TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
Keemia instituut
Molekulaartehnoloogia õppetool

Tanel-Sigmar Sildoja

Struktuurselt varieeruvate orgaaniliste ühendite toksilisuse kvantitatiivsed struktuur-aktiivsus sõltuvused vetikal *Pseudokirchneriella subcapitata*

Bakalaureusetöö
Keemia eriala
12 EAP

Juhendaja Uko Maran, PhD

Tartu 2018

Sisukord

Sissejuhatus 3
1. Kirjanduse ülevaade 4
1.1 Vetikate olulisus, *Pseudokirchneriella subcapitata* ja toksilisuse andmed 4
1.2 Kvantitatiivne struktuur-aktiivsus sõltuvus 5
1.3 QSAR mudelid *Pseudokirchneriella subcapitata*'le 8
2. Andmekomplekt ja meetodid 10
2.1 Andmed ja nende jaotamine klassidesse 10
2.2 Struktuuride genereerimine 11
2.3 Konformatsiooniotsing 11
2.4 Poolempiirilised arvutused 11
2.5 Deskriptorite arvutamine ja valimi koostamine 12
2.6 Kvantitatiisete struktuur-aktiivsus sõltuvuste tuletamine 12
2.7 Mudelite valik, nende valideerimine ja mudelite rakenduspiirkond 13
3. Tulemused ja arutelu 15
3.2 Mittepolaarse narkoosi järgi käituvad ühendid 15
3.3 Polaarne narkoos 18
3.4 Mittespetsiifiliselt reageerivad kemikaalid 20
3.5 Spetsiifiliselt reageerivad kemikaalid 23
3.6 Mitteklassifitseeritavad kemikaalid 24
4. Kokkuvõte 28
5. Summary 30
Viited 32
Lisad 41
Infoleht 56
Lihtlitsents 57

Sissejuhatus

Mistahes kemikaali bioloogiline aktiivsus või füüsikalis-keemilised omadused on seotud tema struktuuri ning võimalike interaktsioonide olemusega [1,2]. Seetõttu on kindlate omadustega uute ühendite loomisel üks eesmärkidest molekuli struktuuri ja omaduste vaheliste seoste mõistmine ning nende teadmiste rakendamine struktuur-aktiivsus sõltuvuste uurimise ja/või kasutamise näol [1]. Käesoleva uurimuse eesmärk oli leida struktuurselt varieeruvatele orgaanilistele ühenditele kvantitatiivseid struktuur-aktiivsus sõltuvusi (QSAR) modelleerimaks nende toksilisust vetikale *Pseudokirchneriella subcapitata* [3] ja selle tegevuse käigus omandada oskused ja teadmised meetoditest, kuidas tuletada neid sõltuvusi (nn mudeleid). Toksilisuse uurimist vesikeskkondades, ja just vetikatele, peetakse vajalikuks, kuna vetikad primaartootjatena on toiduahela alguseks ning häiritus nende hulgas mõjutab ka teisi troofilisi tasemeid [4–8]. Toksilised ained võivad bioakumuleeruda vetikates ja sealtkaudu jõuda ka kõrgematesse troofilistesse tasemetesse [3,6,9], seega on vetikad olulised ökosüsteemi koosluse osad [3,6,7]. Ka on vetikad võrdlemisi tundlikud toksilistele ainetele [10–13]. Täiendav põhjus QSAR mudelite loomiseks on nende *in silico* ehk arvutipõhine olemus [14,15], mis aitab vältida eetilisi probleeme, mis kaasnevad *in vivo* katsetega, ehk katsetega elusorganismides [16]. Ühtlasi on eksperimentide sooritamine on ressursside suhtes kulukam [17], mis tihti suunab vaatama *in silico* lahenduste poole.

1. Kirjanduse ülevaade

Käesolevas peatükis keskendutakse uurimuse seisukohast olulistele aspektidele. Kirjeldatud on vetikatele ainete toksilisuse uurimise tähtsust ning olulisemaid punkte, mida seejuures silmas pidada. Kirjeldatud on uurimisobjektiks olevat vetikat *Pseudokirchneriella subcapitata* ning välja on toodud põhjused, miks on tegemist toksilisuse uuringutes populaarse organismiga. Lähemalt vaadeldakse ka QSAR metodoloogiat ning põhjuseid, mis teevad taoliste *in silico* mudelite kasutamise tänapäeval aktuaalseks. Välja on toodud ka mõned ajaloolised verstapostid meetodi algusaastatest kuni tänase päevani.

1.1 Vetikate olulisus, *Pseudokirchneriella subcapitata* ja toksilisuse andmed

Primaartootjatena kuuluvad vetikad toiduvõrgustiku kõige madalamasse kihti [5–7]. Häiritus primaartootjate hulgas mõjutab ka teisi troofilisi tasemeid, seetõttu annab kemikaalide toksilisuse uurimine vetikatele väärtuslikku informatsiooni ainete üldise kahjulikkuse kohta vesikeskkonnas [4,5,8]. Toksilisuse uurimine on tähtis ka ökoloogia seisukohast, kuna vetikate liigirohkusel ja arvukusel on oluline roll ökosüsteemi toetamisel [3,18]. Sellele, kas vetikad on ainete toksilisusele tundlikumad kui loomariiki kuuluvad organismid, leiab nii poolt- [10,12,13] kui ka vastuväiteid [19].

Pseudokirchneriella subcapitata (P. subcapitata), veel tuntud ka *Selenastrum capricornutum*'i, *Raphidocelis subcapitata* jt taksonoomiliste nimetuste all [3,20,21], on sirbikujulise ehitusega, mageveekogudes leiduv ainurakne (välja arvatud rakujagunemise ajal) planktoniline rohevetikas [22]. Suure paljunemiskiiruse, tundlikkuse ning testide hea korratavuse tõttu on *P. subcapitata* bioindikaatorina ökotoksikoloogias enimtuntud ja kõige kasutatavam organism [3,23,24]. *P. subcapitata* võrdlemisi kõrget tundlikkust toksilistele ainetele kinnitavad mitmed uurimused [10–13]. Aktiivse ja laia kasutuse põhjuseks võib pidada ka asjaolu, et sellele liigile on avaldatud enim andmeid ainete toksilisuse kohta [11].

Kemikaalide toksilisuse testimisel on tähtis, et kõik analüüsijad järgiksid kindlalt paikapandud juhiseid, kuna nii toimides on kerge võrrelda laboritevahelisi tulemusi [23]. Laialdaselt on kasutusel Majanduskoostöö ja Arengu Organisatsiooni (OECD) juhised [14,25] kemikaalide testimiseks, kus on antud spetsiifilised instruktsioonid katse ettevalmistamiseks ning

aine mõju määramiseks erinevate lõpp-punktide suhtes. *P. subcapitata* puhul on sobilikud OECD poolt kinnitatud testid olemas [25], mistõttu on nende testide andmete kombineerimine lihtne.

Tuleb märkida, et toksilisuse andmed võivad üksteisest suuresti erineda, kui ignoreerida erinevusi lõpp-punktide, liigile omase tundlikkuse ja kokkupuuteaja vahel [8,11]. Laborite-vahelistel korduskatsetel on märgatud mitme suurusjärgu võrra erinevaid tulemusi füüsikaliste ja/või keemiliste erinevuste tõttu katsetingimustes [26,27]. Tihti jääb just kvaliteetsete, standardse protokolli järgi sooritatud usaldusväärsete katsetulemuste puudumine QSAR mudelite loomisel limiteerivaks faktoriks [26,28]. Kindlasti tuleb olla ettevaatlik, kui luua QSAR mudeleid andmetest, mis on saadud kasutades taksonoomiliselt erinevaid organisme [11]. Ka on eksperimentaalsete andmete põhjal loodud mudel vaid nii täpne, kui kvaliteetsed on vastavad eksperimentaalsed lähteandmed [8].

1.2 Kvantitatiivne struktuur-aktiivsus sõltuvus

QSAR on matemaatiline mudel, mis seob sõltuva ehk Y-muutuja, milleks võib olla mingi füüsikalis-keemiline omadus, toksikoloogiline või bioloogiline vastus, sõltumatu(te) ehk X-muutuja(te)ga, mis iseloomustab molekuli struktuuri, nn deskriptorid [1,2,15]. Deskriptoriks võib olla mistahes eksperimentaalne või arvutuslik antud molekulile iseloomulik näitaja, mis on kvantitatiivselt avaldatav või mõõdetav [1]. Deskriptorid rühmitatakse nende olemuse põhjal, mida nad kirjeldavad; näiteks aatomite arv, laengujaotus, mõni füüsikaline omadus, jne [2]. Need rühmad võib omakorda kategoriseerida nn dimensiooni alusel, mis võtab arvesse molekuli struktuuri eri tasemetel ja mille üks võimalikest viisidest on toodud tabelis 1 [29,30].

Tabel 1. Üks võimalustest deskriptoreid dimensionaalsuse järgi rühmitada.

Dimensioon	Võimalikud deskriptorid
1D	Konstitutsioonilised deskriptorid, mõõdetavad molekuli omadused
2D	Topoloogilised, graafiteoorial põhinevad deskriptorid
3D	Steerilised, ruumilised, elektroonsed, kvantkeemilised deskriptorid
4D	Tuuakse 3D-QSAR-isse sisse konformatsioonilised ansamblid

QSAR on klassikaliselt esitatav multilineaarse regressiooni (MLR) võrrandiga,

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n \,, \quad (1)$$

kus Y on sõltuv muutuja (omadus) ja X_1, X_2, X_3 sõltumatud muutujad (deskriptorid) ning a_1 , a_2 , a_3 on neile vastavad koefitsiendid, mis iseloomustavad iga sõltumatu muutuja panust [1]. MLR on üks tuntumaid, tähtsamaid ja enimrakendatavaid statistilisi meetodeid struktuuri ja omaduse vahelise seose leidmiseks ja kirjeldamiseks, kuna on kergesti kasutatav ja saadavad tulemused on hästi reprodutseeritavad ning lihtsasti interpreteeritavad [1,2,31].

QSAR annab võimaluse detailselt tundma õppida kemikaalidega toimuvaid interaktsioonimehhanisme ja läheneda uue ühendi sünteesile ratsionaalselt, saades suunata aine struktuurset ülesehitust sobiva aktiivsuse saavutamiseni [1]. See annab võimaluse ennustada ühendi käitumist (näiteks ravim-retsetpor ineraktsiooni) antud keskkonnas *in silico* [15–17]. Tulemuseks on vähem vaja sooritada *in vitro* ja *in vivo* katseid, mis võivad võtta kaua aega, olla ressursside suhtes kulukad ning tekitada probleeme ökoloogia ja eetika seisukohtadest vaadatuna [15,32,33]. QSAR kui meetod aitab vähendada elusorganismide kasutamist ökotoksikoloogia uurimisel, nagu seda nõuab näiteks REACH [14,15,34], mille alusel peavad kõikidel kemikaalidel, mida Euroopa Liidus toodetakse või sinna imporditakse mahus üle 1 tonni aastas, olema esitatud toksilisuse andmed. Selliste kemikaalide arv jäi 2009. aastal ennustatavalt 68000 ja 101000 vahele [35] ning QSAR ülesanne on anda informatsiooni ja ennustada andmeid nende kohta, kus eksperimentaalsed andmed puuduvad [28,36].

QSAR mudelite tuletamisel ja kasutamisel on oluliseks teetähiseks arvuti kasutuselevõtt [1] ja paljuski viimasel kümnendil ka andmehulkade suurenemine ning viimasest tulenev laiaulatuslik andmebaaside arendus, sh Internetis [15]. Arvuti lubab meil mitmel viisil molekule ning nende omadusi esitada ning vastavaid andmeid kiiresti töödelda [1,37]. Loodud on mitmeid tarkvaralisi lahendusi, mis võimaldavad arvutada molekulaardeskriptoreid ning tuletada nendest kvantitatiivseid ja kvalitatiivseid struktuur-aktiivsus sõltuvuse mudeleid [2,15,38]. Põhimõte on, kui huvipakkuvat omadust saab mõõta, siis saab seda ka ennustada, [17] või anda prognoos, milline võiks omadus olla, kui teha läbi vastavad muutused struktuuris.

QSAR loomine ei koosne vaid nupulevajutusest ning korrelatsiooni saamisest [39], mida tarkvara tänapäeval tihti suudab. Korrelatsioon QSAR mudelis peab olema usaldusväärselt demonstreeritav, ideaaljuhul mehhanistlikult interpreteeritav ning kindlasti valideeritud [40,41].

Juhuslike korrelatsioonide ja ülesobitamise vältimiseks on tihti abi vaid keemiku enda kogemustest [14,39,42,43].

Meetodi ajalised verstapostid:

Esimest korda oletati seost molekuli struktuuri ja tema bioloogilise aktiivsuse vahel juba 19. sajandi keskpaiku, kui Crum-Brown ja Fraser eeldasid, et mingi aine füsioloogiline mõju (Φ) on tema keemilise koostise (C) funktsioon (2) [44].

$$\Phi = f(\mathrm{C}) \tag{2}$$

Vastavalt 1899. ja 1901. aastal avaldasid Meyer ja Overton üksteisest sõltumatult oma tööd, kus nad olid leidnud seose orgaaniliste ainete narkootilise mõju ning (oliivi)õli/vesi jaotuskoeffitsiendi logaritmi vahel, kuid kumbki neist ei tuletanud sellest seosest valemeid [45–47].

1937. aastal esitles Hammett sõltuvust, mis näitas struktuuri mõju reaktsioonidele uurides neid asendatud benseenides,

$$log\left(\frac{k_x}{k_H}\right) = \rho\sigma\ , \tag{3}$$

kus k_x on asendusrühmaga molekuli kiirus- või tasakaalukonstant kindla reaktsiooni korral ja k_H vastava asendamata molekuli oma, ρ on konstant, mis oleneb reaktsioonist, temperatuurist ja keskkonnast ning σ on konstant, mille väärtus sõltub asendajast (Hammett'i asendajakonstant) [47–49].

50-ndate aastate alguses, uurides alifaatsete estrite happe- ja aluskatalüütilist hüdrolüüsi, tõi Taft sisse lineaarse sõltuvuse vabaenergiast [50]. Tema uurimustest pärineb ka esimene steerilist interaktsiooni kirjeldav deskriptor [1,51,52]. Võimalikuks sai lahutada polaarsete, steeriliste ja resonantsiliste efektide panused mudelisse [1,51,52].

Suur läbimurre QSAR valdkonnas tuli 1962. aastal, kui Hansch jt [53] näitasid erinevate fenoksüäädikhappe derivaatide aktiivsust taimekasvuregulaatorina taksonoomilises perekonnas kaer (ld k *avena*) ning said seose,

$$log\left(\frac{1}{C}\right) = 3.36 + 4.08\pi - 2.14\pi^2 + 2.78\sigma\ , \tag{4}$$

kus *C* on kasvule kindlat mõju avaldav kontsentratsioon, σ on Hammett'i asendajakonstant ning $\pi = log\left(\frac{P_x}{P_H}\right)$, kus P_x on asendatud molekuli oktanool/vesi jaotuskoeffitsient, P_H asendamata molekuli oma [53]. Nimetatud töö oli mitmeti tähtis ja uuenduslik [47], kuna näitas, et füsikokeemilised deskriptorid võivad kirjeldada ka keerulisemaid protsesse kui narkoos, ja demonstreeris, et vaja võib minna rohkem kui ühte deskriptorit. Lisaks sellele kasutati uurimuses isooktanooli lipiidi aseainena ning deskriptorina hüdrofoobsuskonstanti. Kuigi näiteks Fieser ja Richardson [54] olid ennegi märganud, et sõltuvus tõuseb maksimumini ja siis jälle langeb, kasutati alles Hansch'i jt töös [53] selle kirjeldamiseks esimest korda ruutvõrrandit [47]. Hiljem on sellise kujuga sõltuvuste teket seletatud mitmete hüdrofiilsete ja hüdrofoobsete barjääridega retseptorini jõudmisel, mida molekulid peavad ületama, et jõuda sihtmärgini [55]. Kõige tõenäolisemalt jõuavad retseptorini need ühendid, mille hüdrofoobsus on antud olukorras optimaalne [55].

Tänapäevase QSAR-i alguseks peetakse eelneva sajandi 60-ndaid aastaid tihti just Hansch'i jt töö [53] märkimisväärse panuse tõttu sellesse valdkonda [14,30,32,47].

1.3 QSAR mudelid *Pseudokirchneriella subcapitata*'le

Lee ja Chen [56] leidsid bensoehapetele hea ennustamisvõimega mudeli oktanool/vesi jaotuskoefitsiendiga (log K_{OW}) (R^2 = 0.998; 48h EC_{50} (aine kontsentratsioon, kus mõju on märgata pooltel organismidest); kasvu kiiruse inhibeerimine). Kombineerides log K_{OW} ja hüdroksüülrühmade arvu molekulis (N_{OH}) saadi veel üks mudel ilma hälbivate liikmeteta, mis näitas, et toksilisus kasvab hüdrofoobsuse suurenedes (R^2 = 0.965). Mittepolaarsed narkootilised kemikaalid [23,57] omavad reeglina positiivset korrelatsioon hüdrofoobsuse ja toksilisuse vahel ning kirjandusest leiab seda kinnitavaid mudeleid [8,10,23].

QSAR-i mudeleid vetikate toksilisuse erinevatele lõpp-punktidele on tuletatud ka Eestis, näiteks 2013 aastal uurisid Aruoja jt [23] 50 mittepolaarse ja 58 polaarse kemikaali toksilisust. Mittepolaarsete kemikaalide korral saadi hea korrelatsioon log K_{OW}- ga (R^2 = 0.95; 72h EC_{50}; kasvu kiiruse inhibeerimine), polaarsete korral tuletati kolmeparameetriline mudel log K_{OW}-ga, aatomite arvuga normaliseeritud tekkesoojusega ($\Delta H_f/\#_{atoms}$) ja molekulmassiga (MW). (R^2 = 0.92; 72h EC_{50}; kasvu kiiruse inhibeerimine). Paljud neist kemikaalidest on tööstuses olulised ning suurte tootmismahtudega [23]. Veel on asendatud aniliinide ja fenoolide korral leitud, et

toksilisus oleneb asendajate arvust, ehk mida rohkem asendajaid, seda suurem hüdrofoobsus ja ka toksilisus, samuti tüübist, näiteks klooriga asendatud ühendid on toksilisemad kui alküülasendatud ühendid, ja positsioonist, nimelt para-asendis asendaja suurendab toksilisust kõige enam [28], ning et vetikate korral oli aniliinide ja fenoolide toksilisus võrdväärne. Nimetatud töös näidati, et fenoolide toksilisus korreleerub hästi log K_{OW}-ga (R^2 = 0.85), aniliinide korral saadi kasin mudel (R^2 = 0.55).

Fenoolide (v.a. nitrofenoolide) puhul on hea mudeli (R^2 = 0.93; 96h EC_{50}; kasvu kiiruse inhibeerimine) saanud ka Lee jt [58]. Klorofenoolide korral on Chen ja Lin [59] leidnud positiivse korrelatsiooni log K_{OW}-ga ning negatiivse korrelatsiooni happe dissotsatsiooni-konstandi kümnendlogaritmiga (pK_a), R^2 vastavalt 0.94 ja 0.96 (mõlemal 48h EC_{50}; kasvu kiiruse inhibeerimine).

Klorobenseenide korral on Hsieh jt leidnud, et log K_{OW}-ga mudelis (R^2 = 0.943; 48h EC_{50}; kasvu kiiruse inhibeerimine) toksilisus suureneb klooriaatomite arvu suurenemisel, mis langeb kokku Aruoja jt [23] tulemustega. Huang jt [13] on uurinud nitriilide toksilisust ja saanud võrdlemisi hea ennustusvõimega struktuuri ja omaduse vahelise sõltuvuse (R^2 = 0.92; 72h EC_{50}; kasvu kiiruse inhibeerimine). Deskriptoritena on kasutatud mudelis log K_{OW}'i ja madalaima täitmata molekulaarorbitaali energiat (E_{LUMO}). E_{LUMO} koeffitsiendi absoluutväärtus oli palju suurem, kui log K_{OW} oma (koefitsiendid vastavalt -52.96 ja -0.075), mis viitab nitriilide spetsiifilisele toksilisusele, kusjuures nitriilide toksilisus hüdrofoobsuse suurenedes väheneb (log K_{OW} koefitsient on negatiivne).

Pramanik ja Roy [60] tuletasid erinevatele kemikaalidele kaks multilineaarse regressiooni mudelit (R^2 = 0.774 ja 0.782; 48h EC_{50}; kasvu kiiruse inhibeerimine), mis näitasid, et üldiselt on toksilisemad suurema molekulaartihedusega (molekulaarmassi ja molekulaarruumala suhe) ühendid, kondenseerunud polüaromaatsed ühendid ning hereroaromaatsed molekulid. Nendest esimene mudel baseerus topoloogilistel ja fragmendimõhistel deskriptoritel ning molekulaar-tihedusel, teise mudelisse toodi sisse ka log K_{OW}.

Levet jt [33] asetasid uurimisobjektiks erinevatesse aineklassidesse kuuluvad orgaanilised solvendid ja said tulemuseks võrdlemisi hea mudeli (R^2 = 0.744; 72h EC_{50}; kasvu kiiruse inhibeerimine). Nagu Huang'i jt uurimuses [13], olid ka seal deskriptoriteks log K_{OW} ja E_{LUMO}. Viimase positiivset korrelatsiooni EC_{50} negatiivse kümnendlogaritmiga (pEC_{50}) seletati Levet'i töös [33] elektrofiilsemate kemikaalide suurema toksilisusega.

2. Andmekomplekt ja meetodid

2.1 Andmed ja nende jaotamine klassidesse

Töös kasutatud andmeseeria pärineb Jaapani keskkonnaministeeriumi kuni 2016. aasta märtsini sooritatud eksperimentaalsetest mõõtmistest, kus kokku oli andmeid 659 ühendi kohta [61]. Nendest valiti välja orgaanilised ühendid. Saadud valimist eemaldati kõik soolad, hüdraadid ja kvaternaarsed amiinid. Allesjäänud ühenditest valiti välja need, millel oli olemas kasvu kiiruse inhibeerimise eksperimentaalne 72h EC_{50} väärtus meile huvipakkuvale vetikale *P. subcapitata* [25], mis olid antud mg/l ühikutes. Need arvutati ümber ühikutesse mmol/l ning tulemuse pöördväärtuse kümnendlogaritmi (log ($1/EC_{50}$)) kasutati töös modelleeritava omadusena. Duplikaatide esinemise korral andmeseerias valiti ajaliselt hiljem sooritatud katse tulemus. Pärast neid etappe jäi valimisse alles 342 ühendit (Tabel L3).

Allesjäänud ühenditest moodustati valimid nn Verhaar'i klasside [57,62,63] järgi (Tabel 2), mis olid programmi Toxtree versioon 2.6.6 [64,65] abil juba enamikele ainetele Furuhama jt poolt määratud [66]. Kasutatud oli Verhaar'i klassifikatsiooni modifitseeritud versiooni, mis spetsiaalselt tuletatud reegleid kasutades molekulid struktuuri põhjal klassidesse jaotab [67]. Ühenditele 195, 222 ja 301 (Tabel L3), millel klassifikatsioon puudus, määrati Verhaar'i klass ise sama tarkvara kasutades. Iga klassi ühendid jaotati treening- ja valideerimisvalimiks nii, et nad reastati log ($1/EC_{50}$) järgi ning loeti neljaks. Iga teine (kolmanda klassi puhul neljas) ühend valiti valideerimisvalimisse, kusjuures selliselt, et treeningvalimisse jääks suurima ja vähima log ($1/EC_{50}$) väärtusega ühendid [1].

Tabel 2. Verhaar'i klasside selgitused

Klass	Kirjeldus
1	Inertsed kemikaalid, mille vastastikune mõju ei ole spetsiifiline. Sellist toksilist efekti, mis on täielikult mittespetsiifiline ning sõltub vaid ühendi hüdrofoobsusest nimetatakse mittepolaarseks narkoosiks, või ka nn baassirge toksilisus (*baseline toxicity*).
2	Vähem inertsed kemikaalid, mis on toksilisemad, kui võiks eeldada vaid hüdrofoobsusele tuginedes. Seda efekti nimetatakse polaarseks narkoosiks.
3	Reaktsioonivõimelised kemikaalid, mis reageerivad mitteselektiivselt spetsiifiliste struktuuridega biomolekulides või mis metabolismi toimel muunduvad toksilisemateks ühenditeks (nn bioaktivatsioon).
4	Spetsiifilise toimega kemikaalid, mis mõjutavad organismis kindlaid retseptormolekule.
5	Kemikaalid, mida ei saa antud reeglite järgi eelnevatesse klassidesse jagada ning vajavad toimemehhanismi kindlaksmääramiseks edasist uurimist alternatiivsete eeskirjade järgi.

2.2 Struktuuride genereerimine

Andmeseerias oli iga molekuli jaoks vastav CAS (*Chemical Abstracts Service*) number [68], mille järgi laeti alla vastav SMILES (Lihtsustatud molekulaarsisendi joonkirje süsteem - *Simplified Molecular-Input Line-Entry System*) kood [69] *EPI* (*Estimation Program Interface*) *Suite* [70] tarkvarapaketist. Kui tarkvarapaketis SMILES koodi ei olnud või saadi sobimatu tulemus, kontrolliti selle olemasolu või otsiti korrektset koodi keemiaandmebaasist *PubChem* [71]. SMILES koodist genereeriti igale struktuurile vastav Z-maatriks [37] kasutades *Open Babel* [72] tarkvara. Saadud ühendite vastavust andmeseerias toodutele kontrolliti visuaalselt programmidega *Maestro* (*Schrödinger*-i tarkvarapakett) [73] ja *Molden* [74]. Vigade esinemisel Z-maatriksis parandati need seda käsitsi korrigeerides (*Molden*) või ehitati molekul aatomhaaval üles (*Maestro*). Juhul, kui andmeseerias ei olnud ainel eristatud optilist ja/või geomeetrilist isomeeriat [75], ei määratud seda ka töös.

2.3 Konformatsiooniotsing

Saadud 342 ühendile teostati konformatsiooniruumi otsing *MacroModel* (*Schrödinger*'i tarkvarapakett) [76] programmiga. Konformatsiooniruumi analüüsiks kasutati PRCG meetodit [77], gradiendi normiks valiti 0,05 kJ mol^{-1} $Å^{-1}$ ning maksimaalne analüüsitavate konformeeride arv oli 2500. Nendest valiti kõige madalama energiaga konformeer, nn globaalses miinimumis [37] olev struktuur, kuna see on kõige tõenäolisem molekuli kuju [1]. Globaalses miinimumis oleva konformeeri kasutamine tagab ka arvutustulemuste reprodutseeritavuse [23].

2.4 Poolempiirilised arvutused

Poolempiirlised kvantkeemilised arvutused teostati tarkvarapaketiga *MOPAC* [78], kasutades geomeetria optimeerimiseks BFGS (*Broyden–Fletcher–Goldfarb–Shanno*) meetodit [79–82] ja AM1 (*Austin Method 1*) parametrisatsiooni ning selle laiendusi halogeenidele, väävlile ja fosforile [83–86]. Selleks lisati konformatsiooniruumi otsingust saadud sisendfailile UNIX skripti abil tabelis L1 väljatoodud märksõnad ning käivitati tarkvara. Esimeses poolempiiriliste arvutuste etapis teostatud geomeetria optimiseerimisele järgnes täiendav arvutus uute märksõnadega (Tabel L2), mille abil arvutati erinevad kvantkeemilised, elektrostaatilised, energeetilised ja steerilised parameetrid [78], mis olid aluseks järgnevas etapis molekulaardeskriptorite arvutamisel.

2.5 Deskriptorite arvutamine ja valimi koostamine

Saadud optimeeritud geomeetriad koos poolempiiriliselt arvutatud parameetritega ning log ($1/EC_{50}$) väärtustega sisestati programmi *CODESSA PRO* [87], mille abil arvutati välja mitmed konstitutsioonilised, topoloogilised, geomeetrilised, elektrostaatilised ja kvantkeemilised deskriptorid [2,87], millest omakorda moodustati valimid. Deskriptorivalimite moodustamiseks ei ole enamasti kindlaid eeskirju ning otsustamaks deskriptori sobivust antud situatsioonis peab keemik tihtipeale kasutama omaenda kogemustest tulenevaid ratsionaalseid otsuseid ja intuitsiooni [42,87]. Kuna deskriptorite valim peab olema kooskõlas uuritavate ühendite omadustega [2], valiti käesolevas töös välja nn kogu molekuli kirjeldavad deskriptorid [2], mis omakorda jagati kahte rühma selle alusel, kas rühmas esineb laengujaotust kirjeldavaid deskriptoreid *MOPAC*-ist või Zefirov'i eeskirjade põhjal [2,87]. Kuna Zefirov'i laengujaotusskeemil põhinevaid deskriptoreid sisaldavad valimid andsid süstemaatiliselt korrelatsioonikordaja ruudu R^2 [1,2,87] põhjal paremaid tulemusi, otsustati edasises analüüsis deskriptorivalimites kasutada just neid. Välise deskriptorina kasutatavad eksperimentaalsed log K_{OW} väärtused laeti alla *EPI Suite* [70] tarkvarapaketist kasutades CAS numbreid. Kui vastet ei leitud, kasutati samas paketis olevat programmi *KOWWIN* [88] log K_{OW} väärtuste arvutamiseks.

2.6 Kvantitatiisete struktuur-aktiivsus sõltuvuste tuletamine

Iga treeningvalimi jaoks tuletati programmis *CODESSA* PRO kasutades nn parima multilineaarse regressiooni meetodit, BMLR (*best multilinear regression*) [31,87], mis kasutab regressioonivõrrandi leidmiseks samm-sammulist mudeli deskriptorite valikut (*step-wise forward selection*). BMLR protseduuri korral *CODESSA PRO* tarkvarapaketis [87] moodustatakse kõigepealt kahekaupa korrelatsioonid üksteisega mittekorreleeruvate deskriptorite vahel. Saadud parimatele korrelatsioonidele lisatakse iga järgneva sammuga üks deskriptor, kontrollides statistiliste parameetrite (F-test, kollineaarsuse kontroll, jne) [1,2] alusel nende sobivust võrrandisse. Iga deskriptorite arvu puhul antakse välja parim võrrand, kuni kas saavutatakse etteantud korrelatsiooni paranemise piirväärtus (ΔR^2) kahe järjestikuse deskriptorite arvuga korrelatsioonivõrrandi vahel või etteantud maksimaalne deskriptorite arv.

Üheparameetrilise sõltuvuse leidmiseks kasutati HMPRO (*Heuristic Method Professional*) lähenemist [89], mis on *CODESSA PRO*-le omane meetod. Meetod kombineerib ta endas heuristilised ja BMLR meetodid [31], modifitseerides HMPRO parameetreid on võimalik

neid ka jäljendada. Meetodis viiakse kõigepealt läbi deskriptorite elimineerimine statistiliste parameetrite (R^2, F, s) [1,2] järgi. Eemaldatakse ka deskriptorid mis korreleeruvad mõne teise deskriptoriga või mille väärtus jääb ühendite lõikes konstantseks. Järgnevalt arvutatakse kahe deskriptoriga mudelid. Statistiliste parameetrite (R^2, F, s) järgi lisatakse neile mudelitele deskriptoreid, kuni saavutatakse nende maksimaalne arv (käesolevas töös sätiti see üheks), mispeale protseduuri läbiviija saab teatud arvu parimaid mudeleid.

Paremate ja usaldusväärsemate sõltuvuste saamiseks oli vajalik ka olenevalt ühendite klassist modifitseerida esmast deskriptorite valimit, et vältida deskriptorite valimis ebaühtlase punktijaotusega (normaalsest jaotusest kaugel) deskriptoreid ühes mudelis või antud valimi korral ilma sisuta ning mitteinterpreteeritavate deskriptorite tulekut mudelisse [26,90].

2.7 Mudelite valik, nende valideerimine ja mudelite rakenduspiirkond

Iga klassi kohta saadud mudelite hulgast valiti välja parim kriteeriumi järgi, kus järgneva deskriptori lisamise korral mudel enam R^2 järgi märgatavalt ei paranenud, $\Delta R^2 < 0.02$ [23,91] ja seejärel analüüsiti deskriptorite ning uuritava omaduse vahelist seost. Statistilisteks parameetriteks olid korrelatsioonikordaja ruut, R^2, mis näitab regressiooni kvaliteeti ehk ligilähetust ideaalsele korrelatsioonile (R = 1); ristvalideeritud korrelatsioonikordaja ruut, R^2_{CV}, mis näitab mudeli ennustusvõimet; F on Fisheri kriteerium, mis aitab võrrelda eri arvu deskriptoritega võrrandite kirjeldusvõimet; ning s^2 ehk standardhälbe ruut, mis näitab mudeli kvaliteeti (mida väiksem, seda parem). Kuna statistilistest parameetritest mudeli usaldusväärsuse hindamiseks ei piisa, sooritati ka väline valideerimine, kus ühendid jagati treening- ja valideerimisvalimiteks ning uuriti nende statistikuid (R^2_{val}, s^2_{val}) [23,92–94] ja hinnati tuletatud mudelite ennustusvõimet (Graafik L1a-L6a).

Mudelite rakenduspiirkond on antud töös määratud ära Verhaar'i klassidega. Täiendavalt sellele analüüsiti rakenduspiirkonda ja vaadeldi tuletatud mudelite kvaliteeti Williams'i graafikute [95,96] abiga. Williams'i graafikul esitatakse standardiseeritud jäägi (standardhälbega normaliseeritud ennustuse viga) ning nn omapära [95,96]. Iga ühendi omapära arvutatakse mudeli molekulaarsetest deskriptoritest: $h=x^T(X^TX)^{-1}x$ [95,96], kus x on prognoositava ühendi deskriptori reavektor ja X on treeningvalimi ühendite deskriptorite maatriks. Suurema kui kriitilise omapäraga (defineeritud kui $3p/n$, kus p on ühe võrra suurem kui deskriptorite arv ja n on treeningvalimi ühendite arv) [96] ühendid on struktuurselt erilised [23] ja sellistel treeningvalimi

ühenditel võib olla ebaproportsionaalselt suur mõju regressioonisirge kujunemisele ja see mõjutab mudeli usaldusväärsust [96], mistõttu seda tuleb mudeli kasutamisel analoogsete ühendite puhul silmas pidada.

3. Tulemused ja arutelu

3.2 Mittepolaarse narkoosi järgi käituvad ühendid

Modifitseeritud Verhaar'i reeglite järgi osutus mittepolaarse narkoosi (klass 1) järgi käituvaks 80 ühendit, nendest treeningvalimisse kuulus 60 ja valideerimiseks jäi 20 molekuli. Siinsed kemikaalid käituvad eeldatavasti nn baassirge toksilisuse mehhanismi järgi [57], mida võib mõista kui häiritust bioloogiliste membraanide talitluses [97]. Klassi kuulusid mittepolaarsed, alifaatsed (sh tsüklilised) ja ka aromaatsed ühendid, millel esines arvukaid kõrvalrühmi, muutes ühendeid potentsiaalselt polaarsemaks ja/või andes võimaluse klassile mitteomapäraselt vesiniksidemete moodustumiseks. Samuti tuleb ka märkida, et Toxtree's olev modifitseeritud Verhaar'i klassifikatsioon ei jaga aineid täieliku korrektsusega, mistõttu ei pruugi klassides olevad ained täpselt nii käituda, nagu klassi järgi arvata võiks [98].

Parim QSAR saadi nelja deskriptoriga (Tabel 3), st järgnevate deskriptorite lisamisel mudel märgatavalt ei paranenud, $\Delta R^2 < 0.02$ [23,91]. Arvestades, et valimil on suur hajuvus, ei ole selget keemilist trendi, ained kuuluvad paljudesse erinevatesse aineklassidesse ja tegemist ei ole ka homoloogilise reaga, saab mudeli statistikute põhjal (Tabel 3) järeldada, et tegemist on võrdlemisi hea ja ootuspärase tulemusega.

Tabel 2. Nelja deskriptoriga QSAR mittepolaarse narkoosi järgi toimivatele ühendite puhul. *a*-koefitsient, s-koefitsiendi hälve, *t* näitab *t*-testi väärtusi (N = 60; n = 4; R^2 = 0.8145; R^2_{CV} = 0.7619; F = 60.3746; s^2 = 0.2495; R^2_{val} = 0.7381; s^2_{val} = 0.9832).

Nr	*a*	s	*t*	Nimetus	Tähis
1	0.669155	0.0964121	6.94057	Oktanool/vesi jaotuskoeffitsient (*Octanol-water patrition coefficient*)	log K_{OW}
2	0.0464001	0.00987523	4.69863	Keskmine aatommass (*Average atom weight*)	M_r
3	0.0139802	0.00321451	4.34908	ALFA polariseeritavus (*ALFA polarizability (DIP)*)	α
4	37.0707	11.9872	3.09251	Vesiniksideme aktseptoritest sõltuv vesiniksidemete doonorite pindala normaliseeritud kogu molekuli pindalaga (*HA dependent HDSA-2/TMSA (Zefirov PC)*)	$HDSA2^{HAdep}{}_{TMSA(Z)}$
0	-2.4439	0.298289	-8.19305	Vabaliige (*Intercept*)	b

Nagu eeldatud [8,10,23,99], omas võrrandis väga suurt osakaalu nii koefitsiendi kui ka *t*-testi (näitab võrrandi liikme olulisust) [2,100] järgi log K_{OW}. Deskriptoril on omadusega positiivne korrelatsioon ehk mida suurem on deskriptori väärtus, seda suurem on vastava ühendi toksilisus. Sellist korrelatsiooni on analoogsetes andmeseeriates tuvastatud ka varasemalt [8,10,23,101]. Log K_{OW} andis ka kõige parema üheparameetrise mudeli (N = 60; n = 1; R^2 = 0.667; R^2_{CV} = 0.611; F = 115.969; s^2 = 0.425; R^2_{val} = 0.6768; s^2_{val} = 2.7464). Log K_{OW} kirjeldab absorbtsiooni ja biosaadavust, kus oktanool imiteerib bioloogilistes süsteemides esinevat lipiididest koosnevat kaksikkihti (membraani) ja näitab kui suur on tõenäosus, et aine saab tänu lipofiilsetele omadustele läbida rakuseinu ning seal mõjutada organismi [30,33,47].

Olulisuselt järgmiseks, teiseks deskriptoriks võrrandis, osutus konstitutsioonilisete deskriptorite hulka jääv keskmine aatommass (M_r) [87], millel on ka positiivne korrelatsioon omadusega. Deskriptorit täpsemini uurides võib järeldada, et kuna esimese klassi ühendid koosnesid enamasti süsinikest ja vesinikest, võib kõrgema deskriptori väärtuste korral eeldada, et esineb heteroaatomit (O, N, P, S) või halogeeni [36], kusjuures hapniku ja/või lämmastiku esinemist struktuuris võib seostada vesiniksidemete tekke võimalusega, mis omakorda on seotud membraani läbimisvõimega [102] ja mittekovalentsete ligand-membraan/retseptor interaktsioonidega [94].

