

TARTU ÜLIKOOL

Loodus- ja täppisteaduste valdkond

Keemia instituut

Molekulaartehnoloogia õppetool

Iris Tamsalu

**Kvantitatiivsed struktuuri-aktiivsuse sõltuvused *Pseudokirchneriella subcapitata* kasvu  
pärssivatele orgaanilistele ühenditele**

Bakalaureusetöö (12 EAP)

Keemia

Juhendaja Uko Maran, PhD

Tartu 2023

## INFOLEHT

### **Kvantitatiivsed struktuuri-aktiivsuse sõltuvused *Pseudokirchneriella subcapitata* kasvu pärssivatele orgaanilistele ühenditele**

Kemikaalide toksilisuse hindamiseks kasutatakse üha enam ökotoksikoloogilistele testidele alternatiivseid *in silico* modelleerimismeetodeid, mis võimaldavad leida seoseid keemilise aine struktuuri ja selle bioloogilise aktiivsuse vahel. Töö eesmärk oli tuletada kvantitatiivsed struktuuri-aktiivsuse sõltuvuse (QSAR) mudelid vetika *Pseudokirchneriella subcapitata* kasvu pärssivatele orgaanilistele ühenditele. Kokku tuletati viis mõõdukate kuni väga heade statistiliste parameetrite ning ennustusvõimega mudelit. Neist neli mudelit on suunatud sarnase toimeviisiga ühendite klassidele, mis sisaldavad erineva toimemehhanismiga ühendeid, ning üks kogu andmeseeriale.

**Märksõnad:** *Pseudokirchneriella subcapitata*, QSAR, Verhaari klassifikatsioon

**CERCS:** P305 Keskkonnakeemia; P175 Informaatika, süsteemiteooria; P410 Teoreetiline ja kvantkeemia

### **Quantitative structure-activity relationships of organic compounds inhibiting the growth of *Pseudokirchneriella subcapitata***

*In silico* modeling as an alternative to ecotoxicological tests enables us to study the relationships between chemical structure and biological activity with little economic and environmental impact. The present work aimed to develop quantitative structure-activity relationship (QSAR) models for structurally diverse organic compounds that can predict the toxicity of chemicals to algae *Pseudokirchneriella subcapitata*. Altogether five models with proper to excellent statistical parameters and predicting power were presented, including models for sets of compounds with similar modes of action that include compounds with different mechanisms of action and one model for all the modeled compounds.

**Keywords:** *Pseudokirchneriella subcapitata*, QSAR, Verhaar classification

**CERCS:** P305 Environmental chemistry; P175 Informatics, systems theory; P410 Theoretical chemistry, quantum chemistry

# Sisukord

Sissejuhatus .....	4
1. Kirjanduse ülevaade .....	5
1.1. Testorganismid ökotoksikoloogias: <i>Pseudokirchneriella subcapitata</i> .....	5
1.2. Ökotoksikoloogilised testid.....	5
1.3. Toimeviisid: Verhaari klassifikatsioon .....	6
1.4. QSAR <i>Pseudokirchneriella subcapitata</i> .....	7
2. Andmed ja meetodika .....	11
2.1. Andmete analüüs, korrastamine ja klassifitseerimine .....	11
2.2. Struktuuride kirjeldamine.....	11
2.3. Kvantitatiivsete struktuuri-aktiivsuse sõltuvuste tuletamine.....	12
2.4. Valideerimine .....	13
2.5. Rakenduspiirkonna hindamine.....	14
3. Tulemused.....	15
3.1. Mudel inertsetele ühenditele .....	15
3.2. Mudel vähem inertsetele ühenditele.....	17
3.3. Mudel reaktsioonivõimelistele ühenditele .....	19
3.4. Mudel liigitamisele mittealluvatele ühenditele .....	21
3.5. Mudel kogu andmeseeriale.....	23
Kokkuvõte .....	26
Summary .....	27
Kasutatud kirjandus.....	28
Lisad .....	32

# Sissejuhatus

Maailmas toodetavate kemikaalide kogus ja mitmekesisus on kasvutrendis, eriti kiire kasv toimub areneva majandusega Aasia riikides. Ühinenud Rahvaste Organisatsiooni (ÜRO) Keskkonnaprogrammi 2019. aasta raporti järgi prognoositakse 2020. aastate jooksul tootmise kahekordistumist, mis tähendab, et 2030. aastal toodetakse maailmas ligikaudu 4,6 miljardit tonni erinevaid kemikaale. Jõudnud keskkonda, ohustavad kemikaalid muuhulgas inimese elukvaliteeti ja elurikkuse säilimist ökosüsteemides. [1]

Kemikaalide ohtlikkust hinnatakse ökotoksikoloogiliste testidega, kus üheks enim kasutatud testorganismiks on kemikaalidele tundlik mikrovetikas *Pseudokirchneriella subcapitata*. Primaartootjatena on vetikad ökosüsteemi seisundi varased bioindikaatorid, sest muutused vetikakoosluses ja kemikaalide bioakumuleerumine vetikates mõjutavad toiduahela kaudu kõrgemaid troofilisi tasemeid. [2]

Keskkonna reostumise probleem kajastub mitmetes rahvusvahelistes kokkulepetes, sealhulgas ÜRO säästva arengu eesmärkides (SDG) [3] ja Euroopa Liidu kemikaalide registreerimise, hindamise, autoriseerimise ja piiramise määruses (REACH) [4]. Üheks REACH-i eesmärgiks on edendada *in vivo* ja *in vitro* katsetele alternatiivsete *in silico* modelleerimismeetodite kasutamist. Andmebaasidesse nagu PubChem lisandub igal aastal miljonites uusi keemilisi ühendeid, kuid sissekandeid ühendite bioloogilise aktiivsuse kohta kümnetes kordades vähem [5]. Varasemate katseandmete *in silico* modelleerimine võimaldab molekuli struktuuri põhjal ennustada keemilise aine aktiivsust, näiteks toksilisust. Nii saab vähendada aja- ja ressursimahukate *in vivo* ja *in vitro* katsete hulka.

Bakalaureusetöö eesmärk on uurida ja kirjeldada keemilise aine struktuuri ja vetika *Pseudokirchneriella subcapitata* kasvu pärssiva mõju vahelisi seoseid ning koostada neid seoseid kirjeldavaid kvantitatiivseid struktuuri-aktiivsuse sõltuvusi (QSAR). Tuletatud QSAR mudelitega saab numbriliselt hinnata keemiliste ainete toksilisust, mille kohta eksperimentaalsed andmed puuduvad.

# 1. Kirjanduse ülevaade

## 1.1. Testorganismid ökotoksikoloogias: *Pseudokirchneriella subcapitata*

Peamiselt uuritakse ökotoksikoloogias viit taimegruppi: mikrovetikas, makrovetikas, sõnajalg-, sammal- ja õistaimed. Seejuures üle poole, 60% uuringutest on tehