendatakse esialgsete hinnangutega.
2. (M-samm) Andmestikule rakendatakse PCA, arvutatakse laagungid ja skoorid.
3. (E-samm, 1) Esialgne andmemaatriks rekonstrueeritakse valitud arvu peakomponentide alusel.
4. (E-samm, 2) Puuduvad elemendid esialgses andmestikus asendatakse sammus 3 arvutatud väärtustega.

Samme 2-4 korratakse kuni mudeli koondumiseni. Vajadusel lisatakse andmete teisendamise (skaleerimine, tsentreerimine) ja algkujule viimise sammud. Optimaalne peakomponentide arv andmestiku rekonstrueerimiseks sammus 3 leitakse ristvalideerimise abil.

Algoritmi rakendatavus ja saadavate tulemuste kvaliteet sõltuvad andmestiku puuduvate elementide omavahelisest paiknemisest, elemendi puudumise tõenäosuse seosest sellesama või muude elementide väärtustega ja puuduvate väärtuste korrelatsioonist olemasolevatega [29]. Paljudel juhtudel on EM-PCA algoritmi abil võimalik mittetäielike andmestike alusel luua PCA mudeleid, mis praktiliselt ei erine täielike andmestike alusel loodud mudelitest [31]. Seejuures puuduvate andmete osakaal võib olla kuni 30% [31].

2.3.3 Iteratiivne tingväärtuste meetod

Kui teatud andmekogumi põhjal on loodud PCA mudel, siis selle abil on võimalik ennustada uute objektide jaoks mõningate olemasolevate omaduste põhjal uusi omadusi. Eeldusteks on uue

objekti kuuluvus samasse kogumisse (*population*) kui PCA mudeli moodustavad objektid ja vähemalt osa uue objekti omaduste kättesaadavus.

Ennustamiseks võib rakendada mitut erinevat meetodit [34]. Käesolevas töös kasutatakse iteratiivset tingväärtuste meetodit (*iterative imputation method*) [34].

Selle meetodi algoritmi etapid on järgmised:

1. Luuakse uuele objektile vastav vektor (X); puuduvad omaduste väärtused asendatakse esialgsete hinnangutega.
2. Arvutatakse skooride hinnangud (T) kasutades PCA mudeli *k* esimese peakomponendi laadungeid (P): $T = X \cdot P_{1:k}$
3. Uue objekti vektor rekonstrueeritakse uute skooride ja olemasoleva mudeli laadungite abil: $\widehat{X} = T \cdot (P_{1:k})^T$
4. Puuduvad väärtused esialgses vektoris asendatakse sammus 3 arvutatud hinnangutega.

Samme 2-4 korratakse kuni koondumiseni.

Ennustamise kvaliteet sõltub puuduvate omaduste tähtsusest PCA mudelis: mida suuremad on vastavate laadungite väärtused, seda viletsam on ennustus.

3. Arvutuslik osa

3.1 Uue ennustava mudeli kavand

Käesoleva töö eesmärgiks oli luua arvutuslik metodoloogia ekstraktsioonitingimuste valimise hõlbustamiseks olukorras, kus on vaja eraldada huvipakkuv aine lisanditest ning lisandite struktuurid ei ole teada (nt katseline reaktsioonisegu või uurimata taimeekstrakt). Tihti on praktikas võimalik segu komponentide arvu ja polaarsust ainult ligikaudselt hinnata vedelikkromatograafilise analüüsi kaudu (pöördfaaskromatograafias on aine retentsiooniaeg seotud aine hüdrofoobsusega). Üldjuhul saab huvipakkuvate komponentide eraldamiseks segust rakendada sadu orgaanilise solvendi ja veefaasi lisandite kombinatsioone. Katseliselt selgitada, millised tingimused annavad parima tulemuse, võib olla väga aja- ja materjalikulukas.

Probleemi lahendamiseks seati eesmärgiks luua arvutuslik mudel, mille abil saaks ennustada ainete jaotuskoefitsiente mitmekümnes etteantud ekstraktsioonitingimuste komplektis väikese arvu (6-10) eksperimentaalselt määratud (kasutades ainult vedelikkromatograafia aparatuuri) logD väärtuse alusel, kusjuures ainete struktuure ei oleks vaja teada. See vähendaks oluliselt sobivate ekstraktsioonitingimuste leidmiseks vajaminevaid ressursse ja võimaldaks kiirendada ning hõlbustada uurimistööd ja tööstuslike protsesside optimeerimist.

Käesoleva töö käsitlusala on piiratud neutraalsete ja aluseliste ühendite ning neutraalsete ja happeliste vesilahustega.

Töö etapid:

1. Treenimisandmestiku koostamine:
   valida ühendid, mille omadused kataksid võimalikult ühtlaselt kogu ekstraktsiooni kontrollivate faktorite ruumi, ja solvendipaarid, kus realiseeruksid võimalikult mitmekesised tasakaalude mustrid.
2. Jaotustasakaalude eksperimentaalne uurimine:
   määrata eksperimentaalselt treenimisandmestiku ühendite logD väärtused valitud solvendipaarides.
3. Statistiline analüüs:

uurida statistiliste meetodite abil seoseid logD väärtuste vahel erinevates ekstraktsioonitingimustes, ja tuvastada 6-10 solvendipaari, millele vastavate logD väärtuste alusel võib kõige täpsemalt ennustada aine käitumist ülejäänud solvendipaarides.

4. Ennustava mudeli valideerimine treenimisandmestikku mittekuuluvate ainete abil.

3.2 Treenimisandmestiku koostamine COSMO-RS abil

Esialgselt mudeli koostamiseks väljapakutud solvendipaare kajastab Tabel 1.

Tabel 1. Esialgselt väljapakutud solvendipaaride komponendid.

Orgaanilised sovendid [a,b]	**Happed**[b,c]	**Soolad**[b,c]
1. Tolueen	Etaanhape (6%)	
2. (Tolueen-heptaan 1:1)	Sidrunhape (18%)	NaCl (13%)
3. 2-metüültetrahüdrofuraan	Soolhape (3.5%)	$MgCl_2$ (19.7%)
4. Isopropüülatsetaat	(Metaansulfoonhape)	
5. (Metüülisobutüülketoon)		
6. Butanool		
7. (Heptaan-tolueen-metanool 5:4:1)	Vesi	
8. Tolueen-metanool 9:1		
9. (Isopropüülatsetaat-isopropanool 9:1)		
10. (2-metüültetrahüdrofuraan-metanool 9:1)		
11. (Metüülisobutüülketoon-metanool 19:1)		
12. Diklorometaan		

[a] Segasolventide koostised on väljendatud ruumalaosades.
[b] Sulgudesse on võetud solvendipaaride komponendid, mida eksperimentaalselt ei uuritud.
[c] Komponentide kontsentratsioonid vesilahustes on toodud massiprotsentides.

Orgaaniliste solventide valiku tingis nende kasutatavus farmaatsiatööstuses ja võimalikult madal mürgisus. Samal ajal sooviti saada võimalikult mitmekesiseid lahustuvus- ja ekstraktsiooniomadusi.

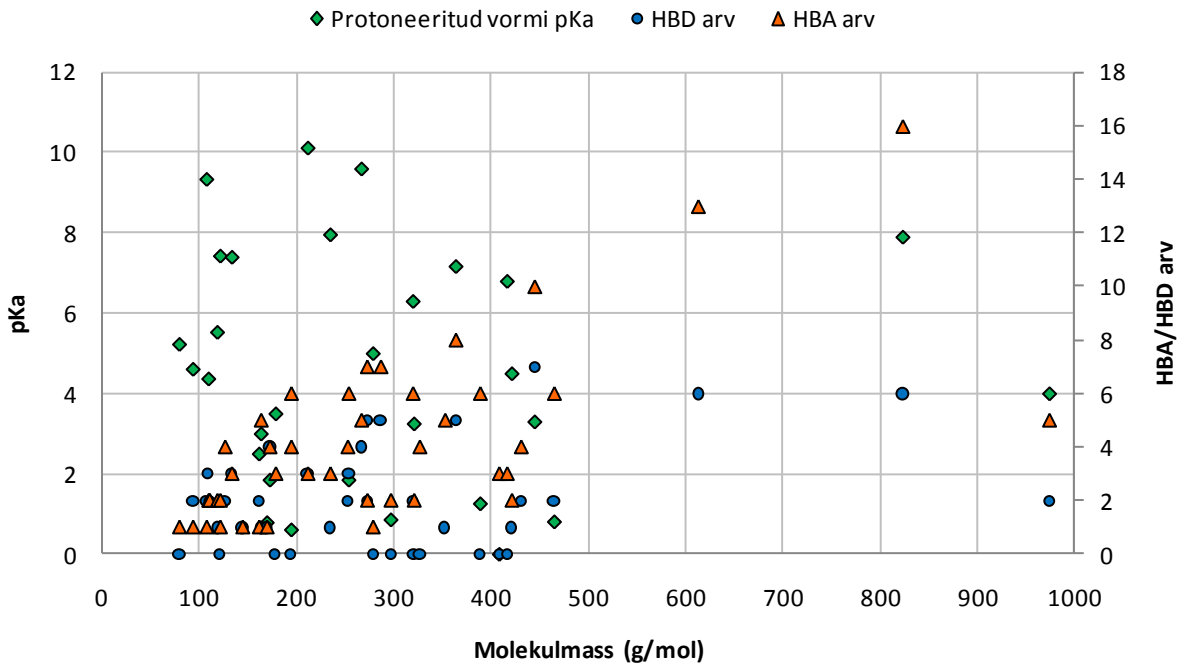
Analüüdid pidid vastama järgmistele kriteeriumidele:

1. Molaarmass 100…800 $g \cdot mol^{-1}$.
2. Kromofoori(de) olemasolu molekulis (enamik eksperimente oli plaanis teostada UV-Vis detektoriga kõrgefektiivse vedelikkromatograafia (HPLC) instrumendi abil).
3. Madal reaktsioonivõime kasutatavates tingimustes.
4. Stabiilsus õhu, vee ja valguse suhtes (analüüdi lahus stabiilne vähemalt ööpäeva jooksul).
5. Kasutamine ei ole seadusega piiratud.
6. Saadavus vastuvõetava puhtuse ja hinnaga.

Treenimisandmestikku valiti keemiliselt võimalikult mitmekesised molekulid, mis lisaks sisaldaksid ravimomadustega ühendites tihti esinevaid funktsionaalseid rühmi ja struktuurseid fragmente. Esialgses ainete nimekirjas esinesid ka molekulid, mis ei vasta kõikidele ülaltoodud kriteeriumidele. Samuti muutus treenimisandmestik eksperimenditöö käigus, kui mõnede ühendite jaotuskonstantide määramisega tekkisid praktilised raskused.

Treenimisandmestiku mitmekesisuse esmaseks hindamiseks kasutati mõnede ekstraktsiooni mõjutavate molekulide omadusi kajastavaid graafikuid (näide Joonisel 1). Tähtsaimad ühendi omadused, mis mõjutavad selle jaotumist eri faaside vajhel, on happelisus/aluselisus, vesiniksideme doonorite ja aktseptorite olemasolu ja arv, molekuli suurus, dipoolmoment ja polariseeritavus. Nende parameetrite väärtused pidid treenimisandmestiku molekulidel võimalikult laias ulatuses varieeruma ja esinema erinevates kombinatsioonides.

Joonisel 1 esindatud ainete kogumi mitmekesisus leiti sobivaks, kuigi molekulide omaduste kombinatsioonide jaotus ei ole ühtlane. Suurema osa esialgsest treenimisandmestikust moodustasid ravimitaolised ühendid [35]. Enamiku ühendite molekulmassid on vahemikus 100-500 $g \cdot mol^{-1}$ ning vesiniksideme doonorite arv on korrelatsioonis molekulmassiga. Paljud ühendid, mis suurendaksid omaduste kominatsioonide varieeruvust, ei sobinud treenimisandmestikku praktilistel kaalutlustel.

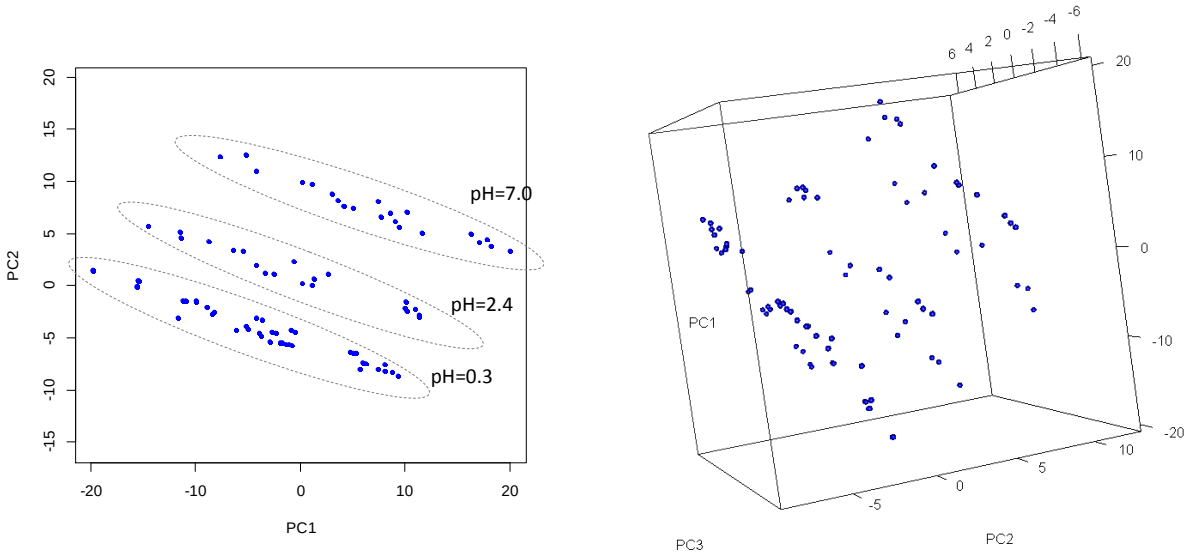


Joonis 1. Ekstraktsiooni mõjutavate molekulide omaduste jaotus lõplikus treenimisandmestikus. Igale ainele vastab skeemil kolm eri omadusi tähistavat punkti.

Treenimisandmestiku mitmekesisuse täpsemaks hindamiseks kasutati ekstraktsioonide modelleerimist COSMO-RS meetodiga (protseduuri täpsem kirjeldus ja kasutatud parameetrid on toodud Lisas 1). COSMO-RS võimaldab arvutada jaotustasakaalusid suvalise koostisega lahuste vahel ja võtta arvesse orgaanilise ja veefaasi vastastikkust küllastumist. Paljude spetsiifiliste intermolekulaarsete vastasmõjude (kelateerumine, dimerisatsioon, püsiva vesiniksidemega struktuurid) modelleerimiseks ei piisa meetodi statistilise termodünaamika osast, ning tekkivaid komplekse tuleb modelleerida kvantkeemiliselt. Käesoleva töö kontekstis oleks see lähenemine liiga aja- ja töökulukas. Seetõttu ei modelleeritud COSMO-RS abil solvendisüsteeme, kus toimub suure tõenäosusega analüütide komplekseerumine või kelateerimine. Need on solvendipaarid, mille veefaas sisaldab sidrunhapet ja/või magneesiumkloriidi. Selle asemel jälgiti, et treenimisandmestikus oleksid esindatud erineva komplekseerumis- ja kelateerimisvõimega ühendid.

On teada, et COSMO-RS arvutuste tulemused ei ühti alati eksperimendiandmetega, kuid on enamasti nendega korrelatiivses sõltuvuses [20]. Oletati, et peakomponentide analüüsi rakendamisel COSMO-RS tulemustele (andmete eelneva skaleerimise ja tsentreerimisega) ei ole meetodi süstemaatilisel veal suurt tähtsust, kuna käesoleval juhul pole COSMO-RS arvutuste eesmärk jaotuskoefitsientide täpne ennustamine vaid treenimisandmestiku mitmekesisuse hindamine.

COSMO-RS abil modelleeriti 130 ühendi jaotumist 96 solvendipaaris. Seejärel rakendati andmestikule peakomponentide analüüsi, et hinnata analüütide ja solvendipaaride mitmekesisust. Sarnaste objektide kogumitest valiti kõige sobivamad: analüütide korral laboratoorse töö, solventide korral tööstusliku protsessi seisukohalt. Puhtaid solvente eelistati segasolventidele mugavuse kaalutlustel.



Joonis 2. Solvendipaaride võrdlus PCA abil. Vasakul on esimese kahe peakomponendi skooride graafik, paremal - 3 esimese peakomponendi skooride 3-mõõtmeline graafik.

Solvendisüsteemide sarnasust hinnati PCA tulemuste ja logD väärtuste vaheliste korrelatsioonide alusel. Kindlat sarnasuse kriteeriumit (nt R^2 piirväärtuse kujul) sel juhul ei kasutatud. Joonis 2 illustreerib solvendisüsteemide peakomponentide analüüsi tulemusi. Punktid graafikutel vastavad erinevatele solvendipaaridele. Vasakul on esimese kahe peakomponendi skooride graafik, paremal - 3 esimese peakomponendi skooride 3-mõõtmeline graafik. Solvendipaarid on mõlemal graafikul rühmitatud veefaasi pH järgi. Suurim rühm (pH väärtusega 0.3) sisaldab soolhapet ja

metaansulfoonhapet sisaldavate veefaasidega solvendipaare. 3-mõõtmeliselt graafikult võib näha, et 3 punktide rühma moodustavad praktiliselt paralleelseid tasandeid.

Modelleerimise tulemusena leiti, et:

1. Veefaaside omavahelised erinevused on põhjustatud peamiselt nende pH väärtustest (Joonis 2). Välja jäeti metaansulfoonhappe vesilahus, kuna see oli ekstraktsiooniomaduste poolest praktiliselt identne HCl vesilahusega (kuid HCl on kättesaadavam ja stabiilsem hape).
2. Tugevate hapete ja soolade olemasolu veefaasis võib nivelleerida erinevusi orgaaniliste solventide vahel.
3. Äädikhapet sisaldavad veefaasid kalduvad esile tooma erinevusi nii soluutide kui ka solventide vahel.
4. Mõnede orgaaniliste solventide ekstraktsiooniomadused on väga sarnased, näiteks:
   - Tolueen ja heptaan-tolueen 1:1 segu, eriti kui veefaas sisaldab soola või tugevat hapet.
   - Isopropüülatsetaat ja metüülisobutüülketoon (MIBK), sõltumata veefaasist. Nendest kahest lahustist valiti eksperimentaalseks uurimiseks isopropüülatsetaat, kuna seda kasutatakse Pfizeri tööstuses rohkem.
   - Isopropüülatsetaat ja isopropüülatsetaat - isopropanool 9:1 segu.
   - 2-metüülTHF-metanool 9:1 ja MIBK-metanool 19:1, praktiliselt sõltumata veefaasist.

Arvestades modelleerimise tulemuste ja praktiliste kaalutlustega (nii arvutusliku mudeli loomise kui ka järgneva rakendamise seisukohalt) valiti eksperimentide läbiviimiseks 72 solvendipaari (toodud Tabelis 1 ja Lisas 4).

4. Eksperimentaalne töö

Kasutatud aparatuuri, töövahendite ja kemikaalide täpsemad karakteristikud on toodud Lisades: kemikaalid: Lisa 2; aparatuur ja töövahendid: Lisa 3.

4.1 Eksperimendimetoodika

Eksperimendimetoodika põhineb protseduuril, mis on kirjeldatud firma Agilent oktanool-vesi jaotuskoefitsientide vedelikkromatograafilise määramine juhendis [36].

Vedelik-vedelik ekstraktsioon viidi läbi standardsetes 2 ml valtskorkidega LC automaatsisestusseadme viaalides. Viaalidesse pipeteeriti ligikaudu võrdsed kogused orgaanilist lahustit ja vesilahust (puhtana või nendes lahustatud analüütidega). Lahustite kogused varieerusid vahemikus 0.5 ml kuni 0.7 ml. Optimaalne kogus (enimkasutatud) on 0.6 ml kumbagi vedelikku, kuna see jätab viaali piisavalt vaba ruumi (*headspace*) faaside efektiivseks segunemiseks loksutamisel.

Ekstraheerimiseks pandi tihedalt suletud viaalid vahtplastist hoidjasse ja asetati horisontaalselt orbitaalse loksuti peale. Loksutati vähemalt 2 tunni jooksul sagedusega 250 ringi minutis.

Pärast loksutamist analüüsiti mõlemat vedelfaasi vedelikkromatograafiliselt. Püsivaid emulsioone lõhuti tsentrifuugimise teel (viaali avamata). Kromatograafiliseks analüüsiks võeti proovid ülemisest ja alumisest solvendikihist, seadistades selle tarvis vastavalt automaatsisestusseadme nõela liikumist proovivõtmisel. logD väärtused arvutati ainetele vastavate piikide pindalade suhetest, võttes arvesse süstide ruumalade erinevust orgaanilise- ja veefaasi analüüsil (võrrand 9):

$$\log D = \log\left(\frac{A_{org} \cdot V_v}{A_v \cdot V_{org}}\right) \quad (9)$$

Kus

A_{org}, A_v – analüüdi piigi pindalad vastavalt orgaanilises ja vesifaasis;

V_{org}, V_v – orgaanilise ja vesifaasi süstide ruumalad.

Suurem osa logD määramise katseid viidi läbi analüütide rühmadega (enamasti 6-9 ainet samasse segusse lahustatult). See vähendab ajakulu võrreldes iga aine jaoks eraldi logD määramisega, võimaldab hinnata metoodika rakendatavust reaalsete ülesannete lahendamisel (mil segudes on enamasti mitmeid aineid) ning tuvastada võimalike probleeme ja metoodika nõrku kohti.

Kasutati analüütide lahuseid ligikaudse kontsentratsiooniga 1-1.5 mg/ml (arvutatud katseks võetud ainete kogumassist ja lahustite koguruumalast). Igal mõõtmispäeval tehti analüütidest värsked lahused ja neid kasutati koheselt. Kogu protseduur (ekstraheerimine + kromatograafiline analüüs) viidi läbi võimalikult kiiresti (enamikul juhtudel <24 tunni jooksul), et vähendada võimalike reaktsioonide mõju tulemustele.

Uuritavate ainete ülekandmiseks ekstraktsiooniviaalidesse kasutati järgmiseid tehnikaid (eraldi või kombinatsioonis):

- Tahked analüüdid kaaluti klaas- või plastanumasse, lisati vajalik kogus lahustit (orgaanilist solventi või vesilahust) ja vedelad analüüdid (nende olemasolul). Vajadusel kasutati ainete lahustumise kiirendamiseks ultrahelivanni. Saadud lahust (või peent suspensiooni) segati ja jaotati automaatsisestusseadme viaalidesse.
- Ained lahustati väikeses koguses võimalikult lenduvas solvendis (diklorometaan hüdrofoobsete, metanool või metanool-diklorometaan segu hüdrofiilsete ühendite jaoks). Lahus pipeteeriti ekstraktsiooniviaalidesse ja lahustid aurustati lämmastikuvoolus.
- Ained jaotati ekstraktsiooniviaalidesse tahkel kujul.

Veefaaside lahuseid (ilma analüütideta) säilitati pimedas, tihedalt suletud anumates, toatemperatuuril kuni 4 kuud.

4.2 Võimalikud tulemust mõjutavad faktorid ja veaallikad ning meetmed nende vastu

Järgnevas osas vaadeldakse võimalikke mõõtmise tulemust mõjutavaid faktoreid ja veaallikaid, kirjeldatakse meetmeid nende kõrvaldamiseks ja hinnatakse nende mõju analüüsi tulemustele.

4.2.1 Loksutusaja mõju tulemustele

Jaotustasakaalude uurimisel soovitatakse tasakaalu saabumise kiirendamiseks kasutatavaid solvente vastastikku küllastada enne neile analüütide lisamist [37]. Käesolevas töös seda ei tehtud, kuna solvendipaaride kombinatsioonide arvukuse tõttu oleks see olnud ebapraktiline. Selle kompenseerimiseks valiti pikk loksutusaeg. Loksutusaja piisavust uuriti katseliselt. 8 erineva hüdrofiilsusega analüüti lahustati kahes orgaanilises solvendis (butanool ja tolueen – mõõdukalt hüdrofiilne ja väga hüdrofoobne), jaotati 0.6 ml kaupa automaatsisestusseadme viaalidesse, lisati igasse 0.6 ml vesilahust (puhas vesi või 1 M HCl) ning loksutati erineva aja jooksul (15 min, 30 min, 1 tund ja 1.5 tundi). Osa tulemustest (logD kujul) on esitatud Tabelis 2.

Tabel 2. Loksutusaja mõju eksperimendi tulemusele.

Solvendipaar→	**vesi – butanool**		**1M HCl – butanool**		**vesi - tolueen**	
Analüüt↓	*15 min*	*1.5 t*	*15 min*	*1.5 t*	*15 min*	*1.5 t*
3-aminofenool	0.49	0.50	-0.17	-0.16	-1.40	-1.42
Resortsinool	1.00	1.00	1.03	1.04	-2.01	-2.02
Kofeiin	0.25	0.25	-0.07	-0.07	-0.25	-0.26
Aniliin	0.89	0.92	---	---	0.87	0.88
Diantipüriilmetaan	1.71	1.67	0.73	0.73	1.21	1.19
1-naftool	2.50	2.36	2.39	2.35	1.82	1.82
Difenüülamiin	2.99	2.72	2.09	2.07	---	---
Mikonasool	3.35	3.07	2.64	2.58	3.30	3.42

Tabelist 2 võib näha, et kui $|\log D|<0.8$, saabub tasakaal juba esimese 15 minuti jooksul. logD muut järgneva tunni jooksul ei ületa 0.01 ühikut, mis ei ole antud juhul statistiliselt oluline. Kõrgema logD absoluutväärtuse korral kuulub tasakaalu saabumiseni rohkem aega. 1.5-tunnise loksutusajale vastavate logD väärtuste võrdlus 2-tunnise loksutusajale vastavate tulemustega näitab, et nendevahelised erinevused ei ole statistiliselt olulised. Käesoleva katseseeria tulemused näitasid, et loksutusaeg 2 tundi on kõigi katsetamiseks valitud analüüt-solvendipaar kombinatsioonide jaoks piisav.

4.2.2 Analüütide poolt põhjustatud veefaasi pH muutused

Käesolevas töös kasutatavate analüütide hulgas oli mitmeid mõõdukalt kuni tugevalt aluselisi ühendeid, mille sisaldumine veefaasis võib tõsta selle pH väärtust ja seeläbi mõjutada analüütide jaotustasakaale. Eeldati, et happeliste veefaaside korral võib pH muudu arvestamata jätta, kuna

nende veefaaside puhverdusvõime on tänu lisatud happele piisavalt kõrge, küll aga võib pH muutus olla oluline puhta vee ning soolade vesilahuste korral. Analüütide mõju hindamiseks neutraalsete veefaaside korral mõõdeti nende pH-d klaaselektroodi abil vahetult pärast valmistamist ja peale ekstraheerimise protseduuri. Paljudel juhtudel täheldati pH tõusu 1-2 ühiku võrra (tugevalt aluseliste ühendite juuresolekul) ja vähestel juhtudel pH mõningat langust (0.2-1 ühiku võrra, võimalikuks põhjuseks on ebasoovitavad reaktsioonid lahuses).

pH muutuste ärahoidmiseks puhverdati happeid mittesisaldavaid veefaase fosfaatpuhvriga. Vee ja 13% NaCl lahuse korral kasutati puhvrit ligikaudse kontsentratsiooniga 0.08 mM ($[H_2PO_4^-]/[HPO_4^{2-}]$ ~1.6). Fosfaatide lahustuvus 19.7% $MgCl_2$ lahuses osutus valitud fosfaatide koguse kasutamiseks ebapiisavaks ning $MgCl_2$ lahus jäeti sel juhul puhverdamata või lisati sellele vähem fosfaate. Teisalt ilmnes, et sedavõrd kõrge kontsentratsiooniga $MgCl_2$ lahusel on ka endal märgatav puhverdusvõime.

Puhverdamine osutus efektiivseks meetmeks pH muutuste vastu. Puhverdatud ja puhverdamata lahustega saadud logD väärtuste võrdlus näitab, et enamiku neutraalsete ainete korral ei ole nende erinevused eksperimendiveast suuremad. See tulemus tõestas ühtlasi, et fosfaatpuhvri mõju jaotustasakaaludele võimalike spetsiifiliste interaktsioonide kaudu ei oma statistilist tähtsust. Seega on enamik neutraalsetele veefaasidele vastavaid logD väärtusi saadud kasutades puhverdatud lahuseid.

4.2.3 Temperatuurikõikumised laboriruumis ja automaatsisestusseadmes

Vedelik-vedelik jaotustasakaalud on keerulises sõltuvuses temperatuurist. Seetõttu jälgiti pidevalt temperatuuri laboriruumis ja vedelikkromatograafi automaatsisestusseadme sees ning tehti katseid temperatuurimuutuste mõju uurimiseks. Temperatuur laboris eksperimentide teostamise perioodil jäi 20.5°C ja 25.4°C piiridesse. Temperatuur kromatograafi automaatsisestusseadmes, kus viaalid asusid kuni 18 tundi enne analüüsi, oli toatemperatuurist keskmiselt 1.5°C kõrgem (22.7-27°C). Seega võib väita, et ekstraheerimise ajal oli temperatuur vahemikus 24 ± 3°C.

Metoodika temperatuuritundlikkuse uurimiseks korrati sama katsete seeriat erinevatel päevadel termostateeritud automaatsisestusseadmes kahel temperatuuril. Määrati 9 erineva hüdrofoobsusega aine jaotuskoefitsiente 12 solvendipaaris, mis koosnesid kahest orgaanilisest

lahustist (2-metüültetrahüdrofuraan, isopropüülatsetaat) ja kuuest erinevaid sooli ja happeid sisaldavast veefaasist. Leiti, et 20 ja 25°C juures saadud logD väärtused ei erine enamikul juhtudel rohkem kui kahel erineval päeval samal temperatuuril, 20°C juures, saadud tulemused. Seega, temperatuuri kõikumise efekt jaotustasakaalule ei ületa muudest faktoritest tingitud päevadevahelist korratavust.

Seetõttu teostati kõik katsed peale ülalkirjeldatud seeria termostateerimata automaatsisestusseadmes.

4.2.4 Ebasoovitavad vastasmõjud segude komponentide vahel

Saavutamaks töö kõrgemat efektiivsust, viidi suurem osa ekstraktsioone läbi ainete rühmadega (enamasti 6-9 ainet). Selle lähenemise peamiseks probleemiks on võimalikud vastasmõjud ainete vahel. Töö käigus täheldati kahte vastasmõjude tüüpi:

1. Reaktsioonid ainete (ka solvendilisandite) vahel uu(t)e ühendi(te) moodustumisega. Reaktsioonide toimumise indikaatoriteks on uute tundmatute piikide ilmumine kromatogrammidele ja/või lahuste värvi muutused.
2. Vastasmõjud, mis ei kutsu esile keemilisi muundumisi, kuid mõjutavad jaotustasakaalu. Töö käigus täheldati, et erinevate ainete kombinatsioonide kasutamisel saadud logD väärtused võivad üksteisest märgatavalt erineda. Võimalusel püüti tuvastada teineteist mõjutavate ainete paare.

Eeltoodust on selge, et üheainsa mõõtmise tulemusena saadud logD väärtus üldjuhul ei ole usaldusväärne, eriti kui uuritakse mitut ainet samas lahuses. Seega on enamik ennustava mudeli loomiseks kasutatud logD väärtusi arvutatud vähemalt 2 korduskatse tulemusest, mis on teostatud erinevatel päevadel erinevate ainete kombinatsioonidega. Üksikmõõtmiste väärtuseid kasutati juhtudel, kus ainega eelnevalt teistes lahustipaarides tehtud korduskatsete kokkulangevused olid väga head ja üksikmõõtmise tulemuse piisav täpsus oli seeläbi tõendatud.

Mitmes katses saadud logD väärtuste maksimaalse vastuvõetava erinevuse leidmiseks kasutati järgmist empiirilist kriteeriumit:

$$|\Delta\text{logD}| \leq 0.05 \cdot (|\text{logD}| + 1) \quad (10)$$

Võrratuse 10 järgi sõltub lubatud lahknevus logD absoluutväärtusest. Võrratuse kuju lähtub praktilistest kaalutlustest: mida suurem on |logD| väärtus, seda väiksem on analüüdi sisaldus ühes faasis ja seda keerulisem on logD täpne määramine ning seda kõrgem on reeglina saadud logD määramatus.

4.3 Lõplik andmemaatriks

Eksperimentaalse töö käigus ilmnenud probleemide tõttu (nt sagedane reaktsioonide esinemine, madal neelduvus, ebapiisav lahustuvus enamikus solventides) jäeti mõned esialgsesse treenimisandmestikku valitud ühendid välja või leiti nendele asendused. Asenduseks valiti mõni probleemse analüüdiga võimalikult sarnane ühend (kas COSMO-RS modelleerimise tulemuste või Joonis 1 - laadse graafiku alusel). Lõplik eksperimentaalsete andmete kogum on toodud Lisas 4.

Osa logD väärtusi ei õnnestunud eksperimentaalselt määrata erinevatel põhjustel:

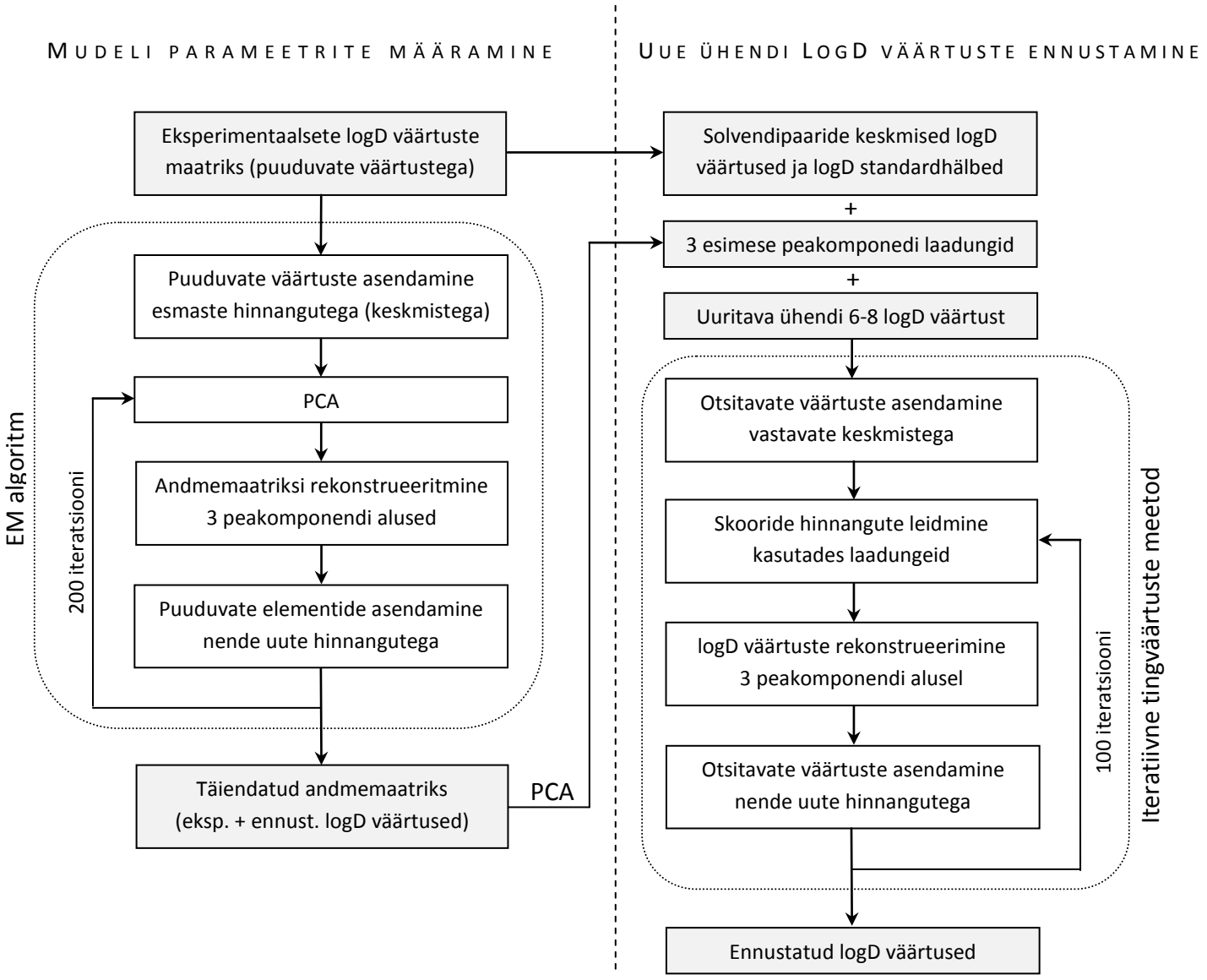
- reaktsioon ühendi ja mõne solvendilisandi vahel;
- liiga suur logD absoluutväärtus (>3.5);
- kordusmõõtmiste suur lahknevus, mille põhjus polnud selge.

Mõned logD väärtused (mitme korduskatse keskmised) kaasati andmestikku vaatamata võrrandiga 10 määratud erinevuse ületamisele. Seda tehti, kui lubatud logD lahknevus oli ületatud vähesel määral või logD keskmine oli arvutatud 4 või rohkem korduskatse tulemusest.

Lõplik andmemaatriks sisaldab 2614 logD väärtust (arvutatud 5304 üksikmõõtmise väärtusest), mis kirjeldavad 44 ühendi jaotust 72 solvendipaaris. Paralleelkatsete tulemuste kogutud standardhälve üle kogu andmestiku on 0.08 log ühikut. Lõplik andmemaatriks sisaldab 82.5% teoreetilisest andmepunktide arvust.

5. Ennustava mudeli loomine

Eksperimendiandmete alusel ennustamist võimaldav algoritm on algselt loodud Dr. Koji Muteki poolt ning realiseeritud MATLAB [38] keskkonnas. Käesoleva töö autor realiseeris selle algoritmi keskkonnas R [39] (programmi tekst keskkonnas R on toodud Lisas 5). Arvutusliku algoritmi üldkuju esitab Joonis 3. Algoritmi kõikide etappide detailsemad kirjeldused on toodud allpool.



Joonis 3. Arvutuslik algoritm mudeli parameetrite määramiseks ja logD väärtuste ennustamiseks.

5.1 Puuduvate väärtuste arvutamine EM algoritmi abil

Puuduvad väärtused eksperimentaalsete andmete maatriksis, mis takistasid PCA rakendamist, arvutati ooteväärtuste tõenäosuse maksimeerimise (EM) algoritmi abil. Algoritm rakendati järgnevalt:

1. Puuduvad väärtused eksperimentaalsete andmete maatriksis asendati vastavas veerus olemasolevate (ehk samale solvendipaarile vastavate) logD väärtuste aritmeetilise keskmisega.
2. Eelnevalt skaleeritud ja tsentreeritud andmetele rakendati peakomponentide analüüsi.
3. Andmemaatriks rekonstrueeriti esimese kolme peakomponendi alusel.
4. Elemendid, mille väärtused puudusid esialgsest maatriksist, asendati sammus 3 arvutatud väärtustega.
5. Arvutati esimese kolme peakomponendi poolt kirjeldatav andmestiku varieeruvuse osa ja korrati tsüklit alates sammust 2.

Leiti, et mudeli koondumiseks (maksimaalse saavutatava kirjeldatud varieeruvuseni jõudmiseks) piisab 200 iteratsioonist.

Saadud PCA mudeli esimest 3 peakomponenti kirjeldavad 93.5% andmete varieeruvusest. See tähendab, et erinevatele solvendipaaridele vastavad logD väärtused on omavahel tugevalt korreleerunud ning PCA on sobiv meetod ennustava algoritmi loomiseks käesoleva andmestiku põhjal.

5.2 Uute ühendite logD väärtuste ennustamine

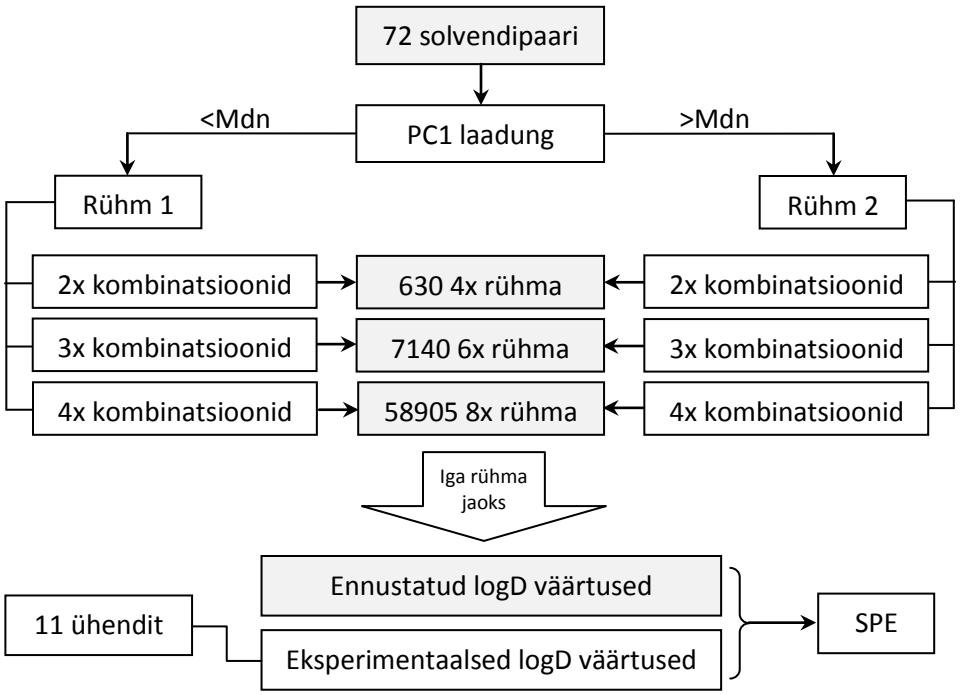
Uute ühendite logD väärtuste ennustamiseks võeti kasutusele nn iteratiivne tingväärtuste meetod. Algoritmi parameetrid optimeeriti mitme juhuslikult valitud andmepunkti abil. Katsetati andmete rekonstrueerimist 3 ja 4 peakomponendi alusel ning andmete skaleerimist fikseeritud ja fikseerimata väärtustega. Leiti, et parimaid tulemusi annab andmete rekonstrueerimine kolme peakomponendi alusel kombinatsioonis skaleerimisega fikseeritud parameetritega (esialgsest eksperimendiandmete maatriksist leitud veergude keskmiste ja standardhälvetega). Samuti tehti kindlaks, et ennustatud logD väärtuste koondumiseks piisab 100 iteratsioonist.

5.3 Ennustamiseks sobivate solvendipaaride valimine

Esialgselt oli plaanis kasutada 6-10 eksperimentaalselt määratud logD väärtust 62-66 logD väärtuse ennustamiseks. Selleks kõige paremini sobivaid solvendipaaride kombinatsioone otsiti Monte Carlo meetodil põhineva algoritmi abil (Joonis 4).

Valiti 11 ühendit, mille jaoks olid eksperimentaalselt määratud kõik (72) logD väärtust. Skooride graafiku abil tehti kindlaks, et need ühendid ei ole omaduste poolest liiga sarnased (ehk vastavad punktid PC1 vs PC2 graafikul ei ole üksteisele lähedal).

Arvutusmahu vähendamiseks jagati solvendipaarid kaheks rühmaks, mediaanist madalama ja mediaanist kõrgema esimese peakomponendi laadungi väärtusega solvendipaarideks. Mõlema rühma solvendipaaridest moodustati kõik võimalikud 2, 3 ja 4-liikmelised alamrühmad, mis ühendati omavahel, saades vastavalt 4, 6 ja 8-liikmelised solvendipaaride rühmad.



Joonis 4. Solvendipaaride kombinatsioonide koostamine ja nende headuse esmane hindamine.

Kõiki moodustatud solvendipaaride kombinatsioone kasutati ülalkirjeldatud 11 ühendi logD väärtuste ennustamiseks ning arvutati ennustusvigade ruutude summa (SPE, võrrand 11) iga solvendipaaride kombinatsiooni jaoks.

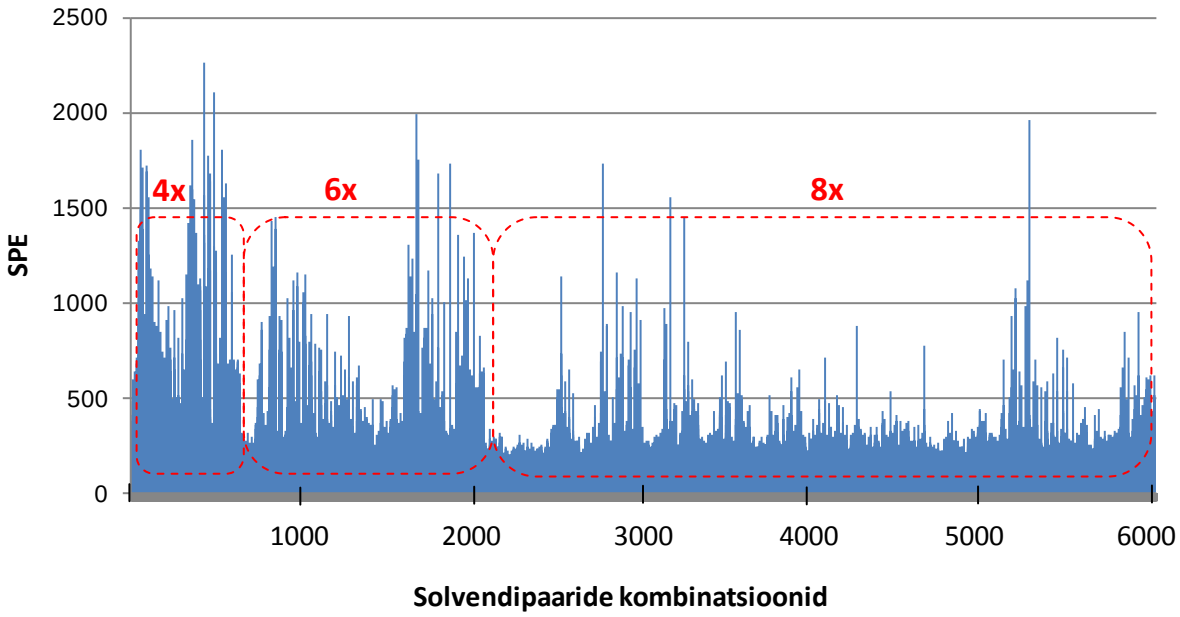
$$SPE = \sum_{n=1}^{11} \sum_{s=1}^{72} \left(\log D_{eksp}^{n,s} - \log D_{ennust}^{n,s} \right)^2 \quad (11)$$

Kus

SPE – ennustusvigade ruutude summa,

$\log D_{eksp}^{n,s}$, $\log D_{ennust}^{n,s}$ – ühendi n jaotuskoefitsiendi logaritmid solvendipaaris s, vastavalt eksperimentaalne ja ennustatud.

Leiti, et kuigi SPE suurus sõltub ennustamiseks kasutatavate solvendipaaride arvust, tähtsaim tegur on siiski solvendipaaride kombinatsiooni sobivus. Sama solvendipaaride arvu juures võivad eri kombinatsioonide SPE väärtused nii erineda ligi 10 korda kui ka olla väga lähedased (Joonis 5).



Joonis 5. Erinevatele solvendipaaride kombinatsioonidele vastavad SPE väärtused (suure andmemahu tõttu on 6x ja 8x kombinatsioonid kajastatud osaliselt).

Kaaludes eksperimentaalse töö mahu ja ennustamise täpsust, otsustati kasutada ennustamiseks 8 solvendipaari. Eelkirjeldatud meetodil tuvastatud 3000 parima 8 solvendipaari kombinatsiooniga teostati kaheetapiline lisauuring.

Esimese etapina arvutati iga solvendipaaride kombinatsiooni abil välja kõikide treenimisandmestiku ühendite jaotuskoefitsiendid ülejäänud solvendipaarides. Arvutusmahu vähendamise eesmärgil kasutati kõikide ennustuste tegemiseks samu parameetreid (keskmised, standardhälbed, laadungid), mis olid arvutatud täielikust treenimisandmestikust. Solvendipaaride kombinatsiooni headust hinnati ennustatud ja eksperimentaalsete andmete vaheliste kogutud standardhälvete abil.

Eelneva testi käigus leitud 150 parima solvendipaaride kombinatsioonidega viidi läbi uuringu teine etapp. 12 ühenditest, mille jaoks olid eksperimentaalselt määratud kõik logD väärtused (toodud Tabelis 6), moodustati kõikvõimalikud paarid. Iga ainete paari jaoks leiti eksperimentaalsetest andmetest 10 solvendipaari, mis sobiksid kõige paremini nende ainete eraldamiseks üksteisest (ehk selliseid, milles ainete logD väärtuste vahe on suurim). Iga solvendipaaride kombinatsiooni korral arvutati, mitu sobivamat solvendipaari eelmainitud kümnest suudab mudel ennustada iga ainepaari jaoks. Suurima õigesti ennustatud „heade“ ekstraktsioonitingimuste arvuga solvendipaaride kombinatsioon on toodud Tabelis 3. Seda kasutati järgnevaks mudeli valideerimiseks.

Tabel 3. Ennustamiseks sobivaimad solvendipaarid.

	Orgaaniline solvent	**Veefaas**
1	2-metüültetrahüdrofuraan	3.5% HCl, 13% NaCl lahus
2	isopropüülatsetaat	6% etaanhappe, 19.7% $MgCl_2$ lahus
3	isopropüülatsetaat	vesi
4	tolueen	18% sidrunhappe lahus
5	tolueen-metanool (9:1)	vesi
6	tolueen-metanool (9:1)	3.5% HCl, 13% NaCl lahus
7	diklorometaan	6% etaanhappe lahus
8	diklorometaan	6% etaanhappe, 19.7% $MgCl_2$ lahus

6. Mudeli valideerimine

Mudeli valideerimiseks kasutati ristkontrolli ja seejärel valideerimist sõltumatu andmestiku abil.

6.1 „Jäta-üks-välja“ ristkontroll

Loodud arvutusliku mudeli headuse hindamiseks kasutati „jäta-üks-välja“ (*Leave-One-Out*, edaspidi LOO) ristkontrolli. Aineid jäeti ühekaupa treenimisandmestikust välja ja rakendati ülejäänud andmetele käesolevas töös loodud andmetöötlusalgoritm (kaasa arvatud puuduvate väärtuste lähendamine EM-PCA meetodiga). Välja jäetud ühendi jaoks ennustati jaotuskoefitsiendid kõikides solvendipaarides ja hinnati ennustuse üldist headust eksperimentaalsete ja arvutatud andmete vahelise ruutkeskmise hälve kaudu.

Ristkontrolli tulemused on toodud Lisas 6A. Kogutud standardhälbeks üle kogu treenimisandmestiku on 0.75 log ühikut. Üksikute ainete ruutkeskmised hälbed varieerusid vahemikus 0.31-1.58 log ühikut.

Mudeli tundlikkuse uurimiseks treenimisandmestiku muutmise suhtes jäeti andmestikust välja 6 ühendit, mille ruutkeskmised hälbed esimeses LOO testis olid suuremad kui 1 log ühik. LOO test korrati ülejäänud 38 ühendiga ja saadi kogutud standardhälbe väärtuseks üle kogu andmetabeli 0.66 ühikut. Väärtused varieerusid vahemikus 0.32-0.98 log ühikut. Seejärel lisati treenimisandmestikule 5 valideerimisühendit (Tabel 4) ja korrati LOO test 49 ühendiga. Kogutud standardhälve väärtuseks saadi 0.79 log ühikut (üksikutel ainetel 0.28-1.90).

Nende modifikatsioonide korral leitud kogutud standardhälbed üle kogu andmestiku varieeruvad vahemikus 0.66 kuni 0.79 log ühikut. See varieerumine on tagasihoidlik, mis võimaldab lugeda mudelit stabiilseks („robustseks“) ja mõnede üksikute ainete poolt vähe mõjutatavaks.

6.2 Valideerimine sõltumatu andmestiku abil

Loodud mudeli valideerimiseks valiti 2 ainete rühma – keskmise ja kõrge aluselisusega – mis ei olnud mudeli loomisel kasutuses (Tabel 4). Kummaski rühmas on ühendite aluselisus rühmasiseselt sarnane, kuid hüdrofoobsused on märgatavalt erinevad.

Tabel 4. Mudeli valideerimiseks kasutatud ühendid (andmed: ACDLabs).

Analüüt	M ($g \cdot mol^{-1}$)[a]	V ($cm3 \cdot mol^{-1}$)[b]	pKa [c]	logP [d]
N-metüülaniliin	107.15	108.8	4.70	1.71
N-tsükloheksüülaniliin	175.27	171.3	5.46	3.68
N-bensüülmetüülamiin	121.18	131.0	9.75	1.43
N-bensüülisopropüülamiin	149.23	164.4	9.77	2.30
Dibensüülamiin	197.28	191.7	8.76	3.03

[a] Molaarmass, g/mol
[b] Molekulruumala, cm^3/mol
[c] Protoneeritud vormi pKa vesikeskkonnas
[d] Oktanool-vesi jaotuskoefitsiendi logaritm

Valideerimisandmestiku jaotuskoefitsientide eksperimentaalseks määramiseks kasutati sama metoodikat kui treenimisandmestiku korral. Et hinnata ennustava mudeli efektiivsust reaalsete segude korral, teostati mõõtmised 5 aine seguga, v.a juhtudel, kui praktiliste raskuste (jaotuskoefitsientide suur erinevus, reaktsioonid) tõttu oli üheaegset määramist keeruline teostada. Seega on osa logD väärtusi saadud kasutades ühte analüüti sisaldavaid lahuseid. Iga solvendipaari korral teostati vähemalt 2 korduskatset ja nende tulemused keskmistati.

Nagu ka treenimisandmestiku korral, osa logD väärtusi ei õnnestunud eksperimentaalselt määrata. Valideerimisandmestiku eksperimentaalsed ja osa ennustatud logD väärtustest on toodud Lisas 6B.

Valideerimisühendite jaotuskoefitsiente ennustati kasutades:

A. täielikku treenimisandmestikku;
B. treenimisandmestikku, kust on välja jäetud 6 ühendit, millele vastavad ruutkeskmised hälbed LOO testi tulemusena olid suuremad kui 1 log ühik;
C. treenimisandmestikku sellele lisatud valideerimisühendite eksperimentaalandmetega;
D. punktis B kirjeldatud andmestikku, täiendatud valideerimisühendite eksperimentaalandmetega;
E. punktis B kirjeldatud andmestikku, täiendatud vaatlusaluse analüüdiga.

Saadud ennustuste ja eksperimentaalandmete vahelised ruutkeskmised hälbed on toodud Tabelis 5. Andmete detailsem analüüs on toodud allpool (jaotused 7.1 ja 7.2).

Tabel 5. Mudeli valideerimise tulemused: kogutud standardhälbed ja R^2 (sulgudes).*

Analüüt	A {44 üh.}	B {38 üh.}	C {44+5 üh.}	D {38+5 üh.}	E {38+1 üh.}
N-metüülaniliin	0.48 (0.94)	0.37	0.37	0.26 (0.98)	0.30
N-tsükloheksüülaniliin	0.58 (0.88)	0.50	0.44	0.32 (0.96)	0.40
N-bensüülmetüülamiin	0.85 (0.31)	0.87	0.65	0.60 (0.57)	0.65
N-bensüülisopropüülamiin	0.92 (0.22)	0.91	0.83	0.81 (0.31)	0.85
Dibensüülamiin	1.25 (0.50)	1.31	1.18	1.20 (0.54)	1.19

*Tähistused A-E on seletatud ülalpool. Looksulgudes on näidatud ühendite arv treenimisandmestikus.

6.3 Jääkliikmete analüüs

LOO testi tulemusena saadud ennustatud logD väärtused võrreldi vastavate eksperimentaalsete väärtustega. Jääkliikmete maatriksi uurimine näitas, et ennustusvigade väärtuste jaotus ei ole täiesti juhuslik. Erinevatele ühenditele vastavate jääkliikmete vahel võib esineda märkimisväärne korrelatsioon, nii positiivne kui negatiivne. Korrelatsioonid on tugevaimad nende solvendipaaride korral, kus orgaaniliseks faasiks on tolueen, ja lähedaste omadustega ühendite korral. Näiteks võib tuua korrelatsiooni katehhooli ja resortsinooli (R^2 = 0.81) ning lidokaiini ja difenüülguanidiini (R^2 = 0.52) jääkliikmete vahel.

7. Tulemused ja arutelu

Töö tulemusena valminud ennustusmudel koosneb andmestikust (Lisa 4) ja selle põhjal töötavast algoritmist, mis on esitatud Joonisel 3 ja Lisas 5.

7.1 Ülevaade valideerimise tulemustest

Ristkontrolli tulemused näitavad, et käesoleva töö tulemusena loodud arvutuslik mudel on rahuldavalt robustne nii ühendite treenimisandmestikust väljajätmise kui ka treenimisandmestiku täiendamise suhtes (Lisa 6A). Kuna puudub selgelt väljendunud sõltuvus LOO testi tulemuste (ruutkeskmiste hälvete suuruse) ja ühendi omaduste vahel, võib mudelit lugeda küllaltki universaalselt rakendatavaks erinevatele neutraalsetele ja aluselistele ainetele.

Treenimisandmestiku kohandamine on kasulik, kui selle tulemusena kasvab uuritava ühendiga omaduste poolest sarnaste ühendite osakaal. Seda efekti illustreerib Tabel 5. Viie valideerimisühendi lisamine treenimisandmestikku parandab märgatavalt ennustuse kvaliteeti. Ainult uuritava analüüdi lisamine on vähem efektiivne. Halvasti kirjeldatavate ühendite (kõrge ruutkeskmise hälvega LOO testis) väljajätmine andmestikust parandab tulemuste kvaliteeti aniliinide seeria korral, kuid praktiliselt ei muuda olukorda amiinide seeria korral.

Valideerimisühendite uurimise tulemused (Tabel 5) näitavad, et erinevalt aniliinidest on amiinide korral jaotuskoefitsientide ennustamise usaldusväärsus küllaltki madal sõltumatult sellest, mis treenimisandmestikku ennustamiseks kasutati. Selle täpset põhjust on keeruline määrata, kuna vastavast andmestikust puudub palju eksperimentaalseid logD väärtusi. Juhul, kui puudub ennustamiseks kasutatavale solvendipaarile vastav väärtus, arvutati jaotuskoefitsiendid tegelikkuses mitte 8, vaid väiksema arvu solvendipaaride alusel. Samas raskendab paljude eksperimentaalsete väärtuste puudumine ennustuse kvaliteedi täpsemat hindamist.

7.2 Võimalike veaallikate analüüs

Ennustatud ja eksperimentaalsete väärtuste vahelised erinevused (väljendatult standardhälbena) on enamiku ainete korral küllalt suured. Järgnevalt on vaadeldud selle nähtuse võimalikud põhjused ja arutletud nende tõenäosuse üle.

Analüütide vahelised vastasmõjud segudes. Vaatleme olukorda, kus ühend A on võimeline moodustada püsiva kompleksi ühendiga B (võrrand 12):

$$A + B \rightleftharpoons A\cdots B \qquad (12)$$

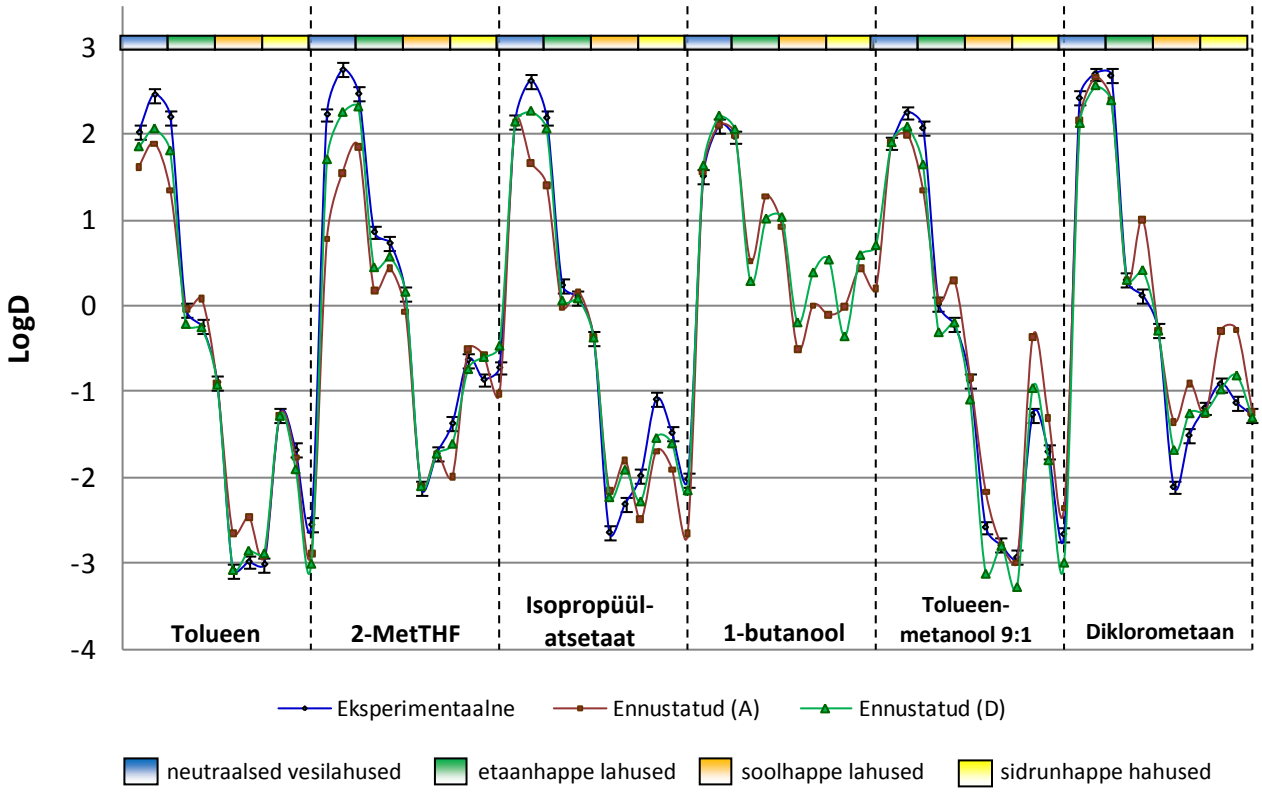
Ühendeid võivad kompleksis koos hoida elektrostaatilised vastasmõjud, vesiniksidemed, hüdrofoobsed jõud jms. Kõige tõenäolisem käesolevas olukorras on vesiniksidemega komplekside ja/või ioonpaaride teke orgaanilises faasis. Vesiniksidemete doonorite/aktseptorite ja laetud osakeste arvukuse tõttu veefaasis võivad seal püsida vaid erakordselt tugevad analüütide kompleksid. Kui reaktsiooni 12 tasakaalukonstandid oleksid sarnased kõikide orgaaniliste faaside korral, on võimalik, et arvutuslik mudel suudaks neid kirjeldada ja kajastada komplekseerumise efekti ennustatud väärtustes. Praktikas on see aga vähetõenäoline, kuna kasutatud orgaaniliste lahustite omadused, mis võivad mõjutada analüütide vahelist vastasmõju (vesiniksideme doonorite/aktseptorite olemasolu, dielektriline läbitavus) on küllalt erinevad. Seega võib tugevate analüütidevaheliste komplekside teke osades lahustites alandada ennustuste kvaliteeti.

LOO testi korral on ülalkirjeldatud interaktsioonide mõju tulemustele praktiliselt välistatud, kuna treenimisühendite eksperimentaalsed logD väärtused arvutati mitmes erinevas ainete segus saadud tulemustest. Valideerimissegu korral aga ei saa välistada ainete omavaheliste vastasmõjude esinemist. See võib olla üheks ennustamiste madala kvaliteedi põhjuseks.

Liiga vähe ühendeid treenimisandmestikus. Selle hüpoteesi lükkavad ümber valideerimise tulemused. Osa ühendite eemaldamisel treenimisandmestikust ennustamiste kvaliteet pigem kasvab kui kahaneb. Ennustamise täpsust aitaks tõsta uuritava ühendiga sarnaste molekulide lisamine treenimisandmestikku või nende osakaalu tõstmine vaatlusalusest analüüdist erinevate molekulide eemaldamise teel. Seda meedet ei ole võimalik rakendada täiesti tundmatu struktuuriga segu komponentide korral. Samas, kui uuritavate ainete struktuurid või omadused on vähemalt osaliselt teada, siis võib olemasolevat infot rakendada mudeli kohandamiseks konkreetse olukorra tarbeks.

Ebapiisav mudeli alusandmete täpsus. Eksperimentaalsete logD väärtuste määramise korratavus väljendatuna korduskatsete tulemuste standardhälvena on kordades madalam kui

vastavad ennustusvead (Lisa 6A). Seega, ei ole ootuspärane, et eksperimendiandmete täpsus limiteeriks ennustustulemuste täpsust. Joonisel 6 on toodud ennustamiste tulemuste võrdlus eksperimentaalsete andmetega N-metüülaniliini näitel.



Joonis 6. N-metüülaniliini eksperimentaalsed ja ennustatud logD väärtused. A – ennustus muutmata treenimisandmestiku alusel, D – ennustus modifitseeritud treenimisandmestiku alusel (eemaldatud halvasti kirjeldatavad ühendid, lisatud valideerimisühendid). Detailid alajaotuses 6.2.

Tähtsamaid solvendipaaride erinevusi suudab arvutuslik mudel hästi esile tuua, kuid nii Jooniselt 6 kui ka valideerimise tulemuste tabelist Lisas 6B võib näha ennustatud ja eksperimentaalsete väärtuste märkimisväärseid lahknevusi ekstreemumkohtades. Selle efekti põhjuseks võib olla treenimisandmestikust eksperimentaalsete väärtuste puudumise tõenäosuse seos puuduvate väärtuste suurusega. Juhul, kui logD absoluutväärtus on üle 3-3.5 ühikut, on selle eksperimentaalne määramine tihti problemaatiline ja vähetäpne. Seega on kõrgete logD absoluutväärtuste korral suurem tõenäosus, et vastav andmepunkt puudub andmemaatriksist, kui nullilähedaste logD väärtuste korral.

EM algoritm, mille abil arvutati puuduvatele andmepunktidele hinnanguid, ei pruugi selles olukorras usaldusväärseid tulemusi anda [29]. Juhul, kui puuduv logD väärtus on väga kõrge või väga madal, võib EM algoritm anda ebapiisava täpsusega või süstemaatilise veaga hinnanguid, mis võib omakorda põhjustada kõrgete |logD| väärtuste suuruse alahindamist ennustava mudeli poolt.

7.3 Mudeli praktilise rakendatavuse hindamine

Loodud ennustava mudeli praktilist rakendatavust ei ole võimalik ainult standardhälvete kaudu hinnata. Mudeli loomise eesmärgiks oli võimaldada leida kõige sobivamad ekstraktsioonitingimused kahe või enama aine eraldamiseks üksteisest. Eraldamiseks sobivaim solvendipaar on see, milles ainete jaotuskoefitsientide erinevused on suurimad.

Arvutusliku mudeli võimet seda eesmärki täita võib hinnata jaotuses 5.3 kirjeldatud protseduuri abil. Treenimisandmestikust valiti ühendeid, mille jaoks on määratud kõik eksperimentaalsed logD väärtused, moodustati nendest kõikvõimalikud paarid ja leiti nii eksperimentaalsetest kui LOO testi käigus ennustatud andmetest 10 solvendipaari, kus kahe aine logD väärtuste vahed on suurimad. Iga ainepaari korral leiti, mitu solvendipaari 10 parima hulgast suutis mudel ennustada. Tulemused on toodud Tabelis 6.

Nagu Tabelist 6 on näha, varieerub mudeli poolt tuvastatud „heade“ solvendipaaride arv nullist üheksani, nende keskmine arv on 5. Samas võib märgata, et aluselisuse ja hüdrofoobsuse poolest küllalt sarnaste ühendite (nt, trifenüülfosfiinoksiid ja dimetüülftalaat) jaoks on mudel tuvastanud enamasti vaid mõned sobivad ekstraktsioonitingimused, kuid omaduste poolest piisavalt erinevate ainete jaoks on paljudel juhtudel leitud rohkem sobivaid solvendipaare. Kuna sarnaste omadustega (ja sellest tulenevalt sarnaste jaotuskoefitsientide väärtustega) ainete üksteisest eraldamine ekstraktsiooni abil ei ole efektiivne ja selle rakendamine tööstusprotsessis on vähetõenäoline, võib mudeli toimetulekut püstitatud ülesandega lugeda rahuldavaks.

Tabel 6. Parimate ekstraktsioonitingimuste leidmine ennustava mudeli abil.

	Resortsinool	Katehhool	Bensamiid	Tümiin	5-nitro-bensimidasool	1-naftool	Difenüülamiin	Dimetüül-ftalaat	Kofeiin	Sulfametoksa-sool	Trifenüül-fosfiinoksiid
Katehhool	4										
Bensamiid	3	4									
Tümiin	3	4	4								
5-nitrobensimidasool	7	7	5	7							
1-naftool	6	6	4	4	7						
Difenüülamiin	8	9	6	5	7	7					
Dimetüülftalaat	5	4	5	5	6	1	4				
Kofeiin	2	6	6	6	6	6	3	5			
Sulfametoksasool	2	4	6	3	5	5	4	3	5		
Trifenüülfosfiinoksiid	4	3	4	5	5	1	2	0	4	3	
Diantipüriilfenüülmetaan	9	8	8	8	3	7	7	8	5	8	9

LogD väärtuste ennustamise edukus sõltub nii uuritava ühendi sarnasusest treenimisandmestiku molekulidega kui ka otsitava jaotuskoefitsiendi väärtusest (Joonis 6): ennustusvead on suurimad ekstreemumkohtades. Kuigi erinevused eksperimentaalsete ja ennustatud logD väärtuste vahel on sel juhul küllalt suured, mudeli praktilise rakendatavuse seisukohalt ei pruugi need põhjustada raskusi. Kui ühendi kontsentratsioonide suhe kahes faasis on üle 100 ($|logD|>2$), ei oma praktilise töö seisukohalt tähtsust, milline see suhe täpselt on. Seega, kui eksperimentaalne logD absoluutväärtus ületab 2 log ühikut, ennustust $|logD|>2$ (kui ennustatud väärtus on sama märgiga kui eksperimentaalne) võib lugeda rahuldavaks sõltumata ennustuse veast.

7.4 Ennustava mudeli edasiarendamise võimalused

Loodud ennustava mudeli parandamiseks ja selle kasutusvaldkonna laiendamiseks on plaanis:

- Uurida lähemalt ennustusvigade seost ennustatud logD väärtustega ning tugevate sõltuvuste avastamise korral tuua mudelisse võimalused vastavate vigade korrigeerimiseks.
- Üritada leida sõltuvust analüüdi struktuuri/omaduste ja vastava ennustuse täpsuse vahel.
- Lisada treenimisandmestikku happelisi analüüte ja aluselisi veefaase kirjeldav andmemassiiv.
- Uurida lähemalt eksperimentaalselt määratud ja COSMO-RS meetodiga ennustatud logD väärtuste sõltuvused ja määrata erinevate solvendipaaride ja/või ainerühmade jaoks parandustegureid. Edu korral võimaldab see täiendada treenimisandmestikku uute andmepunktidega, asendades (vähemalt osaliselt) eksperimente arvutustega.

8. Kokkuvõte

Käesoleva töö käigus on loodud arvutuslik metodoloogia, mis võimaldab ennustada teadmata struktuuriga analüütide jaotuskoefitsiente mitmekümnes solvendipaaris selles töös uuritute hulgast, kasutades selleks 6-8 eksperimentaalselt määratud jaotuskoefitsienti. Arvutusliku mudeli peaeesmärgiks on parimate lahustipaaride ennustamine kahe või enama teadmata struktuuriga ühendi eraldamiseks üksteisest vedelik-vedelik ekstraktsiooni meetodil.

Arvutusliku mudeli rakendusala hõlmab hetkel aluselisi ja neutraalseid analüüte ja 72 solvendipaari, mis koosnevad neutraalsetest ja happelistest vesilahustest ja 6 orgaanilisest lahustist.

Ennustuste kvaliteet varieerub sõltuvalt uuritavast analüüdist. Ennustused on täpsemad, kui mudeli parameetrite määramiseks kasutatud treenimisandmestikus on üks või mitu uuritavale analüüdile omaduste poolest sarnast ainet. Jaotuskoefitsientide (eriti kõrgete ja madalate) täpne ennustamine alati ei õnnestu, kuid mudeli rakendatavust praktikas ei pruugi see segada.

On leitud, et käesoleva metodoloogia võimaldab korrektselt ennustada keskmiselt viit kümnest kahe aine eraldamiseks sobivaimast solvendipaarist. Õigesti ennustatud heade ekstraktsioonitingimuste arv on suurem omadustelt erinevate ainete korral.

Metodoloogia põhiprobleemideks on lahused, kus esinevad tugevad analüütidevahelised vastasmõjud, mis võivad mõjutada tasakaale ekstraktsioonil ja selle tulemusena ka ennustuste kvaliteeti, ning vähesed võimalused ennustatud väärtuste täpsuse hindamiseks mudeli rakendamisel praktikas.

Töös pakutakse välja võimalused loodud mudeli efektiivsuse parandamiseks ja selle kasutusala laiendamiseks.

9. Summary

Extraction prediction model

Sofja Tšepelevitš

A methodology was developed for prediction of extraction behavior (distribution ratios) of compounds with unknown molecular structures in a range of solvent pairs. The algorithm takes as input the experimentally determined distribution ratios in 6-8 solvent pairs, and predicts the distribution ratios in the rest of the solvent pairs studied in this work. The main application of this methodology is to predict the most suitable extraction conditions for selective separation of one or more unidentified compounds from a liquid mixture.

The applicability domain of the predictive model covers neutral and basic analytes and 72 solvent pairs (combinations of 6 organic solvents and 12 neutral and acidic aqueous solutions).

The quality of the results (predicted distribution ratios) varies depending on the analyte. The prediction tends to be more accurate if there are compound(s) among those used to determine the model parameters that have properties similar to the properties of analyte. The prediction of exact logD values (especially in the case of high absolute values) using the model in its present state proved rather unreliable. Nevertheless, the model is usable for solving practical problems. It enables to identify at average 5 of 10 most suitable extraction conditions for separation of the pair of compounds. The number of correctly predicted suitable conditions is higher in case of compounds that have sufficiently different properties.

The main problem of the present methodology are solutions where analytes interact with each other (if the interactions are strong enough to shift the distribution equilibria, the predictions are likely to be biased), and insufficient means to assess the quality of the prediction in case of analytes with totally unknown structures.

The ways were proposed to refine the predictive methodology and expand its applicability.

Kasutatud kirjandus

1. *The art of drug synthesis*, Johnson, D. S., Li, J. J., Eds.; John Wiley & Sons, Inc. Hoboken, New Jersey, 2007.

2. *Solvent Extraction Principles and Practice*; 2nd ed.; Rydberg, J., Cox, M., Musicas, C., Choppin, G. R., Eds.; CRC Press, 2004.

3. Mazzola, P. G.; Lopes, A. M.; Hasmann, F. A.; Jozala, A. F.; Penna, T. C. V.; Magalhaes, P. O.; Rangel-Yagui, C. O.; Pessoa, A., Jr. Liquid–liquid extraction of biomolecules: an overview and update of the main techniques. *J. Chem. Technol. Biotechnol.* **2008**, *83*, 143-157.

4. Gil-Chávez, G. J.; Villa, J. A.; Ayala-Zavala, J. F.; Heredia, J. B.; Sepulveda, D.; Yahia, E. M.; González-Aguilar, G.A. Technologies for Extraction and Production of Bioactive Compounds to be Used as Nutraceuticals and Food Ingredients: An Overview. *Compr. Rev. Food Sci. Food Safety* **2013**, *12*, 1, 5-23.

5. Lv, X.; Hao, Y.; Jia, Q. Preconcentration Procedures for Phthalate Esters Combined with Chromatographic Analysis. *J. Chromatogr. Sci.* **2013**, *51*, 7, 632-644.

6. Kocúrová, L.; Balogh, I. S.; Andruch, V. Solvent microextraction: A review of recent efforts at automation. *Microchem. J.* **2013**, *110*, 599-607.

7. Saraji, M.; Boroujeni, M. K. Recent developments in dispersive liquid–liquid microextraction. *Anal. Bioanal. Chem.* **2014**, *406*, 8, 2027-2066.

8. Majors, R. E. Miniaturized Approaches to Conventional Liquid-Liquid Extraction. *LC GC Eur*. **2006**, *19*, 5, 284-293.

9. Martins, J. G.; Ch ávez, A. A.; Waliszewski, S. M.; Cruz, A. C.; Fabila, M. M. G. Extraction and clean-up methods for organochlorine pesticides determination in milk. *Chemosphere* **2013**, *92*, 233-246.

10. Berendsen, B. J. A.; Stolker L. (A.) A. M.; Nielen, M. W. F. Selectivity in the sample preparation for the analysis of drug residues in products of animal origin using LC-MS. *Trends Anal. Chem*. **2013**, *43*, 229-239.

11. Abraham, M. H.; Acree, W. E., Jr. The transfer of neutral molecules, ions and ionic species from water to benzonitrile; comparison with nitrobenzene. *Thermochim. Acta* **2011**, 526, 22-28.

12. Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*, 3rd ed.; VCH: Weinheim, Germany, 2003.

13. Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev*. **1993**, *22*, 73-83.

14. Abraham, M. H.; Ibrahim, A.; Zissimos, A. M. Determination of sets of solute descriptors from chromatographic measurements. *J. Chromatogr. A* **2004**, *1037*, 29-47.

15. Abraham, M. H.; Acree, W. E., Jr. Equations for the Transfer of Neutral Molecules and Ionic Species from Water to Organic phases. *J. Org. Chem.* **2010**, *75*, 1006-1015.

16. Sprunger L. M.; Gibbs, J.; Acree, W. E., Jr.; Abraham, M. H. Correlation of Human and Animal Air-to-Blood Partition Coefficients With a Single Linear Free Energy Relationship Model. *QSAR Comb. Sci.* **2008**, *27*, 9, 1130-1139.

17. Zhao, Y. H.; Abraham, M. H.; Hersey, A.; Luscombe, C. N. Quantitative relationship between rat intestinal absorption and Abraham descriptors. *Eur. J. Med. Chem.* **2003**, *38*, 939-947.

18. Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier Science Ltd: Amsterdam, 2005.

19. Klamt, A; Eckert, F; Arlt, W. COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annu. Rev. Chem. Biomol. Eng*. **2010**, *1*, 101-122.

20. Tshepelevitsh, S.; Oss, M; Pung, A.; Leito, I. Evaluating the COSMO-RS Method for Modeling Hydrogen Bonding in Solution. *ChemPhysChem* **2013**, *14*, 1909-1919.

21. Eckert, F.; Klamt, A. Validation of the COSMO-RS Method: Six Binary Systems. *Ind. Eng. Chem. Res.* **2001,** *40,* 2371-2378.

22. Oleszek-Kudlak, S.; Grabda, M.; Shibata, E.; Eckert, F.; Nakamura, T. Application of the Conductor-like Screening Model for Real Solvents for prediction the aqueous solubility of chlorobenzenes depending on temperature and salinity. *Environ. Toxicol. Chem.* **2005**, *24*, 6, 1368-1375.

23. Klamt, A. Prediction of the mutual solubilities of hydrocarbons and water with COSMO-RS. *Fluid Phase Equilib.* **2003**, *206*, 223-235.

24. Reithinger, M.; Arlt, W. Prediction of the Partitioning Coefficient in Liquid-Solid Chromatography using COSMO-RS. *Chem. Ing. Tech.* **2011**, *83*, 83–89.

25. Mokrushina, L.; Buggert, M.; Smirnova, I.; Arlt, W.; Schomäcker, R. COSMO-RS and UNIFAC in Prediction of Micelle/Water Partition Coefficients. *Ind. Eng. Chem. Res.* **2007**, *46*, 6501-6509.

26. Ikeda, H.; Chiba, K.; Kanou, A.; Hirayama, N. Prediction of Solubility of Drugs by Conductor-Like Screening Model for Real Solvents. *Chem. Pharm. Bull.* **2005**, *53*, 2, 253-255.

27. Daszykowski, M; Kaczmarek, K; Vander Heyden, Y.; Walczak, B. Robust statistics in data analysis – A review. Basic concepts. *Chemom. Intell. Lab. Syst.* **2007,** *85,* 203-219.

28. Brereton, R. G. *Chemometrics for Pattern Recognition.* John Wiley & Sons Ltd, Chichester, 2009.

29. Muteki, K.; MacGregor, J. F.; Ueda, T. Estimation of missing data using latent variable methods with auxiliary information. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 41-50.

30. Do, C. B.; Batzoglou, S. What is the expectation maximization algorithm? *Nat. Biotechnol.* **2008**, *26*, 8, 897-899.

31. Walczak, B.; Massart, D.L. Dealing with missing data. Part I. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 15-27.

32. Levine, R. A.; Casella, G. Implementations of the Monte Carlo EM Algorithm. *J. Comput. Graph. Stat.*, **2001**, *10*, 3, 422-439.

33. Hong, H.; Schonfeld, D. Maximum-Entropy Expectation-Maximization Algorithm for Image Reconstruction and Sensor Field Estimation. *IEEE Trans. Image Processing* **2008**, *17*, 6, 897-907.

34. Arteaga, F.; Ferrer, A. Dealing with missing data in MSPC: several methods, different interpretations, some examples. *J. Chemometrics* **2002**, *16*, 408-418.

35. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3-26.

36. Huber, U. Determination of octanol-water partition coefficients using the Agilent 220 micro plate sampler. Publication Number 5980-0493E; https://www.chem.agilent.com/Library/applications/59800493.pdf (viimati alla laetud 25.05.2014).

37. Berthod, A.; Carda-Broch, S. Determination of liquid-liquid partition coefficients by separation methods. *J. Chromatogr. A* **2004**, *1037*, 3-14.

38. Programmeerimiskeskkonna MATLAB tutvustus arendaja The MathWorks Inc. ametlikul veebilehel; http://www.mathworks.se/products/matlab/ (viimati alla laetud 25.05.2014)

39. Programmeerimiskeskkonna R tutvustus tarkvara ametlikul veebilehel; http://www.r-project.org/ (viimati alla laetud 25.05.2014)

40. R. Ahlrichs, M. Bär, H.-P. Baron, R. Bauernschmitt, S. Böcker, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, F. Haase, M. Häser, H. Horn, C. Hattig, C. Huber, U. Huniar, M. Kattannek, M. Köhn, C. Kölmel, M. Kollwitz, K. May, C. Ochsenfeld, H. Öhm, A. Schäfer, U. Schneider, O. Treutler, M. von Arnim, F. Weigend, P. Weis, H. Weiss, TURBOMOLE V6.2, 2010.

41. Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187-2193.

42. Schäfer, A.; Huber, C.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys*. **1994**, *100*, 5829-5835.

43. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **1988**, *38*, 3098-3100.

44. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B*, **1986**, *33*, 8822-8824.

45. COSMOtherm Version C3.0 Release 12.01 Release notes. COSMO*logic* GmbH & Co. KG, Leverkusen, Germany, 2011.

46. Buggert, M.; Cadena, C.; Mokrushina, L.; Smirnova, I.; Maginn, E.J.; Arlt, W. COSMO-RS Calculations of Partition Coefficients: Different Tools for Conformational Search. *Chem. Eng. Technol.* **2009**, *32*, 977–986.

Lisad

Lisa 1. Jaotuskoefitsientide modelleerimine COSMO-RS meetodiga.

COSMO-RS arvutus koosneb kahest etapist [18,19]:

1) molekuli oleku (koguenergia, geomeetria ja laengujaotus pinnal) leidmine ideaalses juhis;

2) molekulide omavaheliste vastasmõjude arvutamine statistilise termodünaamika meetoditega kasutades esimeses etapis saadud andmeid laengujaotuste kohta.

1. Kvantkeemilised arvutused

Molekuli geomeetria optimeerimine ja laengujaotuse arvutus pinnal viidi läbi TURBOMOLE V 6.2 tarkvarapaketi abil [40,41]. Valitud meetod on DFT (*Density Functional Theory*), BP funktsionaali (*Becke's hybrid exchange + Perdew-Wang correlation*) [43,44] ja TZVP (*Triple Zeta Valence plus Polarisation*) baasiga [42]. Enamikul juhtudel on kasutatud tarkvara vaikimisi parameetrid (energia kovergeerumisparameeter 10^{-6} Hartree, gradient |dE/dxyz| 10^{-3} Hartree/Bohr, *gridsize* m3), mõnedel juhtudel rangemad parameetrid (energia kovergeerumisparameeter 10^{-7} Hartree, gradient |dE/dxyz| 10^{-4} Hartree/Bohr).

Ühenditele, mis esinevad mitmes stabiilses konformatsioonis, arvutati mitme konformeeri geomeetriaid (2-16). Konformeere otsiti lähtudes molekulide ruumilise struktuuri üldistest põhimõtetest. Rõhuva enamiku molekulide jaoks tehti mitu algset geomeetriat, et kogu konformatsiooniline ruum võimalikult põhjalikult katta. Optimeerimise tulemusena koondusid need geomeetriad väiksema konformeeride arvuni.

2. Statistiline termodünaamika

Jaotustasakaalude arvutamiseks statistilise termodünaamika meetoditega kasutati COSMO*therm* tarkvarapaketti (versioon C30_1201 [45]; parametrisatsioon BP_TZVP_C30_1201).

Molekuli reaalne olek lahuses kujutab endast sageli konformeeride segu. COSMO*therm* võimaldab iga molekuli jaoks arvesse võtta mitu konformeeri, omistades nendele osakaale vastavalt Boltzmanni jaotusele. Erinevate konformeeride stabiilsus ja see, milline neist vastab

globaalsele energiamiinimumile, sõltub keskkonnast, milles molekul asub. Senini tehtud uuringud on näidanud, et ainult madalaima energiaga konformeeri kasutamine ei pruugi parimaid tulemusi anda ning eksperimentaalsete andmete reprodutseerimine on edukaim konformeeride komplekti kasutamisel [24,25,46]. Käesolevas töös kasutati kõigi molekulide korral arvutustes kõiki leitud konformeere.

COSMO-RS arvutustes kasutati temperatuuri 25°C.

Jaotustasakaale arvutati kolme etapina:

1. Kahe faasi koostise arvutamine peale vastastikku küllastumist (kasutades COSMO*therm* moodulit *Liquid Extraction*).
2. Analüütide neutraalsete vormide jaotuse (logP) arvutamine kahe „eelvalmistatud“ faasi vahel (*LLE/VLE* mooduli abil). Kasutati lõpmatut lahjendust (analüütide kontsentratsioonid võeti nulliks).
3. Ionisatsiooni arvessevõtmine (logD arvutus). Molekuli kõikide vormide jaotused arvutati logP ja ionisatsioonikonstandi (pKa) väärtustest. Viimase hindamiseks kasutati eksperimendiandmeid (kui see võimalik oli) või andmebaasist SciFinder saadud ACDLabs tarkvara abil arvutatud väärtusi.

Lisa 2. Kasutatud kemikaalid.

CAS number	Nimetus	Molekulvalem	Molaarmass (g/mol)	Tootja *	Põhiaine sisaldus/ puhtus *
A N A L Ü Ü D I D					
110-86-1	Püridiin	C_5H_5N	79.10	Sigma-Aldrich	99.8%
62-53-3	Aniliin	C_6H_7N	93.13	Sigma-Aldrich	≥99.5%
100-46-9	Bensüülamiin	C_7H_9N	107.15	Aldrich	99%
100-61-8	N-metüülaniliin	C_7H_9N	107.15	Fluka	≥99.5%
591-27-5	3-aminofenool	C_6H_7NO	109.13	NA	NA
120-80-9	Katehhool	$C_6H_6O_2$	110.11	NA	ч.
108-46-3	Resortsinool	$C_6H_6O_2$	110.11	Schering-Kahlbaum AG	p. A.
51-17-2	Bensimidasool	$C_7H_6N_2$	118.14	Aldrich	98%
55-21-0	Bensamiid	C_7H_7NO	121.14	Aldrich	99%
108-75-8	2,4,6-trimetüülpüridiin	$C_8H_{11}N$	121.18	NA	NA
103-67-3	N-bensüülmetüülamiin	$C_8H_{11}N$	121.18	Aldrich	97%
65-71-4	Tümiin	$C_5H_6N_2O_2$	126.11	Sigma	≥99%
934-32-7	2-aminobensimidasool	$C_7H_7N_3$	133.15	Fluka	99.9%
90-15-3	1-naftool	$C_{10}H_8O$	144.17	Merck	p. A.
102-97-6	N-bensüülisopropüülamiin	$C_{10}H_{15}N$	149.23	Aldrich	97%
455-14-1	4-(trifluorometüül)aniliin	$C_7H_6F_3N$	161.12	Aldrich	99%
94-52-0	5-nitrobensimidasool	$C_7H_5N_3O_2$	163.13	Aldrich	98%
122-39-4	Difenüülamiin	$C_{12}H_{11}N$	169.23	NA	ч.д.а.
63-74-1	Sulfanüülamiid	$C_6H_8N_2O_2S$	172.20	NA	NA
1821-36-9	N-tsükloheksüülaniliin	$C_{12}H_{17}N$	175.28	Aldrich	NA
59-26-7	N,N-dietüülnikotiinamiid	$C_{10}H_{14}N_2O$	178.23	Aldrich	99%
131-11-3	Dimetüülftalaat	$C_{10}H_{10}O_4$	194.18	Merck	≥99%
58-08-2	Kofeiin	$C_8H_{10}N_4O_2$	194.19	NA	NA
103-49-1	Dibensüülamiin	$C_{14}H_{15}N$	197.28	Aldrich	97%
102-06-7	1,3-Difenüülguanidiin	$C_{13}H_{13}N_3H$	211.26	NA	NA
137-58-6	Lidokaiin	$C_{14}H_{22}N_2O$	234.34	Sigma	NA
57-41-0	Fenütoiin	$C_{15}H_{12}N_2O_2$	252.268	Sigma	≥99%
723-46-6	Sulfametoksasool	$C_{10}H_{11}N_3O_3S$	253.28	Sigma	NA
29122-68-7	Atenolool	$C_{14}H_{22}N_2O_3$	266.34	Sigma	≥98%

CAS number	Nimetus	Molekulvalem	Molaarmass (g/mol)	Tootja *	Põhiaine sisaldus/ puhtus *
497-76-7	Arbutiin	$C_{12}H_{16}O_7$	272.25	Sigma	≥98%
50-28-2	β-Estradiool	$C_{18}H_{24}O_2$	272.38	Sigma	≥98%
791-28-6	Trifenüülfosfiinoksiid	$C_{18}H_{15}PO$	278.28	Aldrich	98%
138-52-3	D-(–)-salitsiin	$C_{13}H_{18}O_7$	286.28	Sigma	≥99%
97-77-8	Disulfiraam	$C_{10}H_{20}N_2S_4$	296.54	Aldrich	≥97%
70458-96-7	Norfloksatsiin	$C_{16}H_{18}FN_3O_3$	319.3	Fluka	≥98%
1160-28-7	8,8’-dikinolüüldisulfiid	$C_{18}H_{12}N_2S_2$	320.43	Завод "Реагент"	ч.
115-86-6	Trifenüülfosfaat	$C_{18}H_{15}PO_4$	326.28	Aldrich	>99%
85-86-9	Sudaan III	$C_{22}H_{16}N_4O$	352.40	Fluka	NA
50370-12-2	Tsefadroksiil	$C_{16}H_{17}N_3O_5S$	363.39	Sigma	NA
1251-85-0	Diantipüriilmetaan	$C_{23}H_{24}O_2N_4$	388.46	Реахим	ч.д.а.
548-62-9	Kristallviolett	$C_{25}H_{30}N_3Cl$	407.98	Реахим	ч.д.а.
69257-04-1	(8S, 9R)-(–)-N-bensüül-tsinkonidiiniumkloriid	$C_{26}H_{29}ClN_2O$	420.97	Aldrich	98%
125-20-2	Tümoolftaleiin	$C_{28}H_{30}O_4$	430.54	Реахим	ч.д.а.
95255-44-0	Diantipüriilfenüülmetaan	$C_{29}H_{28}O_2N_4$	464.56	Реахим	ч.д.а.
22832-87-7	Mikonasool, nitraat	$C_{18}H_{14}Cl_4N_2O{\cdot}HNO_3$	479.14	Fluka	NA
60-54-8	Tetratsükliin, hüdrokloriid	$C_{22}H_{24}N_2O_8{\cdot}HCl$	480.90	Sigma-Aldrich	97.3%
1476-53-5	Novobiotsiin, naatriumsool	$C_{31}H_{35}N_2NaO_{11}$	634.61	Sigma	≥93%
13292-46-1	Rifampitsiin	$C_{43}H_{58}N_4O_{12}$	822.94	Sigma	≥97%
11121-48-5	Bengali roosa	$C_{20}H_2Cl_4I_4Na_2O_5$	973.67	Chemapol	NA
E K S T R A K T S I O O N I L A H U S T I T E K O M P O N E N D I D					
67-56-1	Metanool	CH_4O	32.04	Sigma-Aldrich	≥99.9%
71-36-3	1-butanool	$C_4H_{10}O$	74.12	Реахим	99.5%
75-09-2	Diklorometaan	CH_2Cl_2	84.93	J.T.Baker	≥99.8%
				EM Science	GR
96-47-9	2-metüültetrahüdrofuraan	$C_5H_{10}O$	86.13	Sigma-Aldrich	≥99%
108-88-3	Tolueen	C_7H_8	92.14	J.T.Baker	≥99.7%
				Sigma-Aldrich	≥99.9%
108-21-4	Isopropüülatsetaat	$C_5H_{10}O_2$	102.13	Sigma-Aldrich	98%
7647-01-0	Soolhape	HCl	36.46	Fluka	NA
64-19-7	Etaanhape	$C_2H_4O_2$	60.05	Fisher Scientific	USP/FCC

CAS number	Nimetus	Molekulvalem	Molaarmass (g/mol)	Tootja *	Põhiaine sisaldus/ puhtus *
77-92-9	Sidrunhape, monohüdraat	$C_6H_8O_7 \cdot 2H_2O$	210.14	Fisher Scientific	99.8% **
7647-14-5	Naatriumkloriid	NaCl	58.44	Реахим	99.8%
				Реахим	≥99.9%
7786-30-3	Magneesiumkloriid, veevaba	$MgCl_2$	95.21	Sigma	≥98%
7791-18-6	Magneesiumkloriid, heksahüdraat	$MgCl_2 \cdot 6H_2O$	203.3	KeboLab AB	puriss
7558-80-7	Naatriumdivesinikfosfaat, monohüdraat	$NaH_2PO_4 \cdot H_2O$	137.99	Merck	>99%
7558-79-4	Naatriumvesinikfosfaat, dodekahüdraat	$Na_2HPO_4 \cdot 12H_2O$	358.14	Реахим	ч.д.а.

*NA – info puudub; kemikaali puhtus hinnati vedelikkromatograafilise analüüsi kaudu.

**Sidrunhapet puhastati täiendavalt vedelik-vedelik ekstraktsioonil tolueeniga.

Lisa 3. Kasutatud aparatuur, töövahendid ja meetodid.

Vedelikkromatograafilise analüüsi aparatuur:

- Vedelikkromatograaf Agilent 1200, mis koosnes automaatsisestussüsteemist, kvaternaarsest pumbast ja 5-kanalilisest UV-Vis detektorist.
- Automaatsisestussüsteemi termostaat Agilent 1290 (kasutati eksperimendimetoodika tundlikkuse määramiseks temperatuuri suhtes).

Vedelikkromatograafilise analüüsi protseduur:

- Elueerimiseks kasutati üle saja erineva meetodi. Nende hulgas oli nii gradient- kui ka isokraatilisi meetodeid.
- Detekteerimiseks kasutati erinevaid lainepikkusi vahemikus 220-580 nm.
- Kromatograafiline kolonn: Agilent Eclipse XDB-C18, 4.6 x 250 mm, täidiseosakese suurusega 5μM; eelkolonn 4.6 x 5 mm samast materjalist.
- Eluentide komponendid:
    - Orgaanilised lahustid – metanool ja atsetonitriil.
    - Ammooniumatsetaatpuhver (pH=4.8) kontsentratsioonivahemikus 5-20 mM.
    - 1 mM ammooniumatsetaat + 0.1% sipelghape puhver (pH=2.8)

Vahendid ekstraktsiooni läbiviimiseks:

- Standardsed 2 ml valtskorkidega viaalid
- Orbitaalloksuti Elpan, tüüp 358S

Lisa 4. Eksperimentide tulemused (mitme korduskatse keskmised logD väärtused).

Veefaaside komponentide tähistused:

w – vesi; **s** – NaCl; **x** – $MgCl_2$; **a** – etaanhape; **h** – soolhape; **c** – sidrunhape.

Orgaaniline solvent →	**Tolueen**											
Analüüt↓ Veefaas→	**w**	**s**	**x**	**a**	**as**	**ax**	**h**	**hs**	**hx**	**c**	**cs**	**cx**
β-estradiool	2.02	2.64	2.64	1.69	2.33	2.56	1.96	2.42	2.56	1.44	1.90	1.94
1-naftool	1.76	2.50	2.63	1.60	2.02	2.27	1.82	2.26	2.47	1.50	1.91	2.06
2,4,6-trimetüülpüridiin	0.97	0.94	1.69			-3.24						
2-aminobensimidasool	-3.04	-2.91	-2.34	-2.95	-2.84	-3.06			-2.41	-3.03	-2.49	-2.66
3-aminofenool	-1.31	-1.12	-1.14	-2.55	-3.22	-3.62	-3.55	-3.01	-2.26	-1.84	-1.93	
4-(trifluorometüül)aniliin	2.28	2.72	2.79	1.62	1.66	1.24	-0.71	-0.85	-1.42	0.72	0.43	-0.51
5-nitrobensimidasool	-0.60	-0.39	-0.46	-1.17	-1.23	-1.70	-2.26	-2.12	-3.28	-2.86	-3.40	-3.84
8,8-dikinolüüldisulfiid	3.93	4.43	3.49	3.47	2.71	0.63				1.78	0.13	-2.32
Aniliin	0.93	1.22	1.18	-0.93	-1.02	-1.60	-3.88	-4.10	-4.23		-2.67	-3.69
Arbutiin												
Atenolool				-1.40	-1.14	-0.73	-1.14	-1.25	-1.22	-1.83	-1.89	-2.11
Bengali roosa				3.28	3.23	3.30	3.19	3.23	3.03	3.43	3.71	3.52
Bensamiid	-0.83	-0.56	-0.59	-0.52	-0.07	0.12	-0.94	-0.67	-0.80	-1.09	-0.87	-1.02
Bensimidasool	-0.85	-0.48	-0.70	-3.64	-3.72		-2.73		-2.45	-2.78	-2.37	-2.49
Bensüülamiin	-1.77	-1.81	-2.26									-2.92
D-(–)-salitsiin						-3.65				-3.81	-3.78	-3.61
Diantipüriilfenüülmetaan	2.74	3.37	3.29	2.05	2.49	1.81	-0.40	-0.69	-0.86	0.59	0.40	-1.28
Diantipüriilmetaan	1.11	1.94	1.57	0.87	1.43	0.77	-1.14	-1.34	-1.99	-0.02	-0.16	-1.60
Difenüülamiin	3.97	4.17	3.74	3.72	3.69	3.65	3.00	2.73	2.18	3.43	3.55	2.81
Difenüülguanidiin	-1.78	-1.27	-1.83	-2.98	-2.73	-1.92	-4.03	-3.44	-2.57	-3.51	-3.16	-3.03
Dimetüülftalaat	2.35	2.81	2.86	2.00	2.45	2.56	2.19	2.60	2.82	1.92	2.32	2.32
Disulfiram	4.07			4.16	3.94			3.74	4.12	3.62	3.83	
Fenütoiin	0.70	1.32	1.12	0.81	1.38	1.91	0.67	1.08	1.25	0.29	0.61	0.75
Katehhool	-1.19	-0.83	-0.40	-1.28	-0.98	-0.56	-1.15	-0.91	-0.56	-1.27	-1.02	-0.69
Kofeiin	-0.27	0.02	-0.31	-0.37	-0.06	-0.24	-0.80	-1.19	-2.26	-0.66	-0.54	-0.68
Kristallviolett				-2.02	-0.80	-1.08				-3.81	-3.96	-3.96
Lidokaiin	1.34	1.31	-0.45	-2.87	-2.58	-2.32	-3.46	-3.38	-3.14	-2.05	-3.52	-3.10
Mikonasool	3.80	3.56	2.60	1.90	2.14	1.88	-0.52	0.81	1.16	0.05	0.13	
N,N-dietüülnikotiinamiid	-0.06	0.35	0.11	-0.73	-0.86	-1.69	-3.72	-3.99	-4.31	-2.19	-2.64	-3.94
N-bensüül-tsinkonidiiniumkloriid	-2.88	-1.50	-2.21	-2.99	-3.03	-2.38	-2.85	-2.99		-2.82	-2.87	-2.81
Norfloksatsiin	-3.10	-3.28		-2.77								
Novobiotsiin	-1.55	-0.52			2.20	-0.16	2.23			1.65	2.46	
Püridiin	0.39	0.69	0.50	-1.28	-1.48	-2.45	-3.74		-1.59	-2.16	-2.10	-1.78
Resortsinool	-1.94	-1.63	-1.35	-2.13	-1.81	-1.41	-1.98	-1.57	-1.36	-2.16	-1.84	-1.56
Rifampitsiin	1.30	3.48		1.00	1.82	2.35		0.46	1.59			-0.31
Sudaan III	4.05	4.33	4.48	3.97	4.08	4.21	4.25	3.90	3.74	3.87	4.03	3.96
Sulfametoksasool	-1.58	-0.97	-0.07	-0.48	-0.17	0.00	-2.17	-2.42	-2.57	-1.08	-1.32	-2.20
Sulfanüülamiid	-2.49	-2.32	-2.26	-2.64	-2.67	-2.93	-3.99	-4.60				
Tetratsükliin												
Trifenüülfosfaat	4.28			4.37	4.18			4.22		3.93		
Trifenüülfosfiinoksiid	2.26	3.24	2.93	2.31	2.92	3.06	2.18	2.77	2.65	1.80	2.27	2.05
Tsefadroksiil												
Tümiin	-3.38	-3.32	-3.30	-2.58	-2.25	-1.88	-3.58	-3.20	-3.66	-3.53	-3.57	-3.78
Tümoolftaleiin	3.85	3.72		2.58	3.87		1.95	3.59		1.34	3.45	

Orgaaniline solvent →	2-metüültetrahüdrofuraan											
Analüüt↓ Veefaas→	**w**	**s**	**x**	**a**	**as**	**ax**	**h**	**hs**	**hx**	**c**	**cs**	**cx**
β-estradiool	3.19	3.18	3.44	2.96	3.24	3.19	3.28	3.48	3.28	2.58	3.41	3.35
1-naftool	2.94	3.09	3.28	2.80	3.05	3.23	3.06	2.40	3.18	1.81	3.02	3.15
2,4,6-trimetüülpüridiin	1.06	0.94	1.08	-2.24	-2.23	-2.65	-3.56	-3.15	-3.50	-2.35	-2.14	-1.67
2-aminobensimidasool	0.06	-0.02	-0.18	-1.80	-0.52	-0.53	-1.49	-0.96	-0.75	-0.87	-0.26	0.43
3-aminofenool	0.97	1.36	1.49	0.34	-0.44	-0.89	-1.73	-1.40	-0.94	-0.39	-0.59	
4-(trifluorometüül)aniliin	3.18	3.31	3.39	2.57	3.08	2.96	0.63	0.63	0.47	1.81	1.91	1.61
5-nitrobensimidasool	1.59	2.01	1.75	1.40	1.60	0.88	-1.39	-1.59	-1.43	0.49	-0.09	-0.08
8,8-dikinolüüldisulfiid	3.34	3.44	3.64	2.93	2.75	2.36				2.18	0.90	-1.34
Aniliin	1.51	1.92	1.97	0.31	0.24	-0.28	-1.61	-1.78	-1.53			
Arbutiin	-0.89	-0.72	-0.99	-0.68	-0.55	-0.63	-0.88	-0.84	-0.80	-0.30	-0.31	-0.24
Atenolool				0.23	0.40	0.58	-0.33	-0.33	-0.34	-0.27	-0.18	0.27
Bengali roosa				3.28	3.32	3.36	3.06	3.09	3.18	3.00	3.02	3.12
Bensamiid	0.61	0.98	0.86	0.69	1.19	1.13	0.43	0.79	0.54	0.57	1.08	1.26
Bensimidasool	0.83	1.28	0.98	-1.27	-1.09	-1.50	-2.49	-2.19	-1.96	-1.41	-1.16	-0.58
Bensüülamiin	-1.14			-0.98	-0.84	-0.84	-1.28	-1.14	-1.19	-1.01		-0.33
D-(–)-salitsiin	-0.97	-0.86	-1.60	-0.77	-0.68	-0.91	-1.19	-1.07	-1.28	-0.60	-0.62	-0.67
Diantipüriilfenüülmetaan	2.25	3.17	2.99	1.59	2.60	1.92	-0.95	-0.04	-0.39	0.46	1.26	2.26
Diantipüriilmetaan												
Difenüülamiin	3.21	3.35	3.17	3.07	3.28	3.24	3.22	3.12	3.28	2.83	3.11	3.22
Difenüülguanidiin	-0.00			-1.60	0.20	0.44	-1.13	-0.84	-0.17	-0.71	0.33	1.10
Dimetüülftalaat	1.89	2.58	2.77	1.67	2.53	2.76	1.72	2.44	2.23	1.34	2.30	2.61
Disulfiram	3.14	3.16	3.23	3.08	3.11	3.17	3.18			2.61		
Fenütoiin	2.82	3.23	3.24	2.39	3.08	3.43	2.65	3.09	2.74	1.89	2.96	2.98
Katehhool	1.79	2.15	2.52	1.57	2.07	2.75	1.70	2.02	2.91	1.37	2.09	2.65
Kofeiin	-0.32	0.03	-0.33	-0.22	0.14	-0.23	-0.65	-1.04	-2.01	-0.18	0.00	-0.35
Kristallviolett	-1.14	0.78			1.51	0.61				-0.25	-0.20	-1.74
Lidokaiin	0.32			-2.24	-0.68	-0.77				-1.21	-0.45	0.07
Mikonasool	3.15	3.26			2.60	2.66	0.77	1.99	1.82	0.80	2.49	
N,N-dietüülnikotiinamiid	-0.04	0.57	0.18	-0.23	-0.19	-1.07				-0.99	-1.58	-1.81
N-bensüül-tsinkonidiiniumkloriid	-2.54		-0.26	-1.98	-0.97	-1.51				-1.94	-1.91	-1.32
Norfloksatsiin	-1.40	-1.31		-2.67	-1.56	-2.49	-2.59	-2.01	-2.80	-1.34	-1.04	-0.81
Novobiotsiin	3.32		3.67	3.72	3.75	3.61		3.78	3.95		3.77	3.74
Püridiin	0.40	0.79	0.55	-0.61	-1.29				-1.58	-1.65	-2.13	-2.20
Resortsinool	1.59	2.21	2.40	1.51	2.15	2.39	1.66	2.34	2.77	1.30	2.12	2.64
Rifampitsiin	0.40	3.59		0.95	3.17		0.72	3.18		0.48		
Sudaan III	3.79		3.79	3.72	3.91	3.67	3.80	3.91	3.71	3.69	3.61	3.76
Sulfametoksasool	1.18	1.88	1.48	1.88	2.68	2.75	0.76	0.67	0.16	1.56	2.19	1.87
Sulfanüülamiid	0.61	0.80	0.82	0.55	0.62	0.35	-1.30	-1.74	-1.61	0.15	-0.14	-0.85
Tetratsükliin	-0.44	-0.43	-1.97	-0.90	-0.51	-1.14		-0.51	0.38			
Trifenüülfosfaat	3.41	3.27	3.60	2.90	2.95	3.15	2.98	3.40	3.27	2.80	3.18	3.42
Trifenüülfosfiinoksiid	1.75	2.79	2.89	1.72	2.88	2.99	1.38	2.28	1.83	1.30	2.58	3.00
Tsefadroksiil				-0.71	-0.58	-0.65	-0.96	-0.82	-0.67	-0.53	-0.53	-0.46
Tümiin	-0.44	-0.44	-0.51	-0.33	-0.12	-0.27	-0.49	-0.31	-0.51	-0.26	-0.05	-0.14
Tümoolftaleiin	3.31	3.38	3.36	3.11	3.14	3.20	3.12	3.17	3.28	2.98	3.24	3.20

Orgaaniline solvent →	Isopropüülatsetaat											
Analüüt↓ Veefaas→	**w**	**s**	**x**	**a**	**as**	**ax**	**h**	**hs**	**hx**	**c**	**cs**	**cx**
β-estradiool	3.16	3.48	3.21	3.01	3.16	3.39	3.01	3.27	3.29	2.73	3.03	3.00
1-naftool	3.10	3.28	3.19	2.42	3.22	3.12	2.88	2.98	2.90	2.59	2.82	2.99
2,4,6-trimetüülpüridiin	0.84	0.79	0.93	-2.71	-2.59	-2.95				-2.33	-2.54	
2-aminobensimidasool	-0.70	-0.71	-0.28	-2.49	-1.62	-1.30	-2.38	-2.03	-1.94	-2.98	-2.51	-1.99
3-aminofenool	0.52	0.80	0.84			-1.65	-2.95		-2.53		-1.24	-2.11
4-(trifluorometüül)aniliin	3.08	3.27	3.27	2.84	2.98	2.72	0.36	0.15	-0.32	1.78	1.44	0.59
5-nitrobensimidasool	1.04	1.32	1.26	0.48	0.23	-0.37	-2.08	-2.40	-2.40	-1.03	-1.68	-2.67
8,8-dikinolüüldisulfiid	2.07	4.00	3.12	3.21	2.81	0.82	-1.77			1.46	-0.05	-2.85
Aniliin	1.42	1.74	1.72	-0.21	-0.29	-0.80		-2.42	-2.06	-1.72	-2.09	-2.33
Arbutiin	-2.47	-2.23	-2.09	-2.23	-2.06	-1.98	-2.28	-2.18	-2.18	-2.45	-2.30	-2.32
Atenolool				-0.26	-0.09	0.14	-0.49	-0.49	-0.31	-1.51	-1.54	-1.16
Bengali roosa		2.47		2.86	3.08	3.50	3.25	2.92	3.08	3.40	3.16	3.44
Bensamiid	0.40	0.67	0.62	0.53	0.92	0.97	0.41	0.60	0.47	0.17	0.40	0.36
Bensimidasool	0.52	0.86	1.13	-1.77	-1.86	-1.99			-2.90	-3.44	-3.28	-2.88
Bensüülamiin												
D-(–)-salitsiin	-2.36	-2.13	-2.27	-2.15	-1.85	-1.82	-2.28	-2.21	-2.21	-2.33	-2.31	-2.52
Diantipüriilfenüülmetaan	2.06	2.85	2.71	1.85	2.54	1.99	-0.65	-0.68	-0.40	0.02	-0.23	-1.00
Diantipüriilmetaan												
Difenüülamiin	3.18	3.26	3.28	3.27	3.18	3.38	2.55	2.96	2.24	3.16	3.12	2.83
Difenüülguanidiin	-0.59	-0.35	-0.53	-2.01	-0.47	0.05	-1.54	-0.75	-0.49	-2.96	-1.24	-1.10
Dimetüülftalaat	2.19	2.72	2.74	1.99	2.56	2.73	2.01	2.49	2.42	1.85	2.26	2.32
Disulfiram	3.83	3.39		3.48	3.34			3.43		2.94	3.15	
Fenütoiin	2.54	3.01	2.92	2.39	2.94	3.19	2.39	2.71		2.08	2.35	2.54
Katehhool	1.12	1.40	1.68	1.06	1.34	1.70	1.20	1.42	1.70	1.04	1.28	1.56
Kofeiin	-0.30	0.00	-0.32	-0.31	0.02	-0.28	-0.77	-1.24	-2.22	-0.66	-0.51	-1.31
Kristallviolett	-0.88	1.61	1.26			-0.01				-3.33	-3.56	
Lidokaiin	1.24	1.01	-0.46	-2.69	-1.42	-1.11	-2.11	-1.64	-1.35	-3.09	-2.24	-1.82
Mikonasool	3.60	3.35	2.66	1.77	2.44	2.44	0.87	1.68	1.78	0.02	1.04	1.41
N,N-dietüülnikotiinamiid	-0.02	0.55	0.13	-0.51	-0.53	-1.40	-3.28	-3.62	-3.93	-1.94	-2.41	-3.53
N-bensüül-tsinkonidiiniumkloriid	-3.20	-0.69		-3.31	-1.78	-2.00			-3.97	-3.11		
Norfloksatsiin	-2.23	-2.44		-3.10	-2.47	-2.67				-2.69	-3.41	-3.65
Novobiotsiin	1.73		3.39	3.78	3.86	3.76	3.69	3.55	3.61	3.68	3.74	3.80
Püridiin	0.45	0.76	0.55	-0.85	-1.27	-2.59						
Resortsinool	0.98	1.34	1.64	0.91	1.27	1.54	1.05	1.39	1.62	0.89	1.17	1.37
Rifampitsiin		2.44	3.16	0.48	2.35	3.32	1.06	2.51	2.04	-0.85	1.09	
Sudaan III	3.98	3.93	3.75	3.33	3.95	3.92	3.68	3.71	3.75	3.66	3.83	3.99
Sulfametoksasool	0.58	1.18	0.87	1.66	1.98	2.14	0.18	-0.08	-0.67	1.19	0.97	0.09
Sulfanüülamiid	0.09	0.19	0.28	0.03	-0.11	-0.36	-1.97	-2.33	-2.31	-0.56	-1.00	-1.87
Tetratsükliin	-1.00	-1.05		-2.01	-1.62	-2.74	-2.40			-3.22	-2.40	-2.09
Trifenüülfosfaat	3.32	3.77	3.83	3.34	3.31	3.33	3.05	3.49	3.48	3.20	3.28	3.39
Trifenüülfosfiinoksiid	2.11	2.82	2.83	2.19	2.86	3.12	2.09	2.65	2.62	1.62	2.20	2.16
Tsefadroksiil				-2.20	-2.00	-1.93	-2.16	-2.14	-2.21	-2.56	-2.43	-1.23
Tümiin	-1.22	-1.19	-1.27	-1.03	-0.91	-0.87	-1.07	-1.14	-1.21	-1.37	-1.41	-1.44
Tümoolftaleiin	3.24	3.51	3.27	3.27	3.33	3.27	3.19	3.40	3.62	3.29	3.22	3.35

Orgaaniline solvent →	1-butanool											
Analüüt↓ Veefaas→	**w**	**s**	**x**	**a**	**as**	**ax**	**h**	**hs**	**hx**	**c**	**cs**	**cx**
β-estradiool	3.05	3.18	3.40	2.65	3.52	3.39	2.83	3.39	3.85	2.28	3.26	3.28
1-naftool	2.54	3.09	3.28	2.30	3.14	3.31	2.34	3.11	3.40	1.88	2.76	3.06
2,4,6-trimetüülpüridiin	1.31	2.23	1.42	-0.92	0.52	-0.21	-0.70	-0.23	-0.22	-0.92	-0.12	-2.21
2-aminobensimidasool	0.97	1.44	1.38	-0.13	1.15	1.35	0.47	1.10	1.38	-0.13	0.91	1.40
3-aminofenool	0.59	0.84	0.75	-0.89	0.01	0.38	-0.14	0.22	0.53	-0.21	0.66	0.89
4-(trifluorometüül)aniliin	2.32	2.90	2.99	1.83	2.17	2.39	0.99			0.94	1.41	1.92
5-nitrobensimidasool	1.56	1.96	1.84	0.98	1.02	0.71	-0.26	0.08	0.26	-0.12	0.24	0.65
8,8-dikinolüüldisulfiid	3.06	3.33	3.69	2.54		1.26	-1.12			1.16	0.23	-0.44
Aniliin	0.98	1.35	1.28	-0.28	0.38	0.52	-0.22	0.21	0.41	-0.50	0.33	0.65
Arbutiin	-0.42	-0.33		-0.37	-0.35	-0.29	-0.31	-0.22	-0.20	-0.35	-0.36	-0.19
Atenolool				0.15	0.31	0.49	-0.61	-0.11	-0.15	-1.02	-0.39	0.07
Bengali roosa				3.34	3.38	3.32	3.09	3.27	3.58	3.01	3.04	3.37
Bensamiid	0.89	1.25	1.16	0.79	1.22	1.20	0.81	1.24	1.21	0.63	1.03	1.36
Bensimidasool	1.30	1.65	1.52	-0.40	0.50	0.57	-0.11	0.37	0.50	-0.45	0.37	0.73
Bensüülamiin	-0.74	0.69	0.45	-0.73	0.31	0.49	-0.24	0.23	0.42	-0.60	0.10	0.43
D-(−)-salitsiin	-0.25	-0.18	-1.10	-0.25	-0.20	-0.41	-0.17	-0.06	-0.15	-0.16	-0.20	-0.05
Diantipüriilfenüülmetaan	2.73	3.01	3.04	2.22	2.89	3.02	1.38	2.69	2.69	0.98	2.31	2.59
Diantipüriilmetaan	1.70	2.65	2.19	1.51	2.53	2.18	0.73	1.66	1.64	0.69	1.56	1.83
Difenüülamiin	2.80	3.20	3.13	2.65	3.02	3.30	2.03	2.48	2.74	2.12	2.83	2.96
Difenüülguanidiin	-0.01	1.83	1.86	-0.07	1.74	1.95	0.93	1.51	1.79	0.04	1.39	1.85
Dimetüülftalaat	1.65	2.21	2.32	1.42	2.19	2.32	1.45	2.10	2.29	1.14	1.83	2.03
Disulfiram	2.93	2.75	3.17	2.66	3.32	3.13	2.73	2.93	3.06	2.26	2.80	2.92
Fenütoiin	2.38	2.89	2.98	2.07	2.87	3.11	2.21	2.82	2.95	1.72	2.54	2.72
Katehhool	1.00	1.38	1.80	0.88	1.29	1.72	0.98	1.35	1.77	0.79	1.26	1.62
Kofeiin	0.24	0.50	0.07	0.24	0.52	0.11	-0.07	-0.20	-0.66	0.10	0.22	-0.00
Kristallviolett	2.45	3.41	2.99	1.77	3.03	2.52	-0.99	-1.05	-1.65	1.05	1.50	-0.01
Lidokaiin	0.44	0.48		-0.43			0.26			-0.21		1.25
Mikonasool	3.14	3.24	3.46	1.82	3.05	3.10	2.47	2.85	2.91	1.45	2.97	3.32
N,N-dietüülnikotiinamiid	0.66	1.22	0.74	0.18	0.24	-0.43	-0.91	-0.68	-0.86	-0.64	-0.57	-0.45
N-bensüül-tsinkonidiiniumkloriid								-0.68	0.61		-0.51	
Norfloksatsiin	-0.71	-0.36		-0.62	0.23		-0.00	0.27	0.00	-0.55	0.04	0.35
Novobiotsiin	3.24		3.81		3.90	3.68	3.58	3.69	3.71	3.13	3.75	3.74
Püridiin	0.82	1.23	0.98	-0.27	-0.61	-1.10	-0.51	-0.71	-1.34	-0.66	-0.42	-0.50
Resortsinool	1.00	1.48	1.88	0.89	1.45	1.75	1.04	1.52	1.84	0.81	1.32	1.54
Rifampitsiin	2.39		3.54	2.19	3.90	3.83	2.82	3.88		1.77	3.73	3.96
Sudaan III	4.05	3.78	3.87		3.86	3.93	3.49		3.54	3.29	4.14	3.63
Sulfametoksasool	0.36	0.93	1.31	1.20	1.63	1.77	0.30	0.71	1.11	0.73	0.97	1.23
Sulfanüülamiid	0.10	0.11	0.18	0.01	-0.02	-0.07	-0.81	-0.79	-0.90			
Tetratsükliin			-0.35	-0.47			0.35					
Trifenüülfosfaat	3.34	3.56	3.37	2.82	3.54	3.30	3.03	3.51	3.56	2.48	3.63	2.67
Trifenüülfosfiinoksiid	2.52	3.29	3.09	2.33	3.32	3.23	2.33	3.39	3.43	1.81	2.74	2.84
Tsefadroksiil				-0.38	-0.34	-0.27	-0.32	-0.23	-0.21	-0.52	-0.37	-0.16
Tümiin	-0.04	-0.03	-0.12	-0.04	-0.05	-0.08	-0.04	-0.04	-0.13	-0.06	-0.11	-0.03
Tümoolftaleiin	2.91	3.44	3.34	2.79	3.34	3.36	3.15	3.30	3.45	2.92	3.31	3.05

Orgaaniline solvent →	Tolueen-metanool 9:1 (v/v)											
Analüüt↓ Veefaas→	**w**	**s**	**x**	**a**	**as**	**ax**	**h**	**hs**	**hx**	**c**	**cs**	**cx**
β-estradiool	1.74	2.35	2.44	1.45	1.96	2.24	1.68	2.13	2.24	1.20	1.58	1.54
1-naftool	1.70	2.24	2.47	1.53	1.90	2.12	1.69	2.05	2.25	1.43	1.72	1.82
2,4,6-trimetüülpüridiin												
2-aminobensimidasool	-2.72	-2.73	-2.78	-3.19		-3.30	-2.92	-3.43	-2.52	-3.24	-2.74	-2.71
3-aminofenool	-1.41	-1.18	-1.19	-1.35								-0.47
4-(trifluorometüül)aniliin	2.11	2.50	2.81	1.71	1.74	1.32	-0.60	-0.82	-1.28	0.77	0.59	-0.55
5-nitrobensimidasool	-0.71	-0.52	-0.52	-1.24	-1.40	-1.87	-4.05	-4.30	-4.86	-2.77	-3.14	-3.80
8,8-dikinolüüldisulfiid	4.25	4.45		3.85	2.87	1.50				2.19	0.66	-1.94
Aniliin	0.85	1.11	1.11	-0.89	-1.09	-1.84	-3.87	-4.07	-4.41	-1.74	-2.68	-3.67
Arbutiin												
Atenolool							-1.26	-1.34		-1.67	-1.73	-1.79
Bengali roosa				3.29	3.44	3.37	3.30	3.29	3.21	3.58	3.29	3.49
Bensamiid	-0.90	-0.66	-0.69	-0.52	-0.19	-0.01	-0.92	-0.70	-0.82	-1.13	-0.94	-1.12
Bensimidasool	-0.83	-0.54	-0.74	-3.96	-3.51	-3.67						
Bensüülamiin	-1.80		-0.78			-3.74						
D-(–)-salitsiin			-3.50	-2.60	-2.20	-1.81					-3.32	-3.17
Diantipüriilfenüülmetaan	2.64	3.23	3.12	2.11	2.44	1.77	-0.54	-0.65	-1.16	0.53	0.30	-1.22
Diantipüriilmetaan	0.93	1.63	1.29	0.76	1.25	0.67	-1.29	-1.53	-2.04	-0.06	-0.26	-1.68
Difenüülamiin	3.61	4.07	4.39	3.60	3.68	3.73	2.84	2.56	1.99	3.38	3.40	2.66
Difenüülguanidiin	-1.77	-1.57	-1.88	-3.61	-2.23	-1.81	-2.21	-2.26	-2.53		-3.57	-3.36
Dimetüülftalaat	2.16	2.54	2.70	1.90	2.25	2.40	2.07	2.42	2.56	1.79	2.10	2.12
Disulfiram	4.43	4.23		4.09	4.43		4.21	3.97		3.54		
Fenütoiin	0.62	1.00	1.06	0.64	1.08	1.56	0.58	0.89	1.09	0.22	0.49	0.55
Katehhool	-1.16	-0.89	-0.63	-1.26	-1.00	-0.62	-1.15	-0.91	-0.60	-1.26	-1.03	-0.74
Kofeiin	-0.28	-0.03	-0.29	-0.38	-0.11	-0.24	-0.68	-1.04	-1.83	-0.68	-0.58	-1.31
Kristallviolett				-1.88	-0.97	-1.13				-3.46	-3.88	-2.95
Lidokaiin	1.43	1.47	-0.38	-2.86	-2.78	-2.50						
Mikonasool	3.59	3.53	2.58	1.81	1.86	1.65	-0.54	0.55	0.79	-0.05	-0.16	-0.39
N,N-dietüülnikotiinamiid	-0.15	0.29	-0.01	-0.73	-0.62	-1.15	-3.61	-4.03		-1.92	-2.51	-3.66
N-bensüül-tsinkonidiiniumkloriid	-4.18	-1.42		-2.69		-3.47		-2.72		-2.97	-2.81	
Norfloksatsiin	-2.98	-3.17										
Novobiotsiin	-1.63	-0.43	-0.20		2.50	3.11				1.42	2.07	
Püridiin	0.34	0.59	0.39	-0.84	-0.51	-1.99	-0.90	-1.47	-0.97			
Resortsinool	-1.84	-1.54	-1.24	-2.01	-1.71	-1.37	-1.96	-1.63	-1.40	-1.96	-1.64	-1.48
Rifampitsiin	0.96	2.30	2.41	0.78	1.20	1.57	0.23		1.75	-0.20		-0.63
Sudaan III	4.47	3.80	4.02	3.85	3.94	3.63	3.86	3.83	4.13	4.05	3.76	3.83
Sulfametoksasool	-1.70	-1.13	-0.95	-0.54	-0.32	-0.15	-2.10	-2.40	-2.95	-1.07	-1.32	-2.14
Sulfanüülamiid	-2.53	-2.35	-2.35	-2.64	-2.65	-2.63	-4.03	-4.49				
Tetratsükliin												
Trifenüülfosfaat	4.84			4.11	4.65		5.12	4.44		3.76	3.82	
Trifenüülfosfiinoksiid	2.22	2.84	2.69	2.26	2.80	3.02	2.08	2.53	2.41	1.73	2.13	1.85
Tsefadroksiil				-1.41	-1.19	-0.76	-1.57	-1.34	-1.89			
Tümiin	-3.18	-3.20	-3.17	-2.66	-2.40	-1.97	-3.43	-3.46	-3.58	-3.56	-3.11	-3.55
Tümoolftaleiin	1.56	2.84		3.64	2.65	3.49	2.78	3.57		3.34	3.65	3.81

Orgaaniline solvent →	Diklorometaan											
Analüüt↓ Veefaas→	**w**	**s**	**x**	**a**	**as**	**ax**	**h**	**hs**	**hx**	**c**	**cs**	**cx**
β-estradiool												
1-naftool	2.09	2.44	2.67	1.85	2.27	2.83	2.05	2.51	2.88	1.86	2.24	2.37
2,4,6-trimetüülpüridiin	1.55			-2.01	-0.97	-0.57	-1.95	-1.26	-1.14		-1.35	-1.25
2-aminobensimidasool	-1.88	-1.74	-1.77	-2.32	-2.09	-1.54	-2.24	-1.88	-2.57	-2.41	-2.31	-3.58
3-aminofenool	-0.64	-0.36	-0.44		-1.48	-1.03	-2.68	-2.80				
4-(trifluorometüül)aniliin	2.52	2.88	2.92	2.13	2.21	2.02	-0.25	-0.37	-0.88	0.98	0.73	-0.22
5-nitrobensimidasool	0.25	0.50	0.34	-0.40	-0.54	-0.90	-2.89	-3.40	-2.28	-1.75	-2.23	-2.38
8,8-dikinolüüldisulfiid	4.12	3.76	3.76	3.48	2.73	1.58				2.41	0.91	-1.40
Aniliin	1.49	1.73	1.74	-0.43	-0.59	-1.11	-2.38	-2.25	-2.30			
Arbutiin				-2.93	-2.98		-2.89	-3.02	-2.69	-2.85	-2.78	-2.99
Atenolool				-0.91	-0.61	-0.21	-1.17	-1.23	-1.29	-1.63	-1.63	-1.70
Bengali roosa		1.90				3.02			3.50	2.95	2.97	2.54
Bensamiid	0.22	0.47	0.46	0.36	0.75	0.90	0.13	0.38	0.24	-0.39	0.17	0.01
Bensimidasool	0.01	0.34	0.13	-1.94	-1.94	-1.98			-2.49	-2.89	-2.85	-2.36
Bensüülamiin	-1.00	-1.06	-1.63	-2.93	-2.43	-1.89	-2.95	-3.21	-3.69	-3.08	-3.16	
D-(–)-salitsiin	-2.81	-2.45	-2.78	-3.03	-2.92	-2.86	-3.44	-3.56	-2.20	-2.07	-2.57	-2.65
Diantipüriilfenüülmetaan	3.14	3.27	3.95	3.30	2.68	3.66	2.09	2.51	2.65	2.37	2.71	1.91
Diantipüriilmetaan	2.80	2.69	2.86	2.78	2.93	2.96	0.87	1.30	0.72	1.65	1.62	0.57
Difenüülamiin	3.14	3.84	3.36	2.88	2.95	3.42	2.98	1.93	2.40	3.49	2.94	3.23
Difenüülguanidiin	-0.79	0.02	0.40	-1.39	0.11	0.83	-0.99	-0.01	0.37	-1.89	-0.36	-0.03
Dimetüülftalaat	2.98	3.39	3.14	2.81	3.10	3.35	2.94	2.92	3.00	2.39	2.85	3.01
Disulfiram		3.64		3.49	3.44		3.73	3.99		3.65	3.65	
Fenütoiin	1.66	2.09	2.30	1.71	2.12	2.70	1.65	2.09	2.31	1.25	1.59	1.67
Katehhool	-0.55	-0.27	-0.00	-0.74	-0.44	0.04	-0.66	-0.41	-0.03	-0.78	-0.53	-0.18
Kofeiin	0.95	1.26	0.96	0.86	1.20	1.02	0.42	-0.08	-0.99	0.52	0.67	-0.30
Kristallviolett	2.11	3.41		2.73	3.18	2.87		-0.27	-1.62		0.80	-0.82
Lidokaiin	1.92	2.56	1.68				-1.85	-0.83	-0.53	-2.22	-1.00	-0.82
Mikonasool	3.44	2.57	2.56	2.76	2.90	2.89	2.09	2.86	3.07	1.06	2.17	2.25
N,N-dietüülnikotiinamiid	1.44	1.84	1.54	0.56	0.52	-0.05	-1.99	-1.95	-2.06	-0.58	-0.95	-1.97
N-bensüül-tsinkonidiiniumkloriid			2.19	-1.56	0.05		-2.54	-1.96	-1.70			
Norfloksatsiin	-0.51	-0.85	-3.12	-2.98	-2.01	-1.62	-2.91	-3.32	-3.06	-2.73		
Novobiotsiin				3.76			3.43			3.02	3.62	3.48
Püridiin	1.00	1.33	1.11	-0.16	-0.38	-0.72	-1.29	-1.30	-1.82	-1.53	-1.60	-1.50
Resortsinool	-1.49	-1.07	-0.78	-1.53	-1.21	-0.79	-1.37	-0.99	-0.71	-1.56	-1.21	-0.88
Rifampitsiin	3.15	3.30	3.28	3.25	3.83	3.55	3.79			2.30	3.41	3.62
Sudaan III	4.63	4.12	3.98	4.14	3.95	3.86	4.20	4.38	3.65	4.39	3.40	4.20
Sulfametoksasool	0.25	0.39	0.98	0.95	1.34	1.42	-0.56	-0.81	-1.24	0.37	0.18	-0.57
Sulfanüülamiid	-0.96	-0.88	-0.73	-1.17	-1.22	-1.36	-2.47	-2.46	-3.24	-1.92	-2.23	-3.14
Tetratsükliin												
Trifenüülfosfaat	3.33	3.87	3.42	3.58	3.10	3.04	4.08	3.73	3.55			
Trifenüülfosfiinoksiid	3.13	2.75	3.14	3.17	3.09	3.11	2.72	3.14	3.26	2.87	3.06	3.38
Tsefadroksiil				-0.91	-0.64	-0.24	-1.23	-1.25	-1.35	-1.52	-1.60	-2.23
Tümiin	-2.12	-2.01	-1.99	-1.54	-1.21	-0.95	-2.00	-1.94	-1.94	-2.05	-2.05	-2.37
Tümoolftaleiin	3.40	3.02	1.67	3.83	2.93		4.26	4.24	3.54	2.73	2.90	

Lisa 5. LogD ennustamiseks kasutatud programmi tekst (keskkond R).