Kolmas deskriptor võrrandis oli α-polariseeritavus (α) [103] ja neljas vesiniksidemete aktseptoritest sõltuv vesiniksidemete doonorite pindala, mis on kaalutud vesiniksidemete doonoraatomite osalaenguga ja normaliseeritud solvendile ligipääsetava molekuli kogu pindalaga ($HDSA2^{HAdep}{}_{TMSA(Z)}$) [30]. Varasemalt on näidatud, et esimene neist on seotud hüdrofoobsuse ning selle kaudu bioloogiliste vastustega organismis [103,104], kusjuures suuremat polariseeritavust seostatakse suurema hüdrofoobsusega [105]. Teine deskriptor neist näitab võimalust osaleda vesiniksidemetes, mõlemad on bioloogiliste membraanide läbimisel olulised omadused [94,102,104]. Vesiniksidemed võivad membraani läbimist nii soodustada kui ka takistada olenevalt konkreetsest süsteemist [102] ja üldiselt peetakse vesiniksidemeid bioloogilistes süsteemides molekulide käitumist oluliselt mõjutavateks teguriteks [30]. $HDSA2^{HAdep}{}_{TMSA(Z)}$ korral on esimese klassi ühenditel näha positiivset korrelatsiooni, mis viitab et mida suurem on vesiniksidemete doonorite suhteline pindala, seda suurem on üheni toksilisus. Vesiniksidemeid moodustavatel ühenditel on peale omavahel ja polaarse solvendi molekulidega seostumise ka tendents akumuleeruda membraani pinnale [94], juba see võib mingil määral

mõjutada organismi talitlust [99]. Ka on teada, et ligand-membraan/retseptor interaktsioonid sõltuvad mittekovalentsetest, st vesiniksidemetest [106].

Valideerimisvalimile arvutati treeningvalimi põhjal saadud parima võrrandiga (Tabel 3) ennustatud väärtused ja esitati need graafikul L1a. Nii mudeli ristvalideeritud korrelatsiooni-koefitsiendi ruudu (R^2_{CV} = 0.7619) ja valideerimisvalimi korrelatsioonikordaja ruudu (R^2_{val} = 0.7381) põhjal võib öelda, et mudelil on hea ennustusvõime, ka on nad lähedased mudeli enda vastavale parameetrile (R^2 = 0.8145) [23,93]. Visuaalsel vaatlusel on näha, et valideerimisvalimi ühendid langevad enamjaolt alasse, kus on treeningvalimi ühendid, kuid leidub ka kõrvalekaldujaid, mille täpsemaks analüüsiks moodustati ka nn Williams'i graafik (Graafik L1b). Graafikut L1b analüüsides selgub, et esimeses klassis on rida tagasihoidlikke kõrvalekaldujaid mis jäävad standardiseeritud jäägi vahemikku 2-3. Nendest viis ühendit: süsiniktetrakloriid (3), orto-ksüleen (139), 4-tert-butüültolueen (170), tridetsaan-1-ool (305) ning trietüülamiin (335) kuuluvad treeningvalimisse ja kaks (metüültsükloheksaan (258) ja triklosaan (551)) valideerimisvalimisse. Standardiseeritud jäägi järgi hälbimisel võib üheks võimalikuks põhjuseks olla eksperimendi täpsus. Samas on need ühendid aga kõik tagasihoidlikud kõrvalekaldujad, mistõttu saavad endiselt olla rakendatavad võrrandi kasutamise juures. Kahe treeningvalimi ühendi (adipiinhape (351) ja indeno[1,2,3-cd]püreen (395)) omapära on natuke suurem kui kriitiline omapära, mis viitab nende struktuuride erinevusele võrreldes teiste treeningvalimi ühenditega. Kõrvalekaldujaid standardiseeritud jäägi alusel ja suurema omapära väärtusega kui kriitiline omapära oli kolm valideerimisvalimi ühendit: oblikhape (389), 2,2-bis[4-(2-hüdroksüetoksü)fenüül]propaan (489) ja N-metüül-N,N-bis(2-dimetüülaminoetüül)amiin (545), kusjuures kaks esimest neist on standardiseeritud jäägi suhtes tugevad kõrvalekaldujad (standardiseeritud jääk > 3) [36] (Graafik L1b, Tabel L3). Mudeli rakendamisel ennustamiseks tuleb tähelepanelik olla viimaste ühendite analoogide suhtes, kuna need väljuvad mudeli rakenduspiirkonnast, mistõttu ennustus ei pruugi olla usaldusväärne ja seda tuleb kontrollida teiste tõenduspõhiste meetoditega.

On näha, et kõrvalekaldujate hulgas olevad ühendid omavad teatud struktuurset või olemuslikku eripära, mida poleks oodata mittepolaarse narkoosi järgi käituvatelt kemikaalidelt. Nendest ühend 3 (süsiniktetrakloriid) on täielikult kloreeritud süsivesinik. Eelnevalt on märgatud kloreeritud benseenide suuremat toksilisust [10], metaani korral on korrelatsiooni toksilisuse ja suurema klooriaatomite arvu vahel samuti varemgi märgatud [107], ning teada on, et

molekulaarpinnal suurt negatiivset osalaengut omavad molekulid on toksilisemad [23]. Lisaks on tegemist võrdlemisi väikese molekuliga, mis võib suurendada membraani läbimise võimet. Orto-ksüleeni (139) ja 4-tert-butüültolueeni (170) puhul on ilmselt tegemist eksperimendi täpsusest tingitud kõrvalekaldumisega, mis toob nad tagasihoidlikke kõrvalekaldujate hulka. 2,2-bis[4-(2-hüdroksüetoksü)fenüül]propaan (489) on selgelt struktuurselt väga eripärane ühend võrreldes teiste andmekomplektis olevate ühenditega, põhjustades nii omapära järgi väljajäämist mudeli rakenduspiirkonnast kui ka tugevat kõrvalekaldumist. Tridetsaan-1-ooli (305) puhul aitab arvatavasti pikast alifaatsest sabast tingitud suurem hüdrofoobsus toksilist efekti suurendada, selliste alkoholide hälbimist on varemgi märgatud [23]. Amiinid trietüülamiin (335) ja N-metüül-N,N-bis(2-dimetüülaminoetüül)amiin (545) võivad tavapärastest esimese klassi ühenditest moodustada rohkem vesiniksidemeid, mis võivad nende ühendite puhul kutsuda esile amiinnarkoosi [33]. Ühendi 545 teeb eriliseks veel tema unikaalne struktuur, olles ka selgelt suurem ühend kui teised selle andmekomplekti ühendid. Metüültsükloheksaan (258) on teadaolevalt veeorganismidele väga toksiline [108], kuigi struktuurilt on ta lihtsalt tsükliline süsivesinik lühikese alifaatse kõrvalrühmaga. Triklosaani (551) korral on teada, et tegemist on bakteri- ja seentevastase kemikaaliga [109] ning on võimalik, et see omab ka vetikate korral toksilist efekti. Adipiinhape (351) ja oblikhape (389), millest esimene ületab kergelt kriitilise omapära, teine kaldub kõrvale ka standardiseeritud jäägi suhtes. Mõlemad on struktuurselt eripärased olles dihapped, mis ka tähendab võimet dissotseeruda (annavad lahusesse prootoni). Dissotseerumisest tuleneb ka töös kasutatud arvutusmeetodiga seotud kõrvalekalle, nimelt *EPI suite* tarkvarapakett arvestab log K_{OW} arvutamisel ainult dissotseerumata vormidega [11], mis toob sisse vea. Ühend indeno[1,2,3-cd]püreen (395) on polütsükliline kondenseerunud aromaatne ühend ja ainuke taoline selles andmekomplektis, mistõttu on ta omapära üle kriitilise piiri. Samuti on seda tüüpi ühendite puhul ennegi märgatud suuremat toksilisust [60].

3.3 Polaarne narkoos

Polaarse narkoosi (klass 2) korral koosnes andmeseeria 67 ühendist, mis jagunesid treening- ja valideerimisvalimisse vastavalt 50 ning 17 ühendi kaupa. Verhaar'i järgi kuuluvad teise klassi võrdlemisi inertsed kemikaalid, samas on nad toksilisemad, kui võiks eeldada ennustades seda vaid hüdrfoobsuse järgi. Struktuurides on domineerivaks lämmastikku ja hapnikku sisaldavad funktsionaalrühmad ja aromaatsed tuumad. Kirjanduses on välja toodud, et erinevus esimese

klassiga võib tulla klassi 2 kuuluvate ühendite võimest olla vesiniksideme doonoriks või akseptoriks [57,62,110].

Parima QSAR mudeli (Tabel 4, Graafik L2a) sai siin kuue deskriptoriga, pärast mida mudeli kirjeldusvõime järgnevate deskriptorite lisamisel enam ei paranenud, $\Delta R^2 < 0.02$.

Tabel 3. Kuue deskriptoriga QSAR polaarse narkoosi järgi käituvatele ühendite puhul: N = 50; n = 6; $R^2 = 0.7886$; $R^2_{CV} = 0.6398$; $F = 26.7385$; $s^2 = 0.1699$; $R^2_{val} = 0.6265$; $s^2_{val} = 1.4772$

Nr	*a*	s	*t*	Nimetus	Tähis
0	-5.51411	1.2159	-4.53499	Vabaliige (*Intercept*)	b
1	1.1836	0.159698	7.41144	Tsüklite arv (*Number of rings*)	N_{rings}
2	0.727378	0.101026	7.19987	Oktanool/vesi jaotuskoeffitsient (*Octanol-water patrition coefficient*)	log K_{OW}
3	17.6142	3.43407	5.12925	Vesiniksidemete aktseptoritest sõltuv vesiniksidemete doonorite pindala normaliseeritud kogumolekuli pindalaga (*HA dependent HDSA-1/TMSA (Zefirov PC)*)	$HDSA1^{HAdep}{}_{TMSA(Z)}$
4	-115.92	29.0372	-3.99212	Vesiniksideme doonorite suhteline positiivse osalaenguga pindala (versioon 2) (*H-donors FCPSA (version 2)*)	${}^{HD}FCPSA^{(2)}$
5	9.37945	2.37211	3.95406	Keskmine strukturaalne informatsoonisisaldus (nullindat järku) (*Average Structural Information content (order 0)*)	${}^{0}ASIC$
6	-0.516459	0.208248	-2.48002	Vesiniksideme aktseptorite pindalakaalutud pinna osalaeng (laengu järgi) (*HACA-2 (Zefirov PC) (all)*)	HACA2

Ka selles võrrandis on *t*-testi järgi üheks olulisemaks deskriptoriks log K_{OW} kui ühendite hüdrofoobsust kirjeldav deskriptor. Võrdväärselt on oluline aga ka tsüklite arv, mis seostub molekuli suurusega [30]. Tsüklid on bioloogilise aktiivsuse määramisel olulised struktuuri osad [30] ja deskriptoril on positiivne korralatsioon uuritava omadusega. Võib eeldada, et süsinikest koosnevate mahukate tsüklite tõttu on ühendil tendents jääda mittepolaarsesse faasi [111]. Järgnevalt aitavad võrrandi kirjeldusvõimet parandada kolm erinevat vesiniksideme moodustamise võimalusega seotud deskriptorit [2,30,87]. Vesiniksidemed on olulised, kuna on seotud käitumisega bioloogilistes süsteemides [30]. Vesiniksidemete olemasolu võib membraani läbimist nii takistada kui ka soodustada [102]. Siin on näha, et vesiniksideme doonorite suhtelise pindala korral (${}^{HD}FCPSA^{(2)}$) on negatiivne korrelatsioon, samuti nagu ka aktseptorite pindalakaalutud pinnalaengut näitaval deskriptoril (HACA2), samas on vesiniksidemete aktseptoritest sõltuva doonorite pindala ($HDSA1^{HAdep}{}_{TMSA(Z)}$) korral aga positiivne korrelatsioon. Selgelt eristub

tendents, et negatiivne korrelatsioon on just osalaenguga seotud deskriptoritel, mis näitab sellise molekuli, kus laeng on ebaühtlaselt jaotunud, tendentsi olla polaarses faasis, st veefaasis [30,75], mis omakorda tähendab väiksemat võimalust sattuda organismi ning seda mõjutada. Vajalik on tasakaal lipofiilsete omaduste ja vesiniksidemete moodustamise võime vahel [112]. Ligand-membraan/retseptor interaktsioonid sõltuvad mittekovalentsetest, st vesiniksidemetest [106]. Sõltuvuse kirjeldusvõimet aitab tõsta ka topoloogiline deskriptor, nullindat järku keskmine strukturaalne informatsioonisisaldus (^{0}ASIC), mis näitab molekuli struktuurset varieeruvust [29,30,113] ja mille positiivne korrelatsioon ütleb, et mida mitmekülgsema atomaarse koostisega on molekul, seda suurem on tema toksilisus. Seda võib interpreteerida kui heteroaatomite ja/või halogeenide esinemist molekulis, kusjuures heteroaatomid võivad anda võimaluse vesinik-sidemete tekkeks ja halogeenid suurendavad negatiivset osalaengut pinnal, viimase korrelatsiooni suurema toksilisusega on märgatud varemgi [23].

Vaadeldes statistikuid (R^2_{CV} = 0.6398; R^2_{val} = 0.6265; R^2 = 0.7886) võib ka teise klassi mudeli ennustusvõimega rahule jääda. Klassil 2 oli vaid 3 kõrvalekaldujat. Treeningvalimis oli kõrvalekaldujaks omapära järgi püridiin (284), kuid kuna punkt on praktiliselt kriitilise omapära joone kõrval, siis võib teda pidada mudeli osaks, kuid viitab täiendava kontrolli vajadusele püridiini analoogidele toksilisuse ennustamise korral. Püridiin on teises klassis üks väga väheseid heteroaromaatseid ühendeid. Valideerimisvalimis oli kaks kõrvalekaldujat nii omapära kui standardiseeritud jäägi järgi: 1,6-heksaandiamiin (353) ja 1,1,1-tris(4-hüdroksüfenüül)etaan (654) (Graafik L2b, Tabel L3). 1,6-heksaandiamiinil (353) on kaks amiinrühma, andes ühendile võimaluse rohketeks vesiniksidemeteks, sealhulgas omavahel, esineb ka nn amiinnarkoos [33]. Seevastu 1,1,1-tris(4-hüdroksüfenüül)etaan (654) on võrreldes teiste teise klassi ühenditega väga unikaalse struktuuriga, tegemist on ka ruumalalt kõige suurema teise klassi ühendiga, mis omakorda eristab ta muust valimist omapära järgi. Antud ühend on siin klassis ainus ülehinnatud toksilisusega ühend, mille tegelikku madalamat toksilisust võib seletada molekuli mõõtmetest ja hargnevusest tulenevate raskustega membraani läbimisel [106].

3.4 Mittespetsiifiliselt reageerivad kemikaalid

Mittespetsiifiliselt reageerivate kemikaalide klass [57] (klass 3) koosnes 44 ühendist, kus treeningvalimi moodustasid 41 ja valideerimisvalimi 13 ühendit. Kolmanda klassi moodustavad molekulid ei ole omavahel keemiliselt ja struktuurselt väga sarnased ning võib eeldada, et ka

nende interaktsioonid organismis *P. Subcapitata* on üsnagi erinevad. Sellest tuleneb ka asjaolu, et struktuure varieeruvuse kirjeldamiseks on vaja rohkem deskriptoreid saavutamaks olukorda, kus uute deskriptorite lisamine märgatavalt võrrandi ennustusvõimet ei paranda ($\Delta R^2 < 0.02$). Seetõttu on võrrandis suur arv deskriptoreid, tervenisti kümme (Tabel 5, Graafik L3a), kuid tegemist on antud metoodikat kasutades parima võimaliku tulemusega.

Tabel 4. Kümne deskriptoriga QSAR kolmandale klassile. N = 41; n = 10; R^2 = 0.8674; R^2_{CV} = 0.7241; F = 19.6259; s^2 = 0.2933; R^2_{val} = 0.3072 (0.1716); s^2_{val} = 18.5132 (20.4001)

Nr	*a*	s	*t*	Nimetus	Tähis
0	11.1869	2.61867	4.27199	Vabaliige (*Intercept*)	b
1	0.158624	0.0186575	8.50189	Sidemete infosisaldus (teist järku) (*Bonding Information content (order 2)*)	^{2}BIC
2	-8.9839	1.54407	-5.81833	Suhteline üksiksidemete arv (*Relative number of single bonds*)	N_{SBrel}
3	5.25915	0.992439	5.29922	Keskmine sidemete infosisaldus (nullindat järku) (*Average Bonding Information content (order 0)*)	^{0}ABIC
4	-16.2609	3.12366	-5.20571	Suhteline süsinikuaatomite arv (*Relative number of C atoms*)	N_{Crel}
5	-174.818	38.7127	-4.51578	Suhteline positiivse osalaenguga pindala (*FPSA3 Fractional PPSA (PPSA-3/TMSA) (Zefirov PC)*)	FPSA3
6	1.15788	0.282444	4.09949	Vesiniksideme aktseptorite pindalakaalutud pinnalaeng (laengu järgi) (*HACA-2 (Zefirov PC)*)	HACA2
7	-909.592	255.042	-3.56644	Vesiniksidemete aktseptoritest sõltuv vesiniksidemete doonorite pindalakaalutud osalaeng normaliseeritud kogumolekuli pindalaga (*HA dependent HDCA-2/TMSA (Zefirov PC)*)	$HDCA2^{HAdep}{}_{TMSA(Z)}$
8	0.0219193	0.0061557	3.56082	Vesiniksidemete doonorite pindala (versioon 2) (*H-donors PSA (version 2)*)	$^{HD}PSA^{(2)}$
9	-4.32091	1.25796	-3.43485	Polaarsusparameeter jagatud vahemaa ruuduga (*Polarity parameter / square distance (Zefirov)*)	P_f
10	-11.7381	4.19473	-2.79829	Minimaalne (kõige negatiivsem) osalaeng (*Min partial charge (Zefirov) for all atom types*)	Q_{min}

t-test järjestab võrrandi olulisemaks deskriptoriks teist järku sidemete informatsioonisisalduse indeksi (^{2}BIC), mis kirjeldab molekuli struktuurset varieeruvust lähtudes hargnevusest ja ka suurusest (nn komplekssust) [2,29,30]. Positiivne korrelatsioon

eksperimentaalse väärtusega näitab, et mida varieeruvamad on ühendid koostise ja hargnevuse poolest, seda toksilisemad nad on. Teine informatsioonisisalduse deskriptor, nullindat järku keskmine sidemete infosisaldus (^{0}ABIC), toob esile struktuurse varieeruvuse lähtudes atomaarsest koostisest [30] ja selle positiivne korrelatsiooni näitab samuti, et mida mitmekesisema atomaarse ülesehitusega on ühend, seda toksilisem ta on. Teine deskriptor võrrandis on suhteline üksiksidemete arv (N_{SBrel}), mis samuti näitab struktuurset varieeruvust [30] lähtudes sidemete mustrist skaalal aromaatsetest või konjugeeritutest puhtalt alifaatsete ühenditeni. Negatiivsest korrelatsioonist omadusega võib järeldada, et mida rohkem on kordseid sidemeid, seda toksilisem on ühend. Vähem alifaatsed ühendid on enamasti reaktsiooni-võimelisemad [99]. Neljas deskriptor võrrandis, suhteline süsinikuaatomite arv (N_{Crel}) [2], võtab arvesse struktuurset varieeruvust atomaarse koostise tasemel, võrreldes omavahel süsiniku aatomite ja kõigi molekulis olevate aatomite arvu. Deskriptori negatiivne korrelatsioon näitab, et mida rohkem on molekulis aatomeid, mis pole süsinikud, seda toksilisem ta on. Võrrandis on ka kolm vesiniksidemete moodustamise võimalust kirjeldavat deskriptorit (HACA2, $HDCA2^{HAdep}{}_{TMSA(Z)}$, $^{HD}PSA^{(2)}$) [2,30,87]. Neist teine omab väärtust vaid siis, kui molekulis leidub nii vesiniksidemete doonorit kui ka aktseptorit. Vesiniksidemeid kirjeldavad deskriptorid on ka käesolevas võrrandis olulised, kuna mõjutavad elektronide jaotust molekulis ning läbi selle reaktsioonivõimet, veel on nad seotud molekuli käitumisega bioloogilistes süsteemides [30]. $HDCA2^{HAdep}{}_{TMSA(Z)}$ negatiivne korrelatsioon võib tulla väga arvukatest vesiniksidemetest nii omavahel kui ka polaarse faasi (vee) molekulidega, mis takistavad organismi jõudmist [94]; on teada, et vesiniksidemete olemasolu võib membraani läbimist nii takistada kui ka soodustada [102]. Lipofiilsete omaduste ja vesiniksidemete moodustamise võime vahel peab olema tasakaal [112]. Teised kaks vesiniksidemete deskriptorit kirjeldavad aktseptor (HACA2) või doonor ($^{HD}PSA^{(2)}$) omadusi ning nende positiivne korrelatsioon näitab, et mida suuremad on vastavad pindalad, seda reaktsiooni- või interaktsioonivõimelisemad ning toksilisemad on need ühendid. Bioloogilise aktiivsuse avaldumiseks on vajalik tasakaal hüdrofiilsete ja hüdrofoobsete omaduste vahel, seda ka vesiniksidemete korral [114]. Võrrandi kirjeldusvõimet aitavad suurendada veel laengujaotusega seotud deskriptorid minimaalne osalaeng (Q_{min}) ja suhteline positiivse osalaenguga pindala (FPSA3) [2]. Negatiivset korrelatsiooni mõlemal võib jällegi seletada vastava osalaenguga molekuli võimega olla veefaasis, nimelt negatiivse osalaenguga osake tõugatakse negatiivse osalaenguga rakumembraani juurest ära tagasi veefaasi, samas liiga tugeva

positiivse laenguga osake aga seostub fosfolipiidi negatiivse osaga ning seetõttu ei sisene rakku [114]. Sisse tulnud polaarsusparameeter (P_f) [87] kirjeldab kõige negatiivsema ja kõige positiivsema osalaengu erinevust normaliseerituna nende vahelise kaugusega. Antud juhul on tegemist negatiivse korrelatsiooniga ja see näitab, et väikesed ühendid, mille laeng paikneb kompaktselt, on toksilisemad. Seda on võimalik seletada samuti polaarsemate ühendite suurema jaotumisega veefaasi [30,75].