<...> # Algandmete sisestamine; **X_in** - eksperimentaalsete logD väärtuste maatriks (osa väärtusi võib sellest puududa)

x<-ncol(X_in) # Andmemaatriksi veergude (solvendipaaride) arv

y<-nrow(X_in) # Andmemaatriksi ridade (molekulide) arv

mean_X<-rep(NA, x); std_X<-rep(NA, x) # Tühjad vektorid veergude keskmiste ja standardhälvete jaoks

Veergude keskmiste ja standardhälvete leidmine, puuduvate logD väärtuste arvesse võtmata:

for (I in 1:x) {mean_X[I]<-mean(X_in[,I], na.rm=TRUE); std_X[I]<-sd(X_in[,I], na.rm=TRUE)}

X<-X_in # Uus muutuja X edasise töö jaoks

Puuduvate logD väärtuste asendamine veergude keskmistega:

for(I in 1:x) {for(J in 1:y) {if (is.na(X[J,I])) {X[J,I]<-mean_X[I]}}}

#EM algoritm

PC<-3; iter<-200 # EM algoritmi parameetrid: PC-peakomponentide arv, iter – iteratsioonide arv

for (K in 1:iter) {

X_sc<-X; for (I in 1:x) {for (J in 1:y){X_sc[J,I]<-(X[J,I]-mean_X[I])/std_X[I]}} #Andmete tsentreerimine ja skaleerimine

pca<-prcomp(X_sc) # Peakomponentide analüüs

loadings<-pca$rotation[,1:PC] #Laadungite maatriksi osa

scores<-pca$x[,1:PC] #Skooride maatriksi osa

Andmemaatriksi rekonstrueerimine:

X_pred_sc<-scores%*%t(loadings)

X_pred<-X_pred_sc; for (I in 1:x) {for (J in 1:y){X_pred[J,I]<-X_pred_sc[J,I]*std_X[I]+mean_X[I]}}

```
# Algsest maatriksist puuduvate logD väärtuste asendamine uute hinnangutega:

X<-X_in; for(I in 1:x) {for(J in 1:y) {if (is.na(X[J,I])) {X[J,I]<-X_pred[J,I]}}}

} #EM algoritmi lõpp

# Uute molekulide logD väärtuste ennustamine

Xnew<-rep(NA, x)  # Tühi vektor uue ühendi jaoks

<...>  # Vektorisse Xnew sisestatakse olemasolevad eksperimentaalsed logD väärtused

Xn<-Xnew  #  Uus muutuja edasise töö jaoks

for(I in 1:x) {if (is.na(Xn[I])) {Xn[I]<-mean_X[I]}}    # Puuduvate väärtuste asendamine algandmete maatriksi veergude keskmistega

# Iteratiivne tingväärtuste meetod

for (K in 1:100) {

# Andmete tsentreerimine ja skaleerimine; saadud väärtused kasutatakse skooride hinnangutena:

Xn_sc<-Xn; for (I in 1:x) {Xn_sc[I]<-(Xn[I]-mean_X[I])/std_X[I]}

# Andmete rekonstrueerimine:

Xn_pred_sc<-Xn_sc%*%loadings%*%t(loadings)

Xn_pred<-Xn_pred_sc;  for (I in 1:x) {Xn_pred[I]<-Xn_pred_sc[I]*std_X[I]+mean_X[I]}

# Puuduvate logD väärtuste asendamine eelnevas sammus saadud hinnangutega:

Xn<-Xnew; for(I in 1:x) {if (is.na(Xn[I])) {Xn[I]<-Xn_pred[I]}}

}  # Iteratsiooni lõpp

Xnew_predicted<-Xn   #Tulemus: uue molekuli eksperimentaalsed ja ennustatud logD väärtused
```

Lisa 6. Arvutusliku mudeli valideerimise tulemused.

A: „Jäta-üks-välja“ ristkontroll.

Tähistused:

Eksp – korduskatsete tulemuste kogutud standardhälbed.

Ennust. – LOO testi käigus ennustatud ja eksperimentaalsete logD väärtuste vahelised standardhälbed:

A – mudeli ehitamiseks kasutatud kogu treenimisandmestik (44 ühendit).

B – mudeli ehitamiseks kasutatud 38 molekuli, mille standardhälved esialgses LOO testis (veerg A) on alla 1 log ühikut.

C – mudel ehitatud 43 ühendi alusel: 38 treenimisandmestiku molekuli, mille standardhälved esialgses LOO testis on alla 1 log ühikut (veerg B), ja 5 valideerimisühendit.

Ühend	**Standardhälved (log ühikutes)**			
	Eksp.	Ennust. A	Ennust. B	Ennust. C
Kogutud:	**0.08**	**0.75**	**0.66**	**0.79**
Treenimisandmestik:				
Fenütoiin	0.06	0.31	0.32	0.60
β-estradiool	0.08	0.31	0.32	0.59
Bensamiid	0.03	0.34	0.37	0.80
1-naftool	0.11	0.38	0.37	0.64
Trifenüülfosfaat	0.13	0.41	0.37	0.48
Sudaan III	0.10	0.43	0.41	0.88
Difenüülamiin	0.09	0.44	0.45	0.88
Katehhool	0.04	0.45	0.41	0.45
Trifenüülfosfiinoksiid	0.08	0.47	0.48	0.73
N,N-dietüülnikotiinamiid	0.07	0.50	0.72	0.55
Tümiin	0.05	0.51	0.56	0.71
Atenolool	0.05	0.51	0.57	0.64
Arbutiin	0.07	0.53	0.90	0.66
Dimetüülftalaat	0.05	0.53	0.55	0.45
Bengali roosa	0.13	0.54	0.45	***1.36***
Aniliin	0.05	0.57	0.44	0.87
Püridiin	0.05	0.58	0.70	0.57
Sulfanüülamiid	0.04	0.58	0.56	0.38
Tsefadroksiil	0.05	0.58	0.67	***1.23***
4-(trifluorometüül)aniliin	0.06	0.59	0.61	0.94

Ühend	Standardhälved (log ühikutes)			
	Eksp.	Ennust. A	Ennust. B	Ennust. C
Disulfiram	0.13	0.59	0.61	0.87
Diantipüriilmetaan	0.06	0.60	0.85	***1.45***
Resortsinool	0.06	0.62	0.52	0.49
2-aminobensimidasool	0.06	0.63	0.74	0.55
5-nitrobensimidasool	0.06	0.65	0.67	0.36
Tümoolftaleiin	0.10	0.65	0.66	***1.90***
Sulfametoksasool	0.04	0.65	0.69	0.68
D-(–)-salitsiin	0.07	0.67	0.70	0.89
Diantipüriilfenüülmetaan	0.08	0.68	0.79	0.67
Norfloksatsiin	0.07	0.69	0.67	0.49
Mikonasool	0.07	0.72	0.75	***1.34***
3-aminofenool	0.06	0.73	0.73	0.46
Bensüülamiin	0.10	0.76	0.76	0.66
Kofeiin	0.04	0.76	0.78	0.47
Bensimidasool	0.11	0.85	0.96	0.67
2,4,6-trimetüülpüridiin	0.05	0.86	0.95	0.79
Difenüülguanidiin	0.07	0.86	0.98	0.31
Novobiotsiin	0.09	0.95	0.98	0.66
Tetratsükliin	0.04	***1.01***		0.70
Rifampitsiin	0.08	***1.26***		0.74
8,8-dikinolüüldisulfiid	0.10	***1.31***		0.92
Lidokaiin	0.09	***1.34***		0.28
N-bensüül-tsinkonidiiniumkloriid	0.09	***1.34***		0.74
Kristallviolett	0.07	***1.58***		0.42
Valideerimisühendid:				
Bensüülmetüülamiin	0.09			0.78
Bensüülisopropüülamiin	0.08			0.79
Dibensüülamiin	0.07			***1.15***
N-metüülaniliin	0.08			0.44
N-tsükloheksüülaniliin	0.10			0.48

B: Valideerimine sõltumatu andmestiku abil.

Tähistused:

Eksp. – eksperimentaalsed logD väärtused.

Arv. – arvutatud väärtused:

I – mudeli ehitamiseks kasutatud kogu treenimisandmestik (44 ühendit),

II – mudel ehitatud 43 ühendi alusel: 38 treenimisandmestiku molekuli, mille ruutkeskmised hälved LOO testi tulemusena on alla 1 log ühikut, ja 5 valideerimisühendit.

Veefaaside komponentide tähistused:

w - vesi **a** - etaanhape

s - NaCl **c** - sidrunhape

x - $MgCl_2$ **h** - soolhape

Punktiirjoontega on tähistatud ennustamiseks kasutatud solvendipaarid.

Ühend→		N-bensüülmetüülamiin			N-bensüül-isopropüülamiin			Dibensüülamiin			N-metüülaniliin			N-tsükloheksüülaniliin		
Solv.paar ↓		*Eksp.*	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*
Tolueen	w	-1.62	-1.78	-1.64	-0.60	-0.91	-0.94	2.01	1.27	1.63	2.02	1.62	1.85	3.59	3.15	3.45
	s	-1.84	-1.81	-1.95	-0.62	-0.93	-1.04	2.09	1.46	1.75	2.46	1.89	2.06	4.12	3.50	3.75
	x	-3.00	-1.95	-2.25	-2.21	-1.05	-1.24	0.99	1.08	1.52	2.20	1.35	1.81	3.93	2.93	3.51
	a		-2.53	-1.50		-1.38	-1.16	-2.18	0.16	0.33	-0.04	-0.03	-0.21	1.29	1.66	1.45
	as		-2.40	-1.18	-2.03	-1.18	-0.89	-1.27	0.34	0.43	-0.24	0.08	-0.25	1.24	1.86	1.45
	ax	-3.06	-3.01	-2.72	-2.05	-1.82	-1.96	-0.55	-0.54	-0.39	-0.89	-0.91	-0.92	0.64	0.80	0.85
	h		-3.39	-1.88	-3.31	-1.94	-1.98	-2.85	-1.64	-1.69	-3.08	-2.66	-3.07	-1.53	-0.96	-1.35
	hs		-2.58	-0.63	-2.55	-1.02	-1.08	-1.40	-1.16	-1.20	-2.98	-2.46	-2.85	-1.00	-0.80	-1.17
	hx	-3.40	-2.93	-2.08	-2.34	-1.35	-1.99	-0.76	-1.57	-1.56	-3.01	-2.93	-2.88	-0.52	-1.25	-1.10
	c		-2.71	-1.10	-1.70				-0.51	-0.33	-1.28			0.10		
	cs		-3.09	-1.08		-1.67	-1.24	-1.84	-0.97	-0.68	-1.68	-1.77	-1.90	-0.20	-0.02	-0.20
	cx		-3.58	-1.93	-2.18	-2.22	-2.05	-1.05	-1.93	-1.66	-2.55	-2.88	-3.00	-0.62	-1.29	-1.27
2-metüültetrahüdrofuraan	w	-1.06	-0.59	-1.01		0.07	-0.03		0.94	1.52	2.23	0.78	1.70		1.80	2.71
	s	-1.33	-0.11	-0.93		0.48	0.17		1.55	1.90	2.75	1.55	2.25		2.55	3.21
	x	-2.14	-0.39	-1.15		0.18	0.10		1.66	1.97	2.47	1.85	2.32	3.15	2.95	3.40
	a	-2.58	-1.54	-1.88		-0.86	-0.79		0.27	0.52	0.86	0.18	0.45	1.67	1.34	1.59
	as	-1.46	-1.11	-1.43		-0.41	-0.38		0.60	0.74	0.73	0.43	0.57	1.89	1.60	1.69
	ax	-1.53	-1.44	-1.59		-0.62	-0.52		0.25	0.47	0.14	-0.07	0.16	1.41	1.19	1.36
	h	-2.43	-1.83	-1.73		-0.68	-1.17		-0.99	-1.07	-2.12	-2.10	-2.10	-0.92	-0.84	-0.79
	hs	-1.87				-0.48	-0.89		-0.67	-0.68	-1.72			-0.17		
	hx	-1.61	-1.73	-1.67		-0.60	-1.01		-0.89	-0.73	-1.37	-1.99	-1.60	0.04	-0.73	-0.33
	c	-2.02	-1.35	-1.25		-0.65	-0.65		-0.14	-0.21	-0.63	-0.51	-0.74	0.09	0.48	0.24
	cs	-1.19	-1.26	-1.02		-0.42	-0.36		-0.03	0.05	-0.86	-0.58	-0.60	0.41	0.56	0.51
	cx	-0.78	-1.03	-0.84		-0.11	-0.21		-0.20	0.19	-0.72	-1.03	-0.46	0.80	0.05	0.64

Ühend→ / Solv.paar ↓		N-bensüülmetüülamiin			N-bensüül-isopropüülamiin			Dibensüülamiin			N-metüülaniliin			N-tsükloheksüülaniliin		
		Eksp.	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*	*Eksp.*	*Arv. I*	*Arv. II*
Isopropüülatsetaat	w	-1.71			-0.75				0.87	1.59	2.14			3.37		
	s	-1.96	-0.69	-1.62	-0.95	-0.04	-0.37		1.49	1.85	2.62	1.66	2.27	3.46	2.84	3.50
	x	-2.68	-0.57	-1.75	-1.79	0.15	-0.46		1.42	1.69	2.19	1.40	2.06	3.35	2.59	3.32
	a		-2.26	-2.27		-1.41	-1.17	-1.84	0.05	0.28	0.24	-0.02	0.06	1.17	1.43	1.45
	as	-2.10	-1.89	-2.14	-1.46	-1.02	-1.03	-0.18	0.31	0.35	0.10	0.17	0.09	1.24	1.60	1.49
	ax	-1.79			-1.08			0.29			-0.37			0.85		
	h	-2.85	-2.20	-1.94	-2.12	-0.91	-1.42	-1.39	-1.03	-1.13	-2.64	-2.15	-2.23	-1.22	-0.71	-0.73
	hs	-2.43	-1.68	-1.64	-1.70	-0.37	-1.14	-0.59	-0.62	-0.82	-2.31	-1.80	-1.91	-0.62	-0.38	-0.39
	hx	-2.01	-2.55	-2.10	-1.44	-1.20	-1.53	-0.60	-1.32	-1.19	-1.98	-2.49	-2.27	-0.32	-0.96	-0.74
	c	-2.98	-2.73	-2.33	-1.00	-1.58	-1.59	-2.76	-1.00	-0.79	-1.09	-1.70	-1.54	-0.13	-0.19	-0.05
	cs	-2.53	-2.60	-2.31	-0.99	-1.37	-1.55	-0.89	-1.04	-0.80	-1.48	-1.91	-1.60	-0.29	-0.38	-0.08
	cx	-2.31	-2.69	-2.27	-0.87	-1.33	-1.62	-0.46	-1.45	-1.12	-2.02	-2.65	-2.14	-0.44	-1.10	-0.57
1-butanool	w		0.00	-0.13		0.51	0.47		1.51	1.63	1.51	1.56	1.63	2.91	2.42	2.51
	s		0.45	0.23		0.88	0.85		1.97	2.09	2.09	2.11	2.21		2.91	3.02
	x		0.37	0.17		0.89	0.87		1.94	2.04	1.97	2.00	2.05	2.91	2.90	2.94
	a		-0.77	-0.63		-0.22	-0.11		0.61	0.59		0.52	0.29		1.42	1.19
	as		-0.14	-0.07		0.46	0.51		1.37	1.31		1.27	1.01	1.86	2.26	1.98
	ax		-0.21	-0.37		0.45	0.38		1.17	1.27		0.92	1.03	2.06	1.92	2.02
	h		-0.57	-0.13		0.17	0.14		0.13	0.41		-0.50	-0.19	0.88	0.35	0.70
	hs		-0.46	-0.19		0.32	0.28		0.55	0.85		0.00	0.39	1.77	0.97	1.36
	hx		-0.35	-0.31		0.46	0.29		0.53	0.95		-0.11	0.54	2.06	0.85	1.54
	c		-0.71	-0.30		-0.12	-0.03		0.30	0.20		-0.01	-0.35	0.16	0.81	0.47
	cs		-0.32	-0.09		0.35	0.38		0.80	0.99		0.44	0.59	1.71	1.36	1.52
	cx		-0.18	-0.31		0.49	0.37		0.68	0.99		0.20	0.70	1.94	1.06	1.56
Tolueen-metanool 9:1	w	-1.50			-0.54			1.91			1.90			3.54		
	s	-1.95	-1.50	-1.66	-0.88	-0.68	-0.88	1.72	1.58	1.79	2.25	2.00	2.08	3.60	3.49	3.68
	x	-3.44	-1.99	-2.40	-2.51	-1.17	-1.41	1.31	0.99	1.30	2.07	1.34	1.64	3.44	2.84	3.25
	a		-2.41	-1.13		-1.28	-0.96	-2.10	0.24	0.33	0.02	0.06	-0.31	1.25	1.70	1.29
	as		-1.76	-0.33		-0.64	-0.34	-1.36	0.60	0.60	-0.22	0.29	-0.19	1.11	1.86	1.31
	ax	-2.76	-2.77	-1.98	-1.95	-1.57	-1.51	-0.62	-0.39	-0.33	-0.87	-0.83	-1.09	0.47	0.86	0.63
	h		-2.12	-0.04		-0.63	-0.85	-3.13	-0.91	-1.32	-2.58	-2.17	-3.11	-1.74	-0.68	-1.57
	hs		-2.50	-0.30		-0.92	-0.98	-1.54			-2.79			-1.23		
	hx		-2.65	-0.96	-2.34	-0.99	-1.45	-0.99	-1.47	-1.58	-2.93	-2.99	-3.27	-0.69	-1.33	-1.56
	c		-1.79	0.24	-1.39	-0.61	-0.25		0.17	0.11	-1.27	-0.37	-0.96	0.01	1.09	0.40
	cs		-2.52	-0.46		-1.19	-0.85	-2.01	-0.56	-0.53	-1.70	-1.31	-1.80	-0.34	0.30	-0.23
	cx		-2.96	-1.01		-1.62	-1.42	-1.41	-1.40	-1.46	-2.66	-2.35	-2.99	-0.80	-0.82	-1.38
Diklorometaan	w	-1.05	-0.71	-0.82	-0.12	-0.04	-0.20	2.41	1.81	1.92	2.42	2.15	2.12	3.52	3.38	3.43
	s	-1.33	-0.20	-0.61	-0.30	0.36	0.04	1.88	2.22	2.23	2.70	2.66	2.56	3.91	3.77	3.78
	x	-1.55	-0.68	-1.62	-0.53	-0.14	-0.61	1.18	1.88	1.88	2.68	2.39	2.39		3.54	3.67
	a	-2.97			-2.17			-1.02			0.30			1.58		
	as	-1.56	-1.75	-1.81	-0.62	-0.98	-1.06	0.67	0.81	0.59	0.12	1.00	0.41	1.60	2.38	1.85
	ax	-0.88			0.06			1.37			-0.29			1.51		
	h		-2.79	-1.92	-1.95	-1.49	-1.59	-0.80	-0.67	-0.69	-2.11	-1.35	-1.68	-0.36	0.34	0.09
	hs	-1.75	-2.82	-2.11	-0.74	-1.68	-1.63	0.60	-0.51	-0.46	-1.51	-0.90	-1.25	0.79	0.72	0.50
	hx	-1.43	-2.75	-2.05	-0.40	-1.60	-1.58	0.98	-0.72	-0.47	-1.19	-1.26	-1.23	1.14	0.29	0.44
	c	-2.89	-2.43	-2.08	-1.02	-1.37	-1.56	-2.68	-0.05	-0.33	-0.91	-0.30	-0.97	0.43	1.27	0.65
	cs	-1.74	-2.38	-1.83	-0.96	-1.32	-1.32	0.25	-0.02	-0.15	-1.13	-0.28	-0.81	0.67	1.30	0.80
	cx	-1.59	-2.76	-1.95	-0.53	-1.54	-1.54	0.67	-0.65	-0.47	-1.28	-1.24	-1.30	0.96	0.40	0.41

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Sofja Tšepelevitš,

(sünnikuupäev: 30.12.1989)

annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Arvutuslik mudel vedelik-vedelik ekstraktsiooni tulemuste ennustamiseks“,

mille juhendajad on Ivo Leito, Karin Kipper, Joel M. Hawkins ja Koji Muteki,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu alates **02.06.2017** kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus **26.05.2014**