Kõrvalekaldujaid (graafik L3b) oli siin vaid üks, nii omapära kui ka standardiseeritud jäägi suhtes kõrvalekalduv valideerimisvalimisse kuuluv tris(2-hüdroksüetüül)isotsüanuraat (666), mis on molekulaarpindalalt klassi suurim ühend ning omab võrreldes teistega väga unikaalset struktuuri; väga palju on vesiniksidemete moodustamise asukohti ning reaktsioonivõimelisi rühmi. Tema toksilisust on tugevalt ülehinnatud, arvatavasti just seetõttu, et reaktsioonivõimelised ja vesiniksidemete moodustamise kohad on olemas, aga molekuli suured mõõtmed takistavad tegelikkuses membraani läbimist [106].

Valideerimisvalimisse kuuluv tris(2-hüdroksüetüül)isotsüanuraat (666) mõjutab märkimisväärselt ka valideerimise tulemust (R^2_{val} = 0.1716, kogu valideerimisvalimile). See ühend otsustati valideerimisvalimist eemaldada, mille tulemusena korrelatsioonikoefitsiendi ruut küll paranes (R^2_{val} = 0.3072), kuid näitab endiselt valimi suurt hajuvust. Siiski tuleb ära märkida, et kõik valideerimisvalimi ühenditega (va 666) tehtud ennustused jäävad mudeli rakendus-piirkonda (Graafik L3b).

3.5 Spetsiifiliselt reageerivad kemikaalid

Spetsiifiliselt reageerivaid kemikaale (klass 4) oli vaid 8 ning seetõttu neid eraldi treening- ja valideerimisvalimiteks ei jagatud, mistõttu ei olnud siin võimalik ka välise valideerimise sooritamine. Klassi kuuluvad ühendid, millel on teada spetsiifiline reaktsioonimehhanism [57]. Ka selle väikese valimi korral moodustati QSAR, kuid ühendite vähese hulga tõttu valiti välja olenemata statistikutest parim kahe deskriptoriga sõltuvus (Tabel 6).

Tabel 5. Kahe deskriptoriga QSAR mudel neljandale klassile. N = 8; n = 2; R^2 = 0.907; R^2_{CV} = 0.7478; F = 24.392; s^2 = 0.101

Nr	*a*	s	*t*	Nimetus	Tähis
0	5.53705	0.766613	7.22274	Vabaliige (*Intercept*)	b
1	0.0466432	0.00834544	5.58906	Vesiniksidemete aktseptorite pindala (*HASA-1 (Zefirov PC)*)	HASA1
2	-15.7211	2.99561	-5.24804	Suhteline negatiivne laeng (*RNCG Relative negative charge (QMNEG/QTMINUS) (Zefirov PC)*)	RNCG

t-test järjestab esimeseks deskriptoriks positiivse korrelatsiooniga vesiniksidemete aktseptorite pindala (HASA1) [2], ehk mida suurem nende pindala, seda suurem toksilisus. Vesiniksidemete moodustamise võime on tihti seotud ensüüm-substraat interaktsioonidega [99]. Väga sarnase *t*-testi absoluutväärtusega on ka teine deskriptor, suhteline negatiivne osalaeng (RNCG) [30], mille negatiivne korrelatsioon ütleb, et mida rohkem on negatiivne osalaeng kogunenud ühele aatomile molekulis, seda vähem toksiline ta on. See näitab, et molekuli polaarsuse suurenedes toksiline efekt väheneb, kuna selline molekul kipub jääma veefaasi [75].

Neljandas klassis ei esinenud kõrvalekaldujaid (Graafik L4b). Eksperimentaalseid ja arvutuslikke väärtusi on võrreldud graafikul L4a, kus on visuaalsel vaatlusel näha head korrelatsiooni arvutatud ja eksperimentaalsete väärtuste vahel. Mudelil on ka hea ennustusvõime (R^2_{CV} = 0.7478).

3.6 Mitteklassifitseeritavad kemikaalid

Märkimisväärset osa kogu andmekomplektis olevatest kemikaalidest ei saa modifitseeritud Verhaar'i reeglite järgi eelnevatesse klassidesse jagada, moodustades viienda klassi [57], kus kokku oli 133 ühendit: treeningvalimis 100 ja valideerimisvalimis 33. Kuna tegemist on kõige heterogeensema valimiga ja ilmselt ka kõige hajuvamaga, siis teeb see viienda klassi jaoks tuletatava võrrandi deskriptorite interpreteerimise eriti keerukaks, kuna andmeseeria koondab endas palju erinevate toksilisuse mehhanismidega ühendeid. Kriteeriumi $\Delta R^2 < 0.02$ järgi saadi kaheteistkümne deskriptoriga võrrand (Tabel 7, Graafik L5a), mis on küll väga palju [1,90], kuid siiski on tegemist statistikuid (Tabel 7) arvestades parima võimaliku mudeliga.

Tabel 6. Kaheteistkümne deskriptoriga QSAR mudel viiendale klassile: N = 100; n = 12; R^2 = 0.6445; R^2_{CV} = 0.5279; F = 13.1454; s^2 = 0.7562; R^2_{val} = 0.7543; s^2_{val} = 1.3214

Nr	*a*	s	*t*	Nimetus	Tähis
0	1.87135	1.02059	1.8336	Vabaliige (*Intercept*)	b
1	0.0298803	0.00346409	8.62571	Molekuli pindala (*Molecular surface area*)	S_{mol}
2	-20.1824	3.26268	-6.18582	Suhteline N aatomite arv (*Relative number of N atoms*)	N_{Nrel}
3	-10.4061	2.05583	-5.06176	Suhteline C aatomite arv (*Relative number of C atoms*)	N_{Crel}
4	-0.148023	0.0304607	-4.85948	H aatomite arv (*Number of H atoms*)	N_H
5	2.99273	0.617347	4.84772	Vesiniksidemete aktseptoritest sõltuv vesiniksidemete doonorite pindala normaliseeritud kogumolekuli pindala ruutjuurega (*HA dependent HDSA-2/SQRT(TMSA) (Zefirov PC)*)	$HDSA2^{HAdep}{}_{sqrt(TMSA)(z)}$
6	-2.84219	0.605355	-4.69508	Keskmine informatsioonihulk (esimest järku) (*Average Information content (order 1)*)	^1AIC
7	10.0686	2.57998	3.9026	Keskmine struktuurne informatsoonihulk (esimest järku) (*Average Structural Information content (order 1)*)	^1ASIC
8	0.517141	0.156815	3.29778	Molekuli kogu elektrostaatilise interaktsiooni energia normaliseeritud aatomite arvuga (*Tot molecular electrostatic interaction / # of atoms*)	$E_{C(tot)}/N_{at}$
9	-1.56606	0.513232	-3.05138	Vesiniksidemete aktseptoritest sõltuv vesiniksidemete doonorite osalaenguga pindala (*HA dependent HDCA-2 (Zefirov PC) (all)*)	HDCA2
10	-0.0565208	0.0192484	-2.93638	Pindalakaalutud negatiivse osalaenguga pindala (*RNCS Relative negative charged SA (SAMNEG*RNCG) (Zefirov PC)*)	RNCS
11	0.116645	0.0435873	2.67611	Lõplik tekkesoojus jagatud aatomite arvuga (*Final heat of formation / # atoms*)	$\Delta H^0_f/N_{at}$
12	-0.234104	0.101351	-2.30982	Hapnikuaatomite arv (*Number of O atoms*)	N_O

Deskriptoritest olulisim on molekuli pindala (S_{mol}), millel on positiivne korrelatsioon. Seda saaks võtta kui suurema molekuli suuremat hüdrofoobsust [111]. Oluline osa varieeruva mehhanistliku ülesehitusega andmekomplekti korral on konstitutsioonilistel deskriptoritel, kuna neljal korral tulid deskriptoritena sisse kindlate aatomite esinemissagedused, nimelt vesinikuaatomite (N_H), suhtelised süsiniku- ja lämmastikuaatomite arvud (N_{Crel}, N_{Nrel}) ning hapnikuaatomite arv (N_O), millel kõigil on negatiivne korrelatsioon. Süsinike suhtelise arvu ja vesinike arvu puhul võib seda võtta kui reaktsioonivõimeliste rühmade puudumist või suurema molekuli raskusi membraani läbimisel, lämmastiku suhtelist arvu ja hapnikuaatomite arvu kui polaarsuse suurenemist ja sealt tendentsi jääda veefaasi [30,75]. Keskmist informatsioonisisaldust (^{1}AIC) ja keskmist struktuurset informatsioonisisaldus (^{1}ASIC) saab vaadelda molekuli komplekssuse näitajatena ja mõlemal juhul on juures ka hargnevuse komponent esimesest aatomist lähtuva sideme tasemel (nn. esimene koordinatsioonisfäär) [2,30]. Veel esineb ka siin mudelis vesiniksidemetega seotud deskriptoreid (HDCA2 ja $HDSA2^{HAdep}_{sqrt(TMSA)(Z)}$) [2,30,87]. Vesiniksidemed on seotud molekuli käitumisega bioloogilistes süsteemides [30] ning nende olemasolu võib membraani läbimist nii takistada kui ka soodustada [102]. Bioloogilise aktiivsuse avaldumiseks on vajalik tasakaal hüdrofiilsete ja hüdrofoobsete omaduste vahel, vesiniksidemete olemasolu tavaliselt suurendab tendentsi jääda veefaasi [36,114], seda on praegu näha osalaengut arvesse võtva deskriptori HDCA2 negatiivsel korrelatsioonil. Positiivset korrelatsiooni teisel vesiniksidemete deskriptoril võib seletada sellega, et doonorrühmade osalaengutele vastavate pindalade suurenemise korral võrreldes kogu molekuli pindalaga suureneb ka toksilisus, kuna siis võib reaktsioonivõimeline vesiniksidemeid moodustav rühm moodustada märkimisväärse osa molekuli pindalast. Ligand-membraan/retseptor interaktsioonid sõltuvad mittekovalentsetest, st vesiniksidemetest [106]. Negatiivset korrelatsiooni on näha osalaenguga seotud deskriptoril (RNCS) [30], mis näitab ebaühtlase laengujaotusega, st polaarse osakese tendentsi olla veefaasis. Molekuli kogu elektrostaatiline interaktsioonienergia normaliseerituna aatomite arvuga ($E_{C(tot)}/N_{at}$) [2] iseloomustab aatomite ja elektronide vahelisi tõmbumisi ja tõukumisi moleklis, tekkesoojus normaliseeritud aatomite arvuga ($\Delta H^0_f/N_{at}$) iseloomustab molekuli stabiilsust, positiivne korrelatsioon näitab vähemstabiilsete ühendite suuremat toksilisust [23].

Viiendas klassis oli neli kõrvalekaldujat (graafik L5b, eksperimentaalseid ja arvutuslikke väärtusi võrreldud graafikus L5a). Treeningvalimis kaldus standardse jäägi suhtes kõrvale orto-atsetoatsetotoluidiid (132) ning omapära suhtes trikloronitrometaan (44). Valideerimisvalimis oli

kaks omapära suhtes kõrvalekaldujat, meta-ksüleenheksafluoriid (410) ja etüültrikloroetanaat (424). Trikloronitrometaani (44) puhul on tegemist väga väikese molekuliga. Ka on ta üks vähestest viiendas klassis, millel esineb nitrorühma. Orto-atsetoatsetotoluidiid (132) ei ole võrreldes teiste viienda klassi ühenditega unikaalne, kõrvalekalle võib tuleneda mõõtmisveast. Meta-ksüleenheksafluoriid (410) on üks väga vähestest fluori sisaldavatest ühenditest andmekomplektis ning etüültrikloroetanaat (424) sisaldab klooriaatomeid, mis on viienda klassi puhul harv, tehes nad andmekomplekti piires unikaalseks.

4. Kokkuvõte

Kemikaalide toksilisus vetikatele on ökoloogia seisukohast oluline vetikate tähtsuse tõttu ökosüsteemis. Juba mõnda aega on tegeletud QSAR mudelite tuletamisega, mis modelleeriks ühendite toksilisust vetikatele ning suudaks seda piisava täpsusega tundmatute ühendite puhul ennustada. *In silico* meetodid, milleks on ka QSAR, säästavad aega ja ressursse ning on eetika seisukohtadest vähem probleeme tekitavad, kui *in vivo* ja *in vitro* katsed.

Käesolevas töös modelleeriti 342 keemiliselt (struktuurselt) erineva aine toksilisust vetikale *Pseudokirchneriella subcapitata* kasutades BMLR meetodit. Kokku tuletati 5 mudelit, igale Verhaar'i klassile (modifitseeritud reeglite järgi) eraldi. Verhaar'i klassid määrasid ka mudelite rakenduspiirkonnad. Modelleeritavaks parameetriks oli log ($1/EC_{50}$). Ennustusvõime hindamiseks sooritati väline valideerimine jagades ühendid treening- ja valideerimisvalimiteks. Kõrvalekaldujaid ja struktuurselt eripäraseid ühendeid määrati ja analüüsiti Williams'i graafikutega.

Esimese klassi korral saadi heade statistikutega ja korraliku ennustusvõimega mudel ($N = 60$; $n = 4$; $R^2 = 0.8145$; $R^2_{CV} = 0.7619$; $F = 60.3746$; $s^2 = 0.2495$; $R^2_{val} = 0.7381$; $s^2_{val} = 0.9832$). Suurema osa andmeseeria struktuursest varieeruvusest kirjeldas oktanool-vesi jaotuskoeffitsiendi logaritm (log K_{OW}). Seda deskriptorit täiendasid keskmine aatommass, polariseeritavuse ja vesiniksidemete deskriptorid. Kõik võrrandi liikmed omasid positiivset korrelatsioon. Rakenduspiirkonna analüüsimisel leiti mudelis 12 kõrvalekaldujat, enamuses tagasihoidlikud, ja mille täpsem analüüs tõi esile nende struktuuride eripärad, mida tuleb arvestada mudeli kasutamisel ennustamiseks.

Teise klassi korral oli vaja kuute deskriptorit hea korrelatsiooni ja ennustusvõimega mudeli saamiseks ($N = 50$; $n = 6$; $R^2 = 0.7886$; $R^2_{CV} = 0.6398$; $F = 26.7385$; $s^2 = 0.1699$; $R^2_{val} = 0.6265$; $s^2_{val} = 1.4772$). Deskriptoritena osutusid oluliseks positiivsete korrelatsioonidega log K_{OW} ja tsüklite arv. Ülejäänud deskriptorid arvestasid struktuurset varieeruvust ning vesiniksidemete olulisust polaarse narkoosi järgi käituvate ühendite toksilisuse kirjeldamisel. Mudeli rakenduspiirkonna analüüs tõi esile vaid 3 kõrvalekaldujat, mis näitab võrrandi head stabiilsust.

Kuna võimalike toksiliste mehhanismide hulk järgnevates andmekomplektides (klassides) suureneb, kasvab nendes ka struktuurse varieeruvuse kirjeldamiseks vajalik parameetrite arv, mis teeb interpretatsiooni keerulisemaks. Nii näiteks saadi mittespetsiifiliselt reageerivate kemikaalide (klass 3) korral sobiv korrelatsioonivõrrand kümne parameetriga ($N = 41$; $n = 10$; R^2

= 0.8674; R^2_{CV} = 0.7241; F = 19.6259; s^2 = 0.2933; R^2_{val} = 0.3072 (0.1716); s^2_{val} = 18.5132 (20.4001)). Andmekomplektis olev struktuurne varieeruvus kirjeldati ära kahe topoloogilise, kahe konstitutsioonilise, kolme vesiniksidemete ning kolme laengut ja laengujaotust kirjeldava deskriptoriga. Problemaatiliseks osutus võrrandi valideerimine, kuna valideerimisvalimis sisaldus üks struktuurselt väga eripärane ühend, mis oli ka ainsaks kõrvalekaldujaks, samas on kõik teised valideerimisvalimi ühendid ennustatud mudeli rakenduspiirkonda.

Viienda, arvatavasti kõige heterogeensema ja hajuvama klassi korral saadi samuti hea mudel, kuid kaheteistkümne deskriptoriga (N = 100; n = 12; R2 = 0.6445; R2cv = 0.5279; F = 13.1454; s^2 = 0.7562; R^2_{val} = 0.7543; s^2_{val} = 1.3214). Oluline on siin märkida, et klass 5 koondab ühendid, mis teistesse klassidesse ei liigitu ja seega on tegemist mehhanistlikus mõistes väga heterogeense andmeseeriaga. See tegi ka võrrandi sügavama analüüsi raskeks, just tänu deskriptorite hulgale: neli konstitutsioonilist, kaks vesiniksidemete, kolm geomeetrilist, kaks laengujaotusega seotud ja kaks kvantkeemilist deskriptorit. Samas on välise valideerimise tulemused head ja rakenduspiikonna analüüs tuvastas vaid neli kõrvalekaldujat, mille analoogidele tuleb mudeli kasutamisel ennustamiseks tähelepanu pöörata.

Spetsiifiliselt reageerivad kemikaalid (klass 4) andsid tänu andmeseeria väiksusele väga heade statistikutega mudeli (N = 8; n = 2; R^2 = 0.907; R^2_{CV} = 0.7478; F = 24.392; s^2 = 0.101). Deskriptoriteks olid laengujaotust ja vesiniksidemeid kirjeldavad parameetrid. Ühendite vähesuse tõttu ei saanud aga seal sooritada välist valideerimist, samas rakenduspiirkonna analüüsil ei leitud ühtegi kõrvalekaldujat.

5. Summary

Toxicity of chemicals towards algae is important from an ecological point of view due to the importance of algae in the ecosystem. For some time QSAR models have been derived that would model toxicity towards algae and could predict it accurately enough for unfamiliar compounds. *In silico* methods, QSAR being one of them, save time and resources and create less ethical problems than *in vivo* and *in vitro* experiments.

In this work, toxicity of 342 chemically (structurally) varying compounds towards algae *Pseudokirchneriella subcapitata* was modelled using the BMLR method. A total of 5 models were derived, one for each Verhaar class (modified rules). Applicability domains were also determined by the Verhaar classes. The parameter for modelling was log ($1/EC_{50}$). To evaluate quality of predictions, external validation was performed by dividing the compounds into training and validation sets. Outliers and structurally unique compounds were determined and analysed with the help of Williams plots.

A model with good statistical parameters and decent predictive power was derived for class 1 ($N = 60$; $n = 4$; $R^2 = 0.8145$; $R^2_{CV} = 0.7619$; $F = 60.3746$; $s^2 = 0.2495$; $R^2_{val} = 0.7381$; $s^2_{val} = 0.9832$). For the most part, structural variation in the series was described by the octanol-water partition coefficient (log K_{OW}). Average atomic weight, polarizability and hydrogen bond descriptors also contributed to the model. All the descriptors had a positive correlation. By analysing the applicability domain, 12 mostly moderate outliers were found, more specific analysis of which made apparent their structural uniqueness, which one must consider when using the model for prediction.

For class 2, six descriptors were needed to get a good correlation and predictive power ($N = 50$; $n = 6$; $R^2 = 0.7886$; $R^2_{CV} = 0.6398$; $F = 26.7385$; $s^2 = 0.1699$; $R^2_{val} = 0.6265$; $s^2_{val} = 1.4772$). The most significant descriptors here were log K_{OW} and the number of cycles with a positive correlation. The rest of the descriptors took into account structural variation and the importance of hydrogen bonding when describing the toxicity of compounds acting by polar narcosis. By analysing the applicability domain, only 3 outliers were observed, which shows the good stability of the model.

The number of possible mechanisms of toxicity in the following classes increases, therefore the number of equation parameters needed to describe the structural variability of the data sets also increases. Fox example, for non-specifically reactive chemicals, a good correlation

with ten parameters was found (N = 41; n = 10; R^2 = 0.8674; R^2_{CV} = 0.7241; F = 19.6259; s^2 = 0.2933; R^2_{val} = 0.3072 (0.1716); s^2_{val} = 18.5132 (20.4001)). The structural variation in this series was described by two topological, two constitutional, three hydrogen bonding and three charge and charge distribution-related descriptors. Validation of the model was problematic due to one structurally very unique compound appearing in the validation set, which was also the only outlier, at the same time, all the other validation set compounds are predicted to be in the applicability domain.

For class 5, which most likely is the most heterogenous and divergent class, a good model was derived, although 12 descriptors were necessary (N = 100; n = 12; R2 = 0.6445; R2cv = 0.5279; F = 13.1454; s^2 = 0.7562; R^2_{val} = 0.7543; s^2_{val} = 1.3214). It must be noted that class 5 includes in itself compounds that cannot be classified into other classes, thus, mechanistically, it is a very heterogeneous series. This also made it difficult to analyse the model more specifically, mostly due to the large number of descriptors: four constitutional, two hydrogen bonding, three geometrical, two charge distribution and two quantum chemical descriptors appeared in the model. There were only 4 outliers.

Specifically reactive chemicals (class 4) gave a model with very good statistical parameters due to the small size of the series (N = 8; n = 2; R^2 = 0.907; R^2_{CV} = 0.7478; F = 24.392; s^2 = 0.101). Charge distribution and hydrogen bonding descriptors appeared in the model. Due to very few compounds in this class, it was not possible to perform an external validation. Still, the applicability domain was analysed and no outliers were found.

Viited

[1] K. Roy, S. Kar, R. N. Das, Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment, Elsevier Academic Press, Amsterdam, 2015.

[2] M. Karelson, Molecular Descriptors in QSAR/QSPR, 2000.

[3] V. Aruoja, Algae Pseudokircheriella subcapitata in environmental hazard evaluation of chemicals and synthetic nanoparticles, Estonian University of Life Sciences, 2011.

[4] J. T. Wootton, Effects of disturbance on species diversity: a multitrophic perspective, Am. Nat. 152 (1998) 803–825.

[5] S. W. Geis, K. L. Fleming, E. T. Korthals, G. Searle, L. Reynolds, D. A. Karner, Modifications to the algal growth inhibition test for use as a regulatory assay, Environ. Toxicol. Chem. 19 (2000) 36–41.

[6] R. G. Wetzel, Limnology: Lake and River Ecosystems, 3rd ed., Academic Press, Elsevier Science, San Diego, California, 2001.

[7] R. L. Chapman, Algae: the world's most important "plants"—an introduction, Mitig. Adapt. Strateg. Glob. Change. 18 (2013) 5–12.

[8] L. Fu, J. J. Li, Y. Wang, X. H. Wang, Y. Wen, W. C. Qin, L. M. Su, Y. H. Zhao, Evaluation of toxicity data to green algae and relationship with hydrophobicity, Chemosphere. 120 (2015) 16–22.

[9] S. Schäfer, G. Buchmeier, E. Claus, L. Duester, P. Heininger, A. Körner, P. Mayer, A. Paschke, C. Rauert, G. Reifferscheid, H. Rüdel, C. Schlechtriem, C. Schröter-Kermani, D. Schudoma, F. Smedes, D. Steffen, F. Vietoris, Bioaccumulation in aquatic systems: methodological approaches, monitoring and assessment, Environ. Sci. Eur. 27 (2015).

[10] S.-H. Hsieh, C.-H. Hsu, D.-Y. Tsai, C.-Y. Chen, Quantitative structure-activity relationships for toxicity of nonpolar narcotic chemicals to Pseudokirchneriella subcapitata, Environ. Toxicol. Chem. 25 (2006) 2920–2926.

[11] X. H. Wang, Y. Yu, L. Fu, H. W. Tai, W. C. Qin, L. M. Su, Y. H. Zhao, Comparison of Chemical Toxicity to Different Algal Species Based on Interspecies Correlation, Species Sensitivity, and Excess Toxicity: Water, CLEAN - Soil Air Water. 44 (2016) 803–808.

[12] K.-P. Tsai, C.-Y. Chen, An algal toxicity database of organic toxicants derived by a closed-system technique, Environ. Toxicol. Chem. 26 (2007) 1931–1939.

[13] C.-P. Huang, Y.-J. Wang, C.-Y. Chen, Toxicity and quantitative structure–activity relationships of nitriles based on Pseudokirchneriella subcapitata, Ecotoxicol. Environ. Saf. 67 (2007) 439–446.

[14] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR Modeling: Where Have You Been? Where Are You Going To?, J. Med. Chem. 57 (2014) 4977–5010.

[15] M. T. D. Cronin, J. C. Madden, In Silico Toxicology- An Introduction, in: M. T .D. Cronin, J. C. Madden (Eds.), Silico Toxicol. Princ. Appl., The Royal Society of Chemistry, 2010.

[16] A. Grover, M. Grover, K. Sharma, A Practical Overview of Quantitative Structure-Activity Relationship, (2015).

[17] C. J. Cramer, Essentials of computational chemistry: theories and models, 2nd ed., Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2004.

[18] J. Stevenson, Ecological assessments with algae: a review and synthesis, J. Phycol. 50 (2014) 437–461.

[19] E. E. Kenaga, R.J. Moolenaar, Fish and Daphnia toxicity as surrogates for aquatic vascular plants and algae, Environ. Sci. Technol. 13 (1979) 1479–1480.

[20] Pseudokirchneriella subcapitata (Korshikov) F.Hindák :: Algaebase. http://www.algaebase.org/search/species/detail/?species_id=47100 viimati alla laetud 22.05.2018.

[21] WoRMS - World Register of Marine Species - Raphidocelis subcapitata (Korshikov) Nygaard, Komárek, J.Kristiansen & O.M.Skulberg, 1987. http://www.marinespecies.org/aphia.php?p=taxdetails&id=610914 viimati uuendatud 26.06.2015.

[22] G. Nygaard, J. Komárek, J. Kristiansen, O. M. Skulberg, Taxonomic designations of the biossay alga NIVA-CHL 1 (Selenastrum capricornutum) and some related strains, Opera Bot. 90 (1986) 1–46.

[23] V. Aruoja, M. Moosus, A. Kahru, M. Sihtmäe, U. Maran, Measurement of baseline toxicity and QSAR analysis of 50 non-polar and 58 polar narcotic chemicals for the alga Pseudokirchneriella subcapitata, Chemosphere. 96 (2014) 23–32.

[24] T. Yamagishi, H. Yamaguchi, S. Suzuki, Y. Horie, N. Tatarazako, Cell reproductive patterns in the green alga Pseudokirchneriella subcapitata (= Selenastrum capricornutum) and their variations under exposure to the typical toxicants potassium dichromate and 3, 5-DCP, PloS One. 12 (2017) e0171259.

[25] OECD, OECD Guidelines for the Testing of Chemicals, Freshwater Alga and Cyanobacteria, Growth Inhibition Test, (2011). https://www.oecd-ilibrary.org/environment/test-no-201-alga-growth-inhibition-test_9789264069923-en viimati uuendatud 28.07.2011.

[26] M. T. D. Cronin, T.W. Schultz, Pitfalls in QSAR, J. Mol. Struct. THEOCHEM. 622 (2003) 39–51.

[27] N. Nyholm, T. Källqvist, Methods for growth inhibition toxicity tests with freshwater algae, Environ. Toxicol. Chem. 8 (1989) 689–703.

[28] V. Aruoja, M. Sihtmäe, H.-C. Dubourguier, A. Kahru, Toxicity of 58 substituted anilines and phenols to algae Pseudokirchneriella subcapitata and bacteria Vibrio fischeri: Comparison with published data and QSARs, Chemosphere. 84 (2011) 1310–1320.

[29] U. Maran, S. Sild, I. Tulp, K. Takkis, M. Moosus, Molecular Descriptors from Two-Dimensional Chemical Structure, in: M.T.D. Cronin, J.C. Madden (Eds.), Silico Toxicol. Princ. Appl., The Royal Society of Chemistry, 2010.

[30] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing, 2nd ed., Wiley-VHC, 2009.

[31] P. Liu, W. Long, Current Mathematical Methods Used in QSAR/QSPR Studies, Int. J. Mol. Sci. 10 (2009) 1978–1998.

[32] H. Kubinyi, QSAR: Hansch analysis and related approaches, VCH, Weinheim ; New York, 1993.

[33] A. Levet, C. Bordes, Y. Clément, P. Mignon, C. Morell, H. Chermette, P. Marote, P. Lantéri, Acute aquatic toxicity of organic solvents modeled by QSARs, J. Mol. Model. 22 (2016).

[34] EC, Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC, Off. J. Eur. Union L. 396 (2006).

[35] C. Rovida, T. Hartung, Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals - a report by the transatlantic think tank for toxicology (t(4))., ALTEX. 26 (2009) 187–208.
[36] N. Basant, S. Gupta, K. P. Singh, Modeling the toxicity of chemical pesticides in multiple test species using local and global QSTR approaches, Toxicol Res. 5 (2016) 340–353.
[37] A. R. Leach, Molecular Modelling, Principles and Applications, 2nd ed., Pearson Education Limited, Essex, England, UK, 2001.
[38] C. Nantasenamat, C. Isarankura-Na-Ayudhya, T. Naenna, V. Prachayasittikul, A practical overview of quantitative structure-activity relationship, (2009).
[39] P. Gramatica, S. Cassani, P. P. Roy, S. Kovarich, C. W. Yap, E. Papa, QSAR Modeling is not "Push a Button and Find a Correlation": A Case Study of Toxicity of (Benzo-)triazoles on Algae, Mol. Inform. 31 (2012) 817–835.
[40] OECD Principles for the Validation, for Regulatory Purposes. of (Quantitative) Structure-Activity Relationship Models, (2007). http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf.
[41] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701.
[42] M. Stone, P. Jonathan, Statistical thinking and technique for QSAR and related studies. Part I: General theory, J. Chemom. 7 (1993) 455–475.
[43] M. Stone, P. Jonathan, Statistical thinking and technique for QSAR and related studies. Part II: Specific methods, J. Chemom. 8 (1994) 1–20.
[44] A. C. Brown, T. R. Fraser, On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia, J. Anat. Physiol. 2 (1868) 224–242.
[45] H. Meyer, Zur Theorie der Alkoholnarkose. Erste Mitteilung. Welche Eigenschaft der Anästhetica bedingt ihre narkotische Wirkung?, Naunyn. Schmiedebergs Arch. Pharmacol. 42 (1899) 109–118.
[46] C. E. Overton, Studies of Narcosis, 1991.
[47] J. C. Dearden, The History and Development of Quantitative Structure-Activity Relationships (QSARs), Int. J. Quant. Struct.-Prop. Relatsh. 1 (2016) 1–46.

[48] L. P. Hammett, The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives, J Am Chem Soc. 59 (1937) 96–103.
[49] C. D. Selassie, History of Quantitative Structure-Activity Relationships, in: D.J. Abraham (Ed.), Burger's Med. Chem. Drug Discov., 6th ed., Wiley, Hoboken, N.J, 2003: pp. 1–48.
[50] R. W. Taft, Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-substituted Benzoate Esters, J. Am. Chem. Soc. 74 (1952) 2729–2732.
[51] R. W. Taft, Polar and Steric Substituent Constants for Aliphatic and o-Benzoate Groups from Rates of Esterification and Hydrolysis of Esters, J. Am. Chem. Soc. 74 (1952) 3120–3128.
[52] R. W. Taft, Linear Steric Energy Relationships, J. Am. Chem. Soc. 75 (1953) 4538–4539.
[53] C. Hansch, P.P. Maloney, R.M. Muir, T. Fujita, Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients, Nature. 194 (1962) 178.
[54] L. F. Fieser, A.P. Richardson, Naphthoquinone antimalarials; correlation of structure and activity against P. lophurae in ducks, J. Am. Chem. Soc. 70 (1948) 3156–3165.
[55] C. Hansch, T. Fujita, p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure, J. Am. Chem. Soc. 86 (1964) 1616–1626.
[56] P. Y. Lee, C.Y. Chen, Toxicity and quantitative structure–activity relationships of benzoic acids to Pseudokirchneriella subcapitata, J. Hazard. Mater. 165 (2009) 156–161.
[57] H. J. Verhaar, C.J. Van Leeuwen, J.L. Hermens, Classifying environmental pollutants, Chemosphere. 25 (1992) 471–491.
[58] Y. G. Lee, S.H. Hwang, S.D. Kim, Predicting the Toxicity of Substituted Phenols to Aquatic Species and Its Changes in the Stream and Effluent Waters, Arch. Environ. Contam. Toxicol. 50 (2006) 213–219. doi:10.1007/s00244-004-1259-6.
[59] C.-Y. Chen, J.-H. Lin, Toxicity of chlorophenols to Pseudokirchneriella subcapitata under air-tight test environment, Chemosphere. 62 (2006) 503–509.
[60] S. Pramanik, K. Roy, Predictive modeling of chemical toxicity towards Pseudokirchneriella subcapitata using regression and classification based approaches, Ecotoxicol. Environ. Saf. 101 (2014) 184–190.
[61] Ministry of the Environment of the Government of Japan, Results of aquatic toxicity tests of chemicals conducted by Ministry of the Environment in Japan (-March 2016), (2016). http://www.env.go.jp/en/chemi/sesaku/aquatic_Mar_2016.pdf viimati alla laetud 12.06.2017.

[62] H. J. M. Verhaar, E. U. Ramos, J. L. M. Hermens, Classifying environmental pollutants. 2: Separation of class 1 (baseline toxicity) and class 2 ('polar narcosis') type compounds based on chemical descriptors, J. Chemom. 10 (1996) 149–162.

[63] H. J. Verhaar, J. Solbé, J. Speksnijder, C. J. van Leeuwen, J. L. Hermens, Classifying environmental pollutants: Part 3. External validation of the classification system, Chemosphere. 40 (2000) 875–883.

[64] G. Patlewicz, N. Jeliazkova, R. J. Safford, A. P. Worth, B. Aleksiev, An evaluation of the implementation of the Cramer classification scheme in the Toxtree software, SAR QSAR Environ. Res. 19 (2008) 495–524.

[65] Toxtree. toxtree.sourceforge.net viimati uuendatud 16.05.2018.

[66] A. Furuhama, K. Hasunuma, Y. Aoki, Interspecies quantitative structure–activity relationships (QSARs) for eco-toxicity screening of chemicals: the role of physicochemical properties, SAR QSAR Environ. Res. 26 (2015) 809–830.

[67] S. J. Enoch, M. Hewitt, M. T. D. Cronin, S. Azam, J. C. Madden, Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree, Chemosphere. 73 (2008) 243–248.

[68] American Chemical Society, CAS REGISTRY - The gold standard for chemical substance information. http://support.cas.org/content/chemical-substances viimati alla laetud 05.04.2018.

[69] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36.

[70] US EPA, Estimation Programs Interface Suite™ for Microsoft® Windows, United States Environmental Protection Agency, Washington, DC, USA, 2012.

[71] The PubChem Project. https://pubchem.ncbi.nlm.nih.gov/ viimati alla laetud 22.05.2018.

[72] Open Babel, or how I learned to love the chemistry file format — Open Babel v2.3.0 documentation. http://openbabel.org/docs/2.3.0/index.html viimati alla laetud 22.05.2018.

[73] Maestro, Schrödinger, LLC, New York, NY, USA, 2017.

[74] G. Schaftenaar, MOLDEN a visualization program of molecular and electronic structure. http://www.cmbi.ru.nl/molden/ viimati alla laetud 22.05.2018.

[75] F. A. Carey, R. M. Giuliano, Organic Chemistry, 8th ed., McGraw-Hill, 2010.

[76] MacroModel, Schrödinger, LLC, New York, NY, USA, 2017.

[77] E. Polak, G. Ribiere, Note sur la convergence de méthodes de directions conjuguées, Rev. Fr. Inform. Rech. Opérationnelle Sér. Rouge. 3 (1969) 35–43.
[78] J. J. Stewart, MOPAC: a semiempirical molecular orbital program, J. Comput. Aided Mol. Des. 4 (1990) 1–103.
[79] C. G. Broyden, The Convergence of a Class of Double-rank Minimization Algorithms 2. The New Algorithm, J. Inst. Math. Its Appl. 6 (1970) 222–231.
[80] D. Goldfarb, A family of variable-metric methods derived by variational means, Math. Comput. 24 (1970) 23–26.
[81] R. Fletcher, A new approach to variable metric algorithms, Comput. J. 13 (1970) 317–322.
[82] D. F. Shanno, Conditioning of Quasi-Newton Methods for Function Minimization, Math. Comput. 24 (1970) 647.
[83] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[84] M. J. S. Dewar, E. G. Zoebisch, Extension of AM1 to the halogens, J. Mol. Struct. THEOCHEM. 180 (1988) 1–21.
[85] M. J. S. Dewar, Y. C. Yuan, AM1 Parameters for Sulfur, Inorg. Chem. 29 (1990) 3881–3890.
[86] M. J. S. Dewar, C. Jie, AM1 Parameters for Phosphorus, J. Mol. Struct. THEOCHEM. 187 (1989) 1–13.
[87] A. R. Katritzky, R. Petrukhin, I. Petrukhina, A. Lomaka, D. B. Tatham, M. Karelson, CODESSA PRO, user's manual, (2005).
[88] US EPA, KOWWIN, United States Environmental Protection Agency, Washington, DC, USA, 2010.
[89] R. Petrukhin, Industrial Applications of the Quantitative Structure-Property Relationships, 2001.
[90] J. C. Dearden, M. T. D. Cronin, K. L. E. Kaiser, How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR), SAR QSAR Environ. Res. 20 (2009) 241–266.

[91] M. S. Karacan, Ç. Yakan, M. Yakan, N. Karacan, S. K. Zharmukhamedov, A. Shitov, D. A. Los, V. V. Klimov, S. I. Allakhverdiev, Quantitative structure–activity relationship analysis of perfluoroiso-propyldinitrobenzene derivatives known as photosystem II electron transfer inhibitors, Biochim. Biophys. Acta BBA - Bioenerg. 1817 (2012) 1229–1236.

[92] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.

[93] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Validation of QSAR models-strategies and importance, Int. J. Drug Des. Discov. 3 (2011) 511–519.

[94] M. Oja, U. Maran, The Permeability of an Artificial Membrane for Wide Range of pH in Human Gastrointestinal Tract: Experimental Measurements and Quantitative Structure□Activity Relationship, Mol. Inform. 34 (2015) 493–506.

[95] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review, ATLA. 33 (2005) 445–459.

[96] T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. M. van de Sandt, W. Tong, G. Veith, C. Yang, Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure–Activity Relationships, ATLA. 33 (2005) 155–173.

[97] T. Öberg, A QSAR for Baseline Toxicity: Validation, Domain of Application, and Prediction, Chem. Res. Toxicol. 17 (2004) 1630–1637.

[98] C. M. Ellison, J. C. Madden, M. T. D. Cronin, S. J. Enoch, Investigation of the Verhaar scheme for predicting acute aquatic toxicity: improving predictions obtained from Toxtree ver. 2.6, Chemosphere. 139 (2015) 146–154.

[99] A. R. Katritzky, D. B. Tatham, U. Maran, Theoretical Descriptors for the Correlation of Aquatic Toxicity of Environmental Pollutants by Quantitative Structure-Toxicity Relationships, J. Chem. Inf. Comput. Sci. 41 (2001) 1162–1176.

[100] T. K. Kim, T test as a parametric statistic, Korean J. Anesthesiol. 68 (2015) 540–546.

[101] A. Kahru, B. Borchardt, Toxicity of 39 MEIC chemicals to bioluminescent photobacteria (the Biotox test): correlation with other test systems, ATLA. 22 (1994) 147–160.

[102] S. Ren, A. Das, E. Lien, QSAR analysis of membrane permeability to organic compounds, J. Drug Target. 4 (1996) 103–107.

[103] M. Karelson, V. S. Lobanov, A. R. Katritzky, Quantum-Chemical Descriptors in QSAR/QSPR Studies, Chem. Rev. 96 (1996) 1027–1044.
[104] C. Hansch, W. E. Steinmetz, A. J. Leo, S. B. Mekapati, A. Kurup, D. Hoekman, On the Role of Polarizability in Chemical−Biological Interactions, J. Chem. Inf. Comput. Sci. 43 (2003) 120–125.
[105] X.-Y. Han, Z.-Y. Wang, Z.-C. Zhai, L.-S. Wang, Estimation of n-octanol/water Partition Coefficients (Kow) of all PCB Congeners by Ab initio and a Cl Substitution Position Method, QSAR Comb. Sci. 25 (2006) 333–341.
[106] A. R. Katritzky, S. H. Slavov, I. S. Stoyanova-Slavova, I. Kahn, M. Karelson, Quantitative Structure–Activity Relationship (QSAR) Modeling of EC50 of Aquatic Toxicities for Daphnia magna, J. Toxicol. Environ. Health A. 72 (2009) 1181–1190.
[107] W. Birge, R. Cassidy, Structure-activity relationships in aquatic toxicology, Fundam. Appl. Toxicol. 3 (1983) 359–368.
[108] Chevron Phillips Chemical Company LP, Methylcyclohexane, (n.d.).
[109] PubChem, Triclosan, (2018). https://pubchem.ncbi.nlm.nih.gov/compound/triclosan viimati uuendatud 19.05.2018.
[110] N. Islam, A. H. Pandith, Toxicity profile of aromatic compounds towards Scenedesmus obliquus: a QSAR study, Can. J. Chem. 91 (2013) 943–950.
[111] G.-N. Lu, X.-Q. Tao, Z. Dang, X.-Y. Yi, C. Yang, Estimation of n-octanol/water partition coefficients of polycyclic aromatic hydrocarbons by quantum chemical descriptors, Open Chem. 6 (2008).
[112] I. Kahn, S. Sild, U. Maran, Modeling the Toxicity of Chemicals to Tetrahymena pyriformis Using Heuristic Multilinear Regression and Heuristic Back-Propagation Neural Networks, J. Chem. Inf. Model. 47 (2007) 2271–2279.
[113] S. C. Basak, V. R. Magnuson, R. R. Regal, G. D. Veith, Topological indices: their nature, mutual relatedness, and applications, Math. Model. 8 (1987) 300–305.
[114] I. Tulp, S. Sild, U. Maran, Relationship Between Structure and Permeability in Artificial Membranes: Theoretical Whole Molecule Descriptors in Development of QSAR Models, QSAR Comb. Sci. 28 (2009) 811–814.

Lisad

Tabel L1. *MOPAC*'is kasutatavad esimesed märksõnad[a]

Märksõna	**Funktsioon ja kommentaare**
AM1	Kasutada AM1 parametrisatsiooni
NOXYZ	Ei esitata väljundfailis Cartesiaani koordinaate
PRECISE	Suurendada vaikimisi täpsuskriteeriumeid 100 korda
NOINTER	Väljundfaili ei lisata aatomitevahelisi kaugusi
GNORM=n.nn	Arvutus peatatakse, kui saavutatakse gradiendi norm alla n.nn kcal mol^{-1} $Å^{-1}$. Siinkohal n.nn = 0.01
NOMM	Ei kasutata molekulaarmehaanika parandust peptiidsidemetele. Kasutati vaid ühendites 132, 207, 256, 555, 637.
GEO-OK	Tavaliselt peatatakse arvutamine veateatega, kui aatomid satuvad üksteisele lähemale kui 0.1 ångstromi. GEO-OK lubab programmil seda ignoreerida. Kasutati vaid ühendites 311, 119, 395, 573.
T=n	Sätestab ajapiirangu n sekundit, mille möödudes arvutamine katkestatakse. Käesolevas töös n = 36000.

[a] Full list of keywords used in MOPAC, (n.d.). http://www.cup.uni-muenchen.de/archive/cicum/software/mopac7/node19.html viimati uuendatud 07.07.2003.

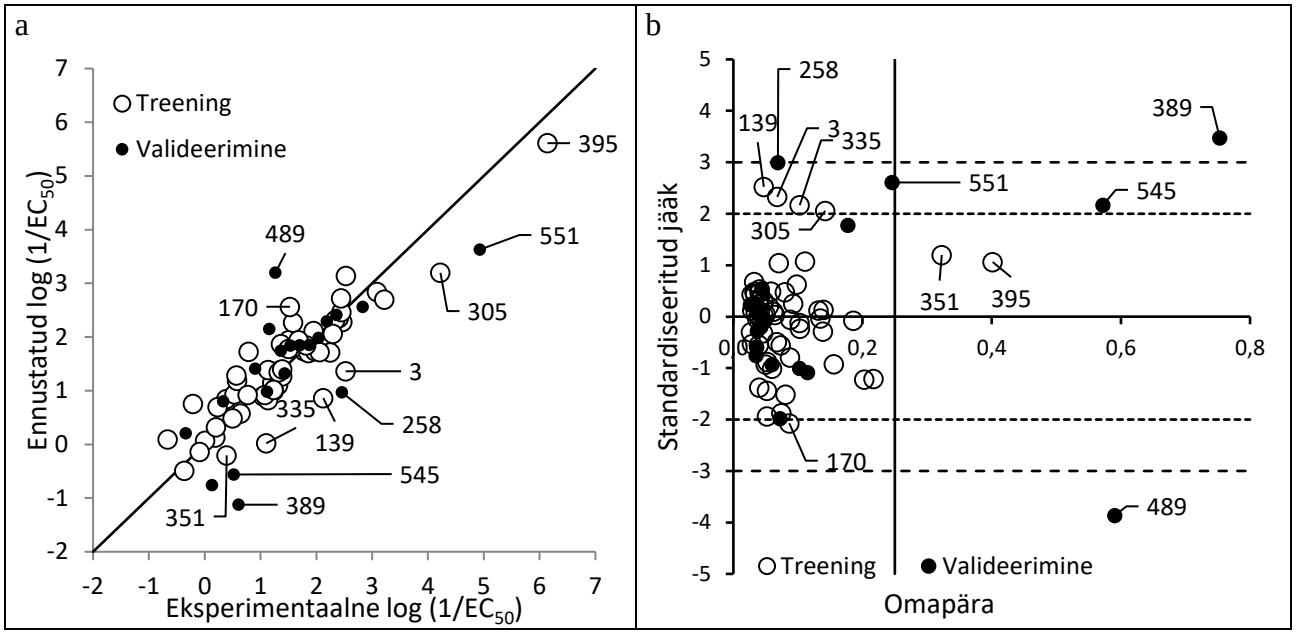
Tabel L2. *MOPAC*'is kasutatavad märksõnad täiendavateks arvutusteks[a]

Märksõna	Funktsioon ja kommentaare
AM1	Kasutada AM1 hamiltoniaani
VECTORS	Kirjutada väljundfaili lõplikud omavektorid
BONDS	Kirjutada väljundfaili lõplik sidemejärkude maatriks
PI	Tihedusmaatriksis kirjutatakse iga aatom-aatom interaktsioon välja eraldi σ, π, ja Δ sidemetena
POLAR	Arvutada esimest, teist ja kolmandat järku polariseeritavused
PRECISE	Suurendada vaikimisi täpsuskriteeriumeid 100 korda
ENPART	Energiad on jaotatud väljudfailis komponentidena
1SCF	Sooritada vaid üks SCF- arvutus
EF	Kasutada Baker'i omavektorjärgnemist[b] BFGS meetodi asemel. Kasutati vaid ühendites 395 ja 573.
NOMM	Ei kasutata molekulaarmehaanika parandust peptiidsidemetele. Kasutati vaid ühendites 132, 207, 256, 555, 637.

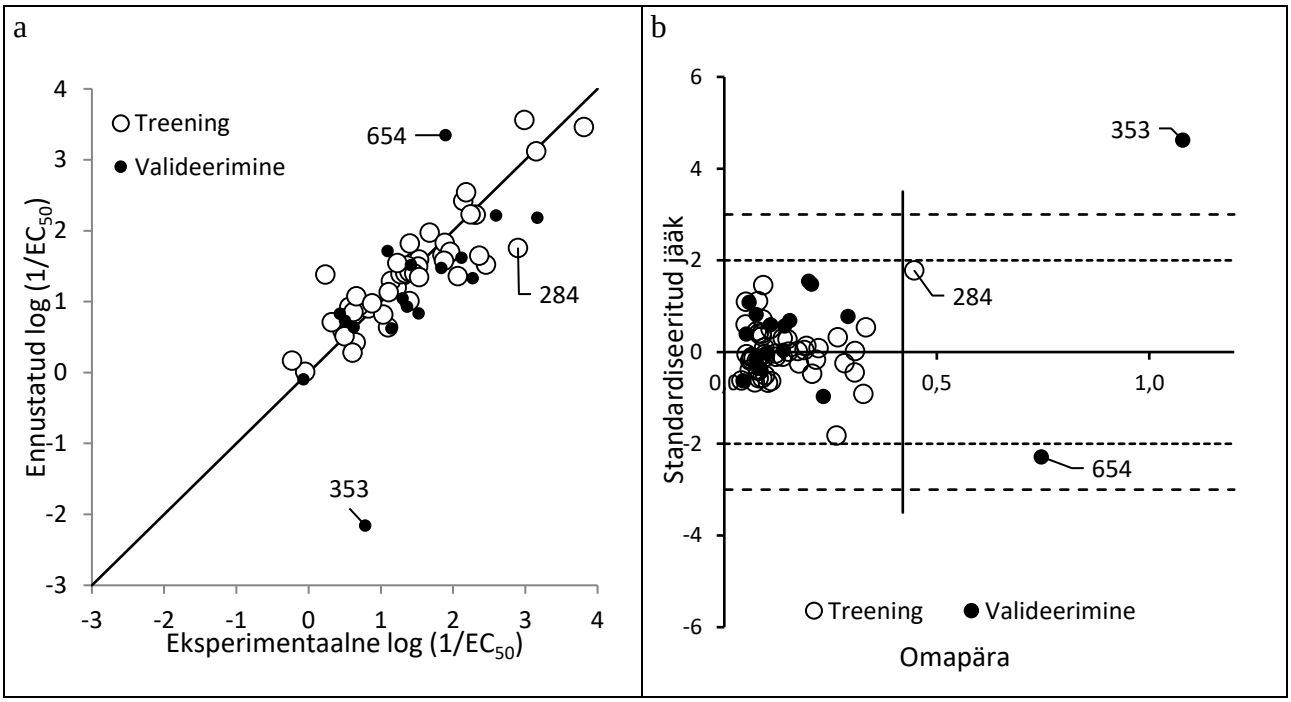
[a] Full list of keywords used in MOPAC, (n.d.). http://www.cup.uni-muenchen.de/archive/cicum/software/mopac7/node19.html viimati uuendatud 07.07.2003.

[b] J. Baker, An Algorithm for the Location of Transition States, J. Comput. Chem. 7 (1986) 385–395.

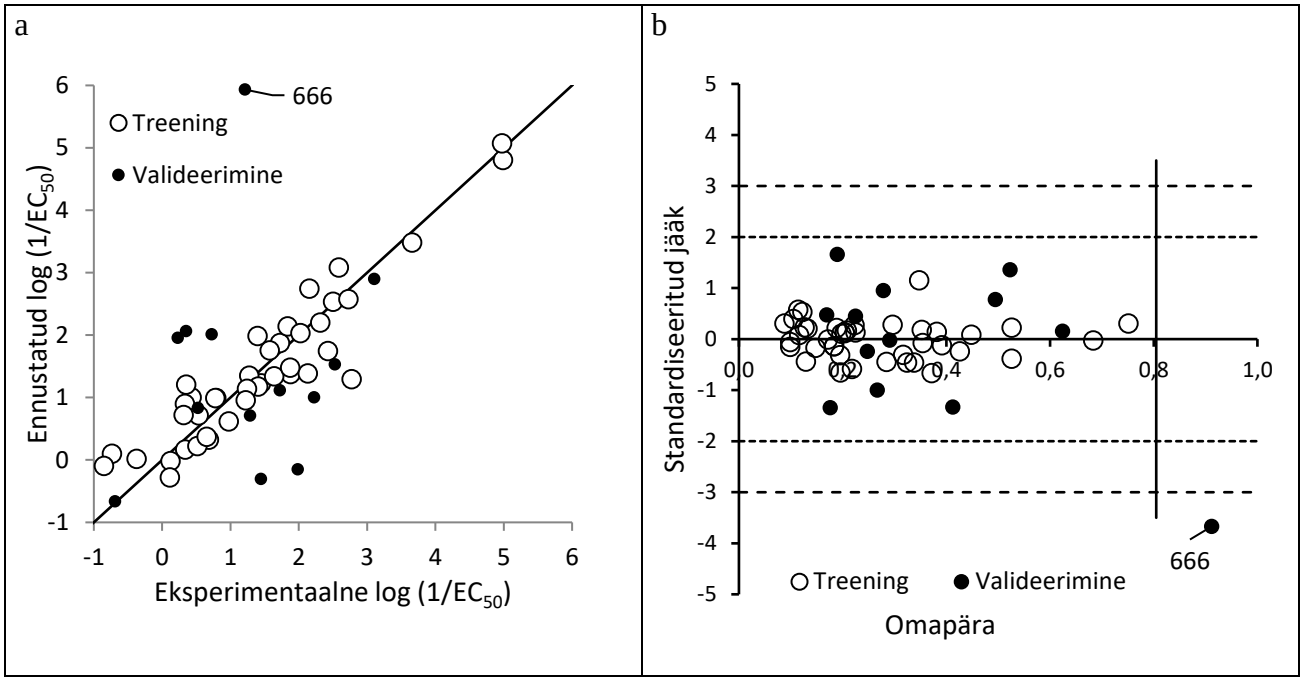
Graafik L1. Klass 1: a) Ennustatud vs eksperimentaalne log (1/EC_{50}); b) Williams'i graafik



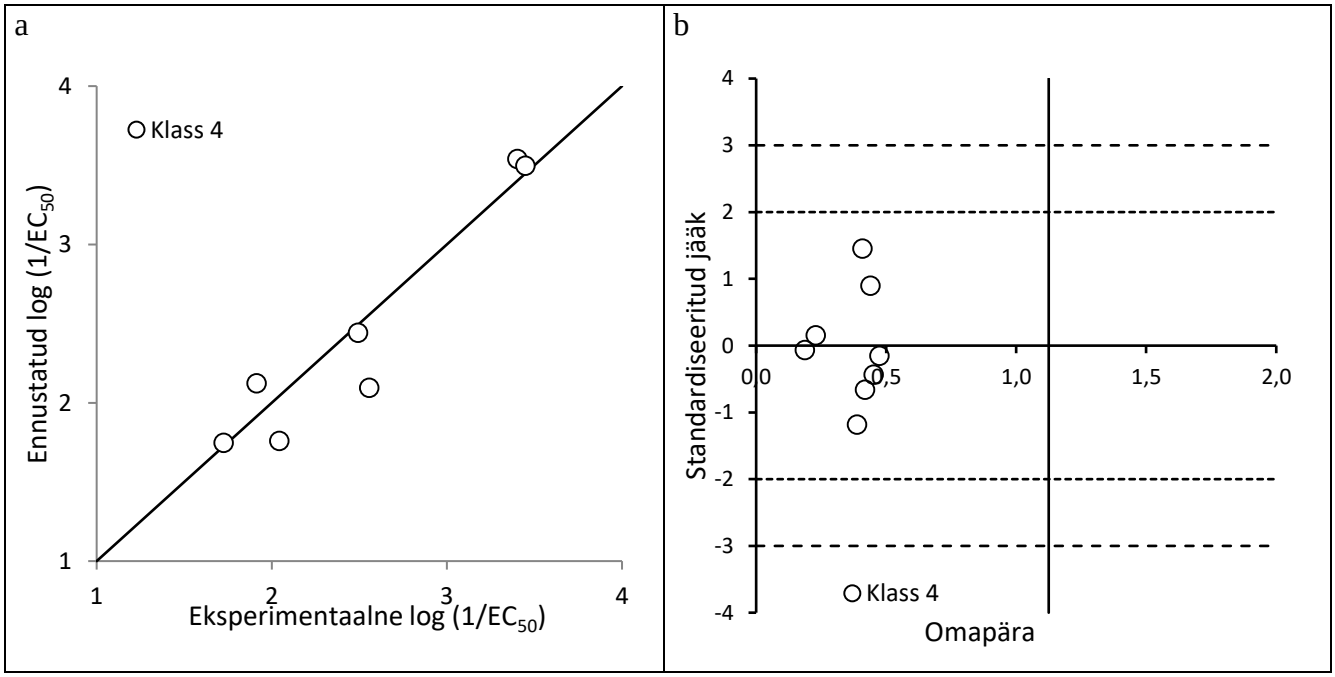
Graafik L2. Klass 2: a) Ennustatud vs eksperimentaalne log (1/EC_{50}); b) Williams'i graafik



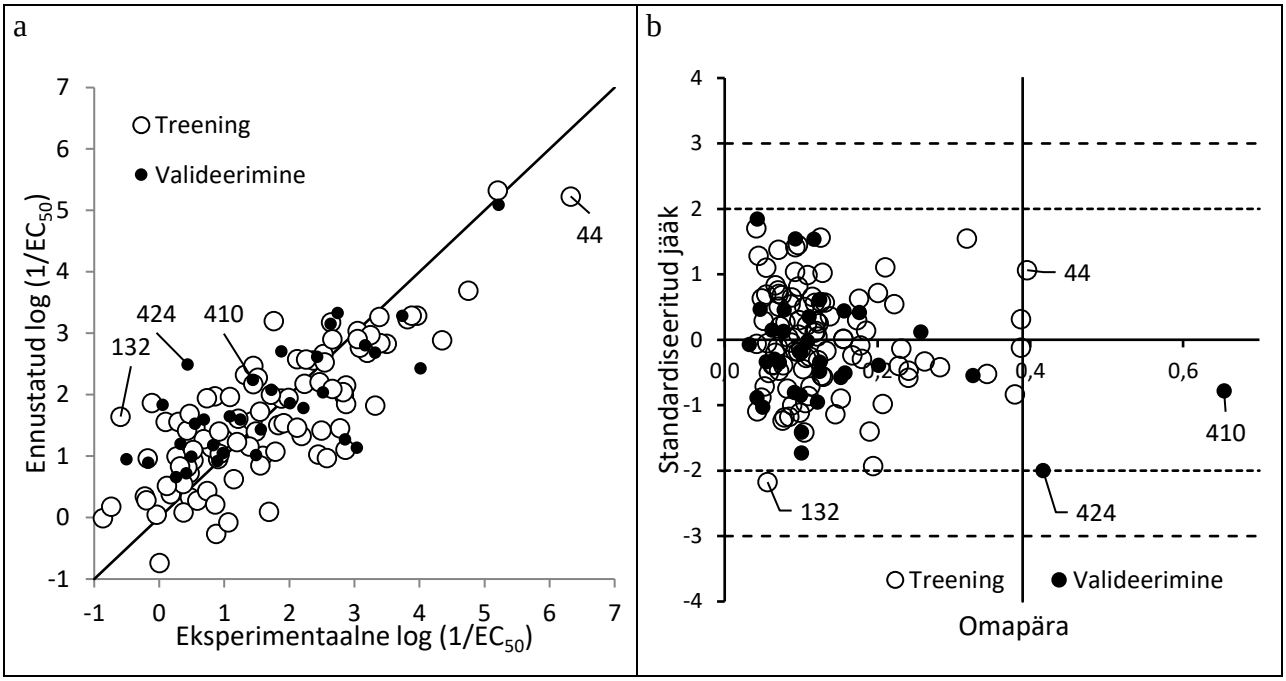
Graafik L3. Klass 3: a) Ennustatud vs eksperimentaalne log (1/EC_{50}); b) Williams'i graafik



Graafik L4. Klass 4: a) Ennustatud vs eksperimentaalne log (1/EC_{50}); b) Williams'i graafik



Graafik L5. Klass 5: a) Ennustatud vs eksperimentaalne log (1/EC_{50}); b) Williams'i graafik



Tabel L3. Uurimuses kasutatud andmeseeria: eksperimentaalne ja ennustatud log ($1/EC_{50}$), valim, nimetus, klass modifitseeritud Verhaar'i reeglite järgi[a]

Nr	log ($1/EC_{50}$) (eksp.)	log ($1/EC_{50}$) (ennust.)	log ($1/EC_{50}$) (vahe)	Valim	Nimetus	Klass
43	0.19	0.129484	0.0605161	Treening	Pivalic acid	1
88	2.055	1.95246	0.102538	Treening	1,2,3-Trichlorobenzene	1
140	1.21	1.14966	0.0603388	Treening	o-Chlorotoluene	1
211	2.468	2.28295	0.185052	Treening	Diphenyl ether	1
224	1.924	1.91432	0.00967845	Treening	Butylbenzene	1
232	1.044	0.895693	0.148307	Treening	p-Xylene	1
233	1.317	1.1004	0.216596	Treening	p-Chlorotoluene	1
266	0.642	0.574373	0.0676269	Treening	1,3-Dibromopropane	1
290	0.385	0.844839	-0.459839	Treening	N,N,N',N'-Tetramethylhexamethylenediamine	1
294	0.534	0.930914	-0.396914	Treening	1-Heptanol	1
296	1.12	0.854346	0.265654	Treening	1,5-Cyclooctadiene	1
318	1.776	1.73951	0.0364864	Treening	2,6-Dichlorotoluene	1
331	1.503	1.94357	-0.440573	Treening	1,2,4-Trichlorobenzene	1
395	6.14	5.60966	0.530336	Treening	Indeno[1,2,3-cd]pyrene	1
416	-0.21	0.756687	-0.966687	Treening	2-Oxabicyclo[2.2.2]octane, 1,3,3-trimethyl-	1
439	1.581	2.26816	-0.687156	Treening	1,2-Dimethylnaphthalene	1
440	2.401	2.34918	0.0518242	Treening	1,3-Dimethylnaphthalene	1
518	2.251	1.71528	0.53572	Treening	Ethylcyclohexane	1
575	3.085	2.84243	0.242572	Treening	1,1'-Biphenyl, 4-ethyl- <4-Ethyl-1,1'-biphenyl>	1
645	1.367	1.87186	-0.50486	Treening	Isodecyl alcohol	1
55	0.007	0.068464	-0.061464	Treening	Isoprene	1
58	0.232	0.696957	-0.464957	Treening	Trichloroethylene	1
85	2.34	2.31709	0.0229052	Treening	Fluorene	1
139	2.123	0.862691	1.26031	Treening	o-Xylene	1
157	0.581	1.1855	-0.604504	Treening	1,2-Dibromo-3-chloropropane	1
170	1.528	2.56302	-1.03502	Treening	4-tert-Butyltoluene	1
173	1.391	1.27028	0.120719	Treening	2-Phenylpropene	1
219	1.684	1.82793	-0.143927	Treening	Dibenzyl ether	1
243	-0.366	-0.490937	0.124937	Treening	1,2-Dichloroethane	1
259	0.502	0.484756	0.0172441	Treening	Toluene	1
302	2.451	2.46735	-0.0163549	Treening	1-Decanol	1
305	4.223	3.1957	1.0273	Treening	Tridecan-1-ol	1
335	1.102	0.020044	1.08196	Treening	Triethylamine	1
355	1.336	1.37473	-0.0387313	Treening	Dibromochloromethane	1
360	0.788	1.72363	-0.935626	Treening	Tetrachloroethylene	1

Nr	log ($1/EC_{50}$) (eksp.)	log ($1/EC_{50}$) (ennust.)	log ($1/EC_{50}$) (vahe)	Valim	Nimetus	Klass
367	2.532	3.14131	-0.609311	Treening	2-Hydroxy-4-methoxybenzophenone	1
404	1.139	1.38518	-0.246179	Treening	4-Bromo-1,2-difluorobenzene	1
505	1.859	1.70538	0.153617	Treening	Divinylbenzene	1
526	1.237	0.997753	0.239247	Treening	2-Chlorohydroquinonedimethylether	1
628	1.976	1.73592	0.240077	Treening	2,5-Dichlorotoluene	1
3	2.524	1.36138	1.16262	Treening	Carbon tetrachloride	1
35	1.135	0.823471	0.311529	Treening	Bromodichloromethane	1
56	-0.093	-0.140103	0.0471026	Treening	1,2-Dichloropropane	1
80	2.445	2.72328	-0.278283	Treening	Phenanthrene	1
94	0.202	0.31705	-0.11505	Treening	Butanoic acid, 2-ethyl-	1
126	2.296	2.05992	0.236082	Treening	Biphenyl	1
148	2.061	1.72253	0.33847	Treening	3,4-Dichlorotoluene	1
250	1.077	0.922241	0.154759	Treening	m-Xylene	1
300	1.952	2.11153	-0.159533	Treening	2-Undecanone	1
351	0.394	-0.204576	0.598576	Treening	Adipic acid	1
352	0.568	1.28569	-0.717692	Treening	Octanoic acid	1
382	0.772	0.922056	-0.150056	Treening	Butane, 1,1'-oxybis-	1
384	1.817	1.84679	-0.0297882	Treening	1-Nonanol	1
425	1.324	1.34662	-0.0226241	Treening	1,2,3-Trimethylbenzene	1
447	3.218	2.6961	0.521897	Treening	2-Tridecanone	1
466	1.504	1.77994	-0.27594	Treening	2-Decanone	1
546	1.235	1.01473	0.220271	Treening	3a,4,7,7a-Tetrahydro-1H-indene	1
616	-0.663	0.094696	-0.757696	Treening	2,2,5,5-Tetramethyltetrahydrofuran	1
618	1.39	1.39946	-0.00946436	Treening	5-Ethylidene-8,9,10-trinorborn-2-ene	1
643	1.679	1.94364	-0.264638	Treening	Diisopropylbenzene	1
74	2.042	1.98179	0.0602105	Valideerimine	Acenaphthene	1
110	-0.337	0.204285	-0.541285	Valideerimine	2-Methoxyphenol	1
111	1.706	1.84272	-0.136725	Valideerimine	1-Methylnaphthalene	1
120	1.874	1.85157	0.0224341	Valideerimine	2-Methylnaphthalene	1
131	0.909	1.41109	-0.502095	Valideerimine	4-Allyl-1,2-dimethoxybenzene	1
183	1.364	1.74209	-0.378085	Valideerimine	p-Cymene	1
236	1.435	1.31875	0.11625	Valideerimine	p-Dichlorobenzene	1
257	1.117	0.984152	0.132848	Valideerimine	Bromobenzene	1
258	2.461	0.96647	1.49453	Valideerimine	Methylcyclohexane	1
268	0.132	-0.756211	0.888211	Valideerimine	Diethylamine	1
287	0.336	0.800743	-0.464743	Valideerimine	Heptanoic acid	1
389	0.612	-1.12327	1.73527	Valideerimine	Oxalic acid	1
402	1.157	2.14652	-0.989516	Valideerimine	Decanoic acid	1

Nr	log (1/EC_{50}) (eksp.)	log (1/EC_{50}) (ennust.)	log (1/EC_{50}) (vahe)	Valim	Nimetus	Klass
445	2.194	2.29528	-0.101285	Valideerimine	2,7-Dimethylnaphthalene	1
480	2.366	2.40967	-0.0436654	Valideerimine	Cyclohexylbenzene	1
489	1.27	3.20106	-1.93106	Valideerimine	2,2-Bis[4-(2-hydroxyethoxy)phenyl]propane	1
545	0.523	-0.561097	1.0841	Valideerimine	N-Methyl-N,N-bis(2-dimethylaminoethyl)amine	1
551	4.93	3.62699	1.30301	Valideerimine	Triclosan	1
638	1.54	1.83678	-0.296777	Valideerimine	Butylated hydroxyanisole	1
657	2.833	2.56037	0.272629	Valideerimine	Isopropylnaphthalene	1
87	0.471	0.587042	-0.116042	Treening	2,3-Dimethylaniline	2
98	1.937	1.67659	0.260414	Treening	6-tert-Butyl-m-cresol	2
104	1.308	1.37244	-0.0644411	Treening	4-Chloro-2-nitroaniline	2
105	1.447	1.5496	-0.102596	Treening	4-Chloro-2-nitrophenol	2
113	3.81	3.46435	0.345647	Treening	1-(N-phenylamino)-naphthalene	2
127	1.851	1.6579	0.193098	Treening	p-Phenylphenol	2
143	-0.049	0.0141731	-0.0631731	Treening	o-Toluidine	2
161	2.315	2.22853	0.0864653	Treening	2,4-Di-tert-butylphenol	2
208	1.218	1.19942	0.0185847	Treening	4,4'-Methylenedianiline	2
221	0.565	0.927325	-0.362325	Treening	N-Ethylaniline	2
230	1.1	0.641981	0.458019	Treening	2,4-Xylenol	2
239	0.65	0.429525	0.220475	Treening	p-Toluidine	2
251	0.827	0.907893	-0.0808934	Treening	m-Chloroaniline	2
284	2.898	1.761	1.137	Treening	Pyridine	2
327	1.528	1.59507	-0.0670658	Treening	o-Tolidine	2
333	2.139	2.4214	-0.2824	Treening	2,4-Di-tert-pentylphenol	2
562	1.395	1.00967	0.385328	Treening	4-(1-Methylethenyl)phenol	2
67	1.677	1.97562	-0.298624	Treening	Bisphenol A	2
93	0.7	0.942593	-0.242593	Treening	2,4,6-Trimethylaniline	2
95	2.063	1.36316	0.699841	Treening	2-tert-Butylphenol	2
107	1.031	0.821286	0.209714	Treening	Thymol	2
121	2.457	1.52561	0.931389	Treening	β-Naphthylamine	2
130	1.514	1.49568	0.0183179	Treening	Biphenyl-4,4'-diol	2
177	1.885	1.82953	0.0554725	Treening	3,4-Dichloronitrobenzene	2
187	1.137	1.29348	-0.156477	Treening	p-Nitrotoluene	2
188	0.507	0.765265	-0.258265	Treening	p-Nitroaniline	2
193	1.359	1.51042	-0.151421	Treening	4-Vinylpyridine	2
231	1.274	1.39026	-0.116261	Treening	4-Bromophenol	2
234	0.318	0.710411	-0.392411	Treening	p-Cresol	2
254	0.621	0.858825	-0.237825	Treening	3,5-Dimethylaniline	2
262	-0.23	0.171316	-0.401316	Treening	Phenol	2

Nr	log ($1/EC_{50}$) (eksp.)	log ($1/EC_{50}$) (ennust.)	log ($1/EC_{50}$) (vahe)	Valim	Nimetus	Klass
319	2.241	2.23087	0.0101271	Treening	2,4,6-Tribromophenol	2
454	3.149	3.11984	0.0291578	Treening	Di-p-tolylamine	2
523	1.4	1.82197	-0.421972	Treening	2-(1,1-Dimethylethyl)-4,6-dimethylphenol	2
99	0.497	0.514807	-0.0178074	Treening	o-Nitroaniline	2
106	1.338	1.40445	-0.0664524	Treening	o-sec-Butylphenol	2
142	0.659	1.07784	-0.418844	Treening	o-Chloroaniline	2
149	1.397	1.42073	-0.0237308	Treening	3,4-Dichloroaniline	2
150	0.606	0.284329	0.321671	Treening	2,5-Dimethylaniline	2
152	1.461	1.39604	0.0649565	Treening	2-Chloro-5-methylaniline	2
153	1.227	1.54832	-0.321318	Treening	2,5-Dichloroaniline	2
184	0.876	0.976312	-0.100312	Treening	4-Isopropylaniline	2
202	0.229	1.3851	-1.1561	Treening	2-Vinylpyridine	2
238	1.109	1.13557	-0.0265693	Treening	p-Chlorophenol	2
332	1.531	1.34777	0.183229	Treening	2,4-Dichlorophenol	2
448	2.181	2.54242	-0.361415	Treening	4-(α,α-Dimethylbenzyl)phenol	2
451	2.36	1.65131	0.708686	Treening	2,4-Dibromophenol	2
493	2.985	3.56237	-0.577368	Treening	2-Phenylindole	2
532	1.96	1.70462	0.255384	Treening	2-tert-Butyl-p-cresol	2
615	1.873	1.57996	0.293037	Treening	4-Pentylphenol	2
19	-0.072	-0.0963949	0.0243949	Valideerimine	Aniline	2
100	1.365	0.927998	0.437002	Valideerimine	o-Nitrophenol	2
146	1.149	0.62355	0.52545	Valideerimine	3,4-Dimethylaniline	2
154	0.625	0.639088	-0.0140876	Valideerimine	2,5-Xylenol	2
155	2.119	1.6204	0.498597	Valideerimine	2,4,5-Trichlorophenol	2
176	0.507	0.729287	-0.222287	Valideerimine	Benzenamine, 3-nitro- <m-Nitroaniline>	2
180	1.302	1.04754	0.25446	Valideerimine	p-sec-Butylphenol	2
237	1.526	0.835295	0.690705	Valideerimine	p-Chloroaniline	2
337	2.595	2.21529	0.379705	Valideerimine	Diphenylamine	2
353	0.786	-2.15946	2.94546	Valideerimine	1,6-Hexanediamine	2
372	1.837	1.4748	0.362204	Valideerimine	2-Naphthol	2
442	0.434	0.831379	-0.397379	Valideerimine	2,6-Xylenol	2
453	1.097	1.71281	-0.615806	Valideerimine	4,4'-Dihydroxydiphenylmethane	2
519	3.168	2.1837	0.9843	Valideerimine	p-Octylphenol	2
527	1.416	1.51669	-0.100692	Valideerimine	6-tert-Butyl-o-cresol	2
571	2.273	1.32938	0.943622	Valideerimine	2,6-Di-sec-butylphenol	2
654	1.895	3.34658	-1.45158	Valideerimine	Phenol, 4,4',4"-ethylidynetris-	2
60	2.771	1.29106	1.47994	Treening	Chloroacetic acid	3
82	0.338	0.165239	0.172761	Treening	Phthalic anhydride	3

Nr	log (1/EC_{50}) (eksp.)	log (1/EC_{50}) (ennust.)	log (1/EC_{50}) (vahe)	Valim	Nimetus	Klass
156	0.682	0.322565	0.359435	Treening	Styrene oxide	3
160	1.444	1.22506	0.218936	Treening	Methyl acrylate	3
274	0.12	-0.0171319	0.137132	Treening	Sorbic acid	3
291	1.722	1.87525	-0.153245	Treening	Glutaraldehyde	3
361	2.498	2.5324	-0.0344048	Treening	2,3,3,3,2',3',3',3'-Octachlorodipropyl ether	3
380	1.877	1.3751	0.501902	Treening	n-Butyl acrylate	3
419	1.273	1.34747	-0.074468	Treening	2,4,6-Trimethylbenzaldehyde	3
456	0.512	0.223417	0.288583	Treening	Dimethyl disulphide	3
485	-0.737	0.0995896	-0.83659	Treening	2-Hydroxyethyl methacrylate	3
512	0.972	0.619259	0.352741	Treening	2-(1'-Cyclohexenyl)cyclohexanone	3
605	2.151	2.74295	-0.591949	Treening	Hexamethylene diacrylate	3
675	4.989	4.80257	0.186434	Treening	2-Chloro-2',6'-diethyl-N-(2-propoxyethyl)acetanilide <Pretilachlor>	3
18	0.112	-0.280577	0.392577	Treening	3-Amino-1,2,4-triazole	3
108	1.406	1.17765	0.228352	Treening	2-Hydroxybenzaldehyde	3
165	0.791	0.988803	-0.197803	Treening	n-Butyl methacrylate	3
194	-0.853	-0.0977141	-0.755286	Treening	Benzyl alcohol	3
242	0.648	0.371878	0.276122	Treening	Glycidyl methacrylate	3
340	2.425	1.74313	0.681873	Treening	Benzalacetone	3
359	0.429	1.00032	-0.571317	Treening	Methacrylonitrile	3
379	1.639	1.33936	0.299638	Treening	Ethyl acrylate	3
450	2.128	1.38459	0.743412	Treening	2-Chlorobenzyl chloride	3
470	0.333	0.89954	-0.56654	Treening	3,4-Dichlorobut-1-ene	3
542	1.242	1.1384	0.103602	Treening	2-(Dimethylamino)ethyl methacrylate <DMMA>	3
560	1.873	1.47865	0.394351	Treening	Crotonaldehyde	3
636	4.975	5.07185	-0.0968468	Treening	N-(Butoxymethyl)-2-Chloro-2',6'-diethylacetanilide	3
659	2.723	2.57431	0.148687	Treening	Dibromocresyl glycidyl ether	3
14	1.833	2.14056	-0.307559	Treening	Aniline, p-(phenylazo)- <p-Aminoazobenzene>	3
174	0.776	0.993022	-0.217022	Treening	alpha,alpha-Dichlorotoluene	3
189	3.655	3.48481	0.170193	Treening	alpha-Chloro-4-nitrotoluene	3
292	-0.376	0.0174169	-0.393417	Treening	Bis(2-chloroethyl) ether	3
313	0.532	0.712311	-0.180311	Treening	Tetrachlorophthalic anhydride *3	3
343	1.223	0.956367	0.266633	Treening	2-Methylvaleraldehyde	3
400	0.351	1.20701	-0.856011	Treening	Glyoxylic acid	3
436	2.585	3.08234	-0.497335	Treening	Methyl isothiocyanate	3
464	1.573	1.75326	-0.180264	Treening	2-Ethylhexyl methacrylate	3

Nr	log ($1/EC_{50}$) (eksp.)	log ($1/EC_{50}$) (ennust.)	log ($1/EC_{50}$) (vahe)	Valim	Nimetus	Klass
473	2.313	2.2031	0.109904	Treening	2-Butenedinitrile, (E)-	3
498	1.396	1.98458	-0.588579	Treening	2-Butenenitrile, 3-amino-	3
547	2.022	2.03084	-0.00883983	Treening	3-(Methylthio)propionaldehyde	3
604	0.312	0.717418	-0.405418	Treening	Tetrahydromethylphthalic anhydride	3
30	0.229	1.9538	-1.7248	Valideerimine	Acetaldehyde	3
59	1.983	-0.150834	2.13383	Valideerimine	Acrylic acid	3
158	-0.689	-0.663407	-0.0255929	Valideerimine	1,3-Dichloro-2-propanol	3
195[b]	0.521	0.831043	-0.310043	Valideerimine	Benzaldehyde	3
244	0.725	2.01041	-1.28541	Valideerimine	2-Propenenitrile	3
280	1.444	-0.304925	1.74893	Valideerimine	Diethyl disulfide	3
341	2.224	1.00263	1.22137	Valideerimine	Hydrazobenzene	3
342	0.349	2.06371	-1.71471	Valideerimine	p-Methoxybenzaldehyde	3
412	2.527	1.52748	0.999522	Valideerimine	4,4,4-Trifluorocrotonitrile	3
430	1.723	1.11494	0.608064	Valideerimine	1,3-Dichloropropene	3
476	1.287	0.708021	0.578979	Valideerimine	2-Hydroxyethyl acrylate	3
490	3.103	2.8999	0.203103	Valideerimine	2-Propenenitrile, 2-chloro-	3
666	1.212	5.93292	-4.72092	Valideerimine	Tris(2-hydroxyethyl)isocyanuric acid acrylate	3
11	2.043	1.7586	0.2844	-	2,3,4,6-Tetrachlorophenol	4
92	2.491	2.44191	0.0490862	-	Pentachlorophenol	4
311	1.912	2.12213	-0.210125	-	Triphenyl phosphate	4
458	1.725	1.74591	-0.020909	-	2,4,6-Trichloroaniline	4
555	0.798	1.17482	-0.376817	-	2-sec-Butylphenyl N-methylcarbamate <Fenobucarb>	4
567	3.401	3.53934	-0.138337	-	2,4,6-Trichlorophenylhydrazine	4
627	2.556	2.09382	0.462182	-	Isoxathion	4
655	3.447	3.49648	-0.0494799	-	S-4-Chlorobenzyl diethylthiocarbamate <Benthiocarb>	4
31	1.316	2.32404	-1.00804	Treening	Ethanethiol	5
44	6.324	5.22902	1.09498	Treening	Trichloronitromethane	5
52	-0.221	0.346114	-0.567114	Treening	3,5,5-Trimethyl-2-cyclohexen-1-one	5
64	1.758	3.20108	-1.44308	Treening	Tetrabromobisphenol A	5
101	2.234	2.17999	0.0540061	Treening	Phenol, 2-(1-methylpropyl)-4,6-dinitro- <Dinoseb>	5
123	0.421	1.59217	-1.17117	Treening	2,4-Diamino-6-phenyl-s-triazine	5
145	2.862	1.10834	1.75366	Treening	o-Aminophenol	5
162	1.147	0.62491	0.52209	Treening	2-Nitro-p-anisidine	5
175	0.904	0.956534	-0.0525343	Treening	m-Toluic acid	5
185	0.265	0.996144	-0.731144	Treening	4-Methylbenzoic acid	5

Nr	log (1/EC_{50}) (eksp.)	log (1/EC_{50}) (ennust.)	log (1/EC_{50}) (vahe)	Valim	Nimetus	Klass
186	0.099	1.55714	-1.45814	Treening	4-Hydroxybenzoic acid	5
191	3.05	3.03836	0.0116449	Treening	p-Dinitrobenzene	5
207	3.818	3.23037	0.58763	Treening	3,4,4'-Trichlorodiphenylurea	5
209	0.854	1.97701	-1.12301	Treening	4,4'-Diaminodiphenyl ether	5
213	1.444	2.47072	-1.02672	Treening	1,3-Diphenylguanidine	5
240	2.779	1.45473	1.32427	Treening	p-Phenylenediamine	5
248	0.986	1.28528	-0.299278	Treening	Vinyl acetate	5
252	0.557	1.46385	-0.90685	Treening	m-Phenylenediamine	5
256	-0.867	-0.00818145	-0.858819	Treening	Isocyanuric acid	5
261	0.465	0.727455	-0.262455	Treening	Cyclohexylamine	5
270	-0.116	1.8669	-1.9829	Treening	Thiophene	5
301[b]	0.734	1.94381	-1.20981	Treening	Triethylenetetramine	5
304	3.198	2.68308	0.514923	Treening	Tetraethylenepentamine	5
369	2.119	2.57782	-0.458819	Treening	Dibenzothiophene	5
386	4.349	2.88961	1.45939	Treening	1-Decanethiol	5
391	2.446	1.03017	1.41583	Treening	8-Hydroxyquinoline	5
392	2.524	2.66488	-0.140884	Treening	2-Mercaptobenzothiazole	5
403	0.514	0.299775	0.214225	Treening	o-Fluoroaniline	5
530	1.877	1.93749	-0.0604865	Treening	1,5-Naphthalenediamine	5
535	3.383	3.26737	0.115635	Treening	6-Methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	5
561	0.344	0.771952	-0.427952	Treening	Thiourea dioxide	5
584	1.592	1.00765	0.584351	Treening	2,2'-Dimethyl-4,4'-methylenebis(cyclohexylamine)	5
635	1.514	2.27991	-0.765907	Treening	1-Benzo[b]thien-2-ylethan-1-one	5
677	2.641	3.17739	-0.536394	Treening	Bis(2,2,6,6-tetramethyl-4-piperidyl) sebacate	5
10	2.874	2.15113	0.722866	Treening	Menadione	5
15	2.662	2.90346	-0.241461	Treening	2-Mercaptoethanol	5
21	0.874	-0.265147	1.13915	Treening	Sulphanilamide	5
51	0.801	1.14146	-0.340456	Treening	Tri-n-butoxyethyl phosphate	5
62	0.681	1.27718	-0.596182	Treening	Thiosemicarbazide	5
66	0.472	0.332083	0.139917	Treening	Hydrogenatedbisphenol A	5
78	2.189	1.32849	0.860508	Treening	Diisobutyl phthalate	5
96	0.003	-0.736214	0.739214	Treening	o-Toluenesulfonamide	5
119	2.492	1.42028	1.07172	Treening	6-Ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	5
132	-0.594	1.64519	-2.23919	Treening	o-Acetoacetotoluidide	5
169	1.432	1.56925	-0.137249	Treening	Benzotrifluoride <(Trifluoromethyl)benzene>	5

Nr	log (1/EC_{50}) (eksp.)	log (1/EC_{50}) (ennust.)	log (1/EC_{50}) (vahe)	Valim	Nimetus	Klass
171	0.524	0.93508	-0.41108	Treening	4-Toluenesulfonyl chloride <p-Toluene sulfonyl chloride>	5
181	0.434	0.824908	-0.390908	Treening	Methyl p-hydroxybenzoate	5
212	2.37	2.55915	-0.189149	Treening	1,4-Benzenediamine, N,N'-bis(1-methylpropyl)-	5
222[b]	1.985	1.94761	0.0373927	Treening	N,N-Dimethylbenzylamine	5
263	2.838	1.99914	0.838863	Treening	Benzenethiol	5
283	-0.179	0.966704	-1.1457	Treening	Piperazine	5
285	0.177	0.380435	-0.203435	Treening	Morpholine	5
314	3.499	2.83012	0.66888	Treening	2,3-Dichloro-1,4-naphthoquinone	5
325	1.717	2.00112	-0.284122	Treening	Benzophenone	5
345	3.318	1.82419	1.49381	Treening	Hydroquinone	5
366	1.215	1.61266	-0.397658	Treening	Diallyl phthalate	5
407	0.374	0.0836101	0.29039	Treening	m-Fluoroaniline	5
414	2.576	0.968351	1.60765	Treening	3-Aminopyridine	5
435	0.93	1.03892	-0.108919	Treening	m-Nitroanisole	5
449	1.084	1.96581	-0.881813	Treening	2,6-Dinitrotoluene	5
497	5.206	5.32753	-0.121526	Treening	Tri-n-octylamine	5
511	3.964	3.28708	0.676918	Treening	9-Vinylcarbazole	5
534	3.082	2.76931	0.312689	Treening	Dodecanenitrile	5
603	0.318	0.837977	-0.519977	Treening	5,6,7,8-Tetrahydroquinoline	5
649	1.487	1.40294	0.084055	Treening	Iprobenfos	5
676	1.556	0.853134	0.702866	Treening	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	5
685	1.83	1.50516	0.324844	Treening	3,5-Bis(trifluoromethyl)benzylamine	5
13	1.688	0.0933538	1.59465	Treening	Ethylenediaminetetraacetic acid	5
41	1.913	1.54038	0.372619	Treening	tert-Butylhydroperoxide	5
68	0.585	0.275641	0.309359	Treening	Bis(4-hydroxyphenyl)sulfone	5
115	-0.038	0.0486658	-0.0866658	Treening	Phthalonitrile	5
117	0.292	1.56188	-1.26988	Treening	Quinoline	5
118	0.429	1.41998	-0.990981	Treening	o-Nitroanisole	5
124	2.257	2.57704	-0.320042	Treening	3,3'-Dichlorobenzidine	5
128	0.468	1.68937	-1.22137	Treening	3-Hydroxy-2-naphthoic acid	5
136	3.4	2.84185	0.558149	Treening	N-(tert-Butyl)-2-benzothiazolylsulfenamide	5
138	3.246	2.97106	0.274942	Treening	N-Cyclohexyl-2-benzothiazolylsulfenamide	5
144	2.12	1.46682	0.653181	Treening	o-Phenylenediamine	5
147	1.786	1.07268	0.71332	Treening	2,5-Diaminotoluene	5
159	0.736	0.436172	0.299828	Treening	2-Butanone oxime	5
163	3.051	2.90729	0.143712	Treening	1-Chloro-2,4-dinitrobenzene	5

Nr	log (1/EC_{50}) (eksp.)	log (1/EC_{50}) (ennust.)	log (1/EC_{50}) (vahe)	Valim	Nimetus	Klass
164	1.444	2.17594	-0.731945	Treening	N,N'-Bis(2-methylphenyl)guanidine	5
201	1.062	-0.0749502	1.13695	Treening	Cyclohexanone oxime	5
215	0.917	1.4029	-0.485899	Treening	2-(Dibutylamino)ethanol	5
227	-0.74	0.181326	-0.921326	Treening	Diethyl malonate	5
277	2.872	1.85009	1.02191	Treening	Pentane-1-thiol	5
297	2.826	2.04569	0.780312	Treening	Methyl dodecanoate	5
310	0.369	0.547711	-0.178711	Treening	2-Amino-2-ethylpropanediol	5
312	-0.198	0.28509	-0.48309	Treening	Tris(2-chloroethyl) phosphate	5
381	1.388	1.15873	0.229271	Treening	Monoethanolamine	5
405	2.466	2.20773	0.258266	Treening	2-Chloro-1-fluoro-4-nitrobenzene	5
426	2.544	2.52622	0.0177781	Treening	o-Dinitrobenzene	5
427	1.549	1.7237	-0.174702	Treening	4,6-Dinitro-o-cresol	5
486	0.86	0.21408	0.64592	Treening	o-Chlorobenzonitrile	5
495	3.882	3.28881	0.593187	Treening	2,4-Bis(ethylamino)-6-methylthio-1,3,5-triazine <Simetryn>	5
531	0.117	0.515886	-0.398886	Treening	2,2,6,6-Tetramethylpiperidin-4-ol	5
541	1.193	1.23386	-0.0408579	Treening	3-Amino-4-chlorobenzoic acid	5
573	4.75	3.69657	1.05343	Treening	1-Nitropyrene	5
614	0.523	1.10367	-0.580667	Treening	2-Ethylhexyl hydrogen (2-ethylhexyl)phosphonate	5
660	2.659	2.09479	0.56421	Treening	2,3,4,4'-Tetrahydroxybenzophenone	5
8	1.247	1.5952	-0.348195	Valideerimine	N,N-Dimethylhydrazine	5
16	2.216	1.7834	0.432596	Valideerimine	Methylhydrazine	5
27	0.327	1.20173	-0.874726	Valideerimine	Salicylic acid	5
79	2.013	1.86081	0.15219	Valideerimine	Dibutyl phthalate	5
122	1.727	2.08153	-0.35453	Valideerimine	N,N-Diethylaniline	5
129	2.43	2.61938	-0.189376	Valideerimine	Phenothiazine	5
151	0.832	1.17837	-0.346372	Valideerimine	2,4-Diaminotoluene	5
179	2.635	3.15556	-0.520559	Valideerimine	m-Dinitrobenzene	5
225	0.977	1.04733	-0.0703309	Valideerimine	p-Anisidine	5
271	-0.503	0.949878	-1.45288	Valideerimine	Isobutyl acetate	5
289	0.552	1.52927	-0.977266	Valideerimine	3,3'-Thiodipropionic acid	5
299	4.019	2.43192	1.58708	Valideerimine	1-Mercaptooctane <n-Octylmercaptan>	5
322	1.489	1.01718	0.471824	Valideerimine	4-Amino-2-nitrophenol	5
334	0.258	0.658061	-0.400061	Valideerimine	4-Nitrotoluene-2-sulphonic acid	5
344	3.038	1.13787	1.90013	Valideerimine	4-Aminophenol	5
354	5.218	5.09256	0.125438	Valideerimine	Dimantine	5
410	1.439	2.24346	-0.80446	Valideerimine	Metaxylene hexafluoride	5
417	2.518	2.03668	0.481322	Valideerimine	1,8-Naphthylenediamine	5
421	0.497	0.991572	-0.494572	Valideerimine	4-Aminopyridine	5

Nr	log (1/EC_{50}) (eksp.)	log (1/EC_{50}) (ennust.)	log (1/EC_{50}) (vahe)	Valim	Nimetus	Klass
422	0.895	0.918239	-0.023239	Valideerimine	2-Aminopyridine	5
424	0.437	2.49367	-2.05667	Valideerimine	Ethyl trichloroacetate	5
446	-0.166	0.89138	-1.05738	Valideerimine	3-Aminophenol	5
457	0.055	1.83591	-1.78091	Valideerimine	Octanedinitrile	5
461	1.086	1.64659	-0.560589	Valideerimine	But-3-en-3-olide	5
510	0.687	1.59565	-0.908654	Valideerimine	m-Phenylenebis(methylamine)	5
516	3.319	2.68417	0.634831	Valideerimine	Dodecyldimethylamine oxide	5
522	0.417	0.722456	-0.305456	Valideerimine	1-Cyclohexene-1-carbonitrile	5
528	3.168	2.80433	0.363665	Valideerimine	Dibenzo[b,f]cyclohepten-1-one	5
543	2.858	1.27361	1.58439	Valideerimine	tert-Butyl 2-ethylperoxyhexanoate	5
558	2.743	3.33241	-0.589411	Valideerimine	Pentaethylenehexamine	5
629	1.566	1.43241	0.133593	Valideerimine	3,5-Di-tert-butylsalicylic acid	5
633	3.74	3.28381	0.456195	Valideerimine	Metribuzin	5
637	1.877	2.70463	-0.827631	Valideerimine	Propyzamide	5

[a] Ministry of the Environment of the Government of Japan, Results of aquatic toxicity tests of chemicals conducted by Ministry of the Environment in Japan (-March 2016), (2016). http://www.env.go.jp/en/chemi/sesaku/aquatic_Mar_2016.pdf viimati alla laetud 12.06.2017.
[b] Verhaar'i klass määratud käesolevas töös

Infoleht

Struktuurselt varieeruvate orgaaniliste ühendite toksilisuse kvantitatiivsed struktuur-aktiivsus sõltuvused vetikal *Pseudokirchneriella subcapitata*

Töös tuletati QSAR mudelid 342-le struktuurselt varieeruvale orgaanilisele ühendile. Modelleeriti toksilisust vetikale *Pseudokirchneriella Subcapitata*. Rakenduspiirkonnad olid määratud modifitseeritud Verhaar'i klassifikatsiooni reeglitega. Tuletati kokku viis QSAR mudelit, igale Verhaar'i klassile üks. Mudelite ennustusvõimet analüüsiti välise valideerimisega ning kõrvalekaldujad määrati ja rakenduspiirkondi analüüsiti Williams'i graafikute abiga.

Märksõnad: QSAR, toksilisus, vetikad, *Pseudokirchneriella Subcapitata*
CERCS koodid: P410 Teoreetiline ja kvantkeemia; P305 Keskkonnakeemia

Quantitative structure-activity relationships of the toxicity of structurally varying organic compounds towards algae *Pseudokirchneriella subcapitata*

In this work, QSAR models were derived for 342 structurally varying organic compounds. Toxicity towards algae *Pseudokirchneriella Subcapitata* was modelled. Applicability domains were determined by modified Verhaar classification rules. A total of five QSAR models, one for each Verhaar class, were derived. The models' quality of prediction was assessed using external validation, outliers were determined and applicability domains were analysed with the help of Williams plots.

Keywords: QSAR, toxicity, algae, *Pseudokirchneriella Subcapitata*
CERCS codes: P410 Theoretical chemistry, quantum chemistry; P305 Environmental chemistry

Lihtlitsents

Mina, Tanel-Sigmar Sildoja,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Struktuurselt varieeruvate orgaaniliste ühendite toksilisuse kvantitatiivsed struktuur-aktiivsus sõltuvused vetikal *Pseudokirchneriella subcapitata*“,

mille juhendaja on Uko Maran,

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 kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

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

Tartus, **28.05.2018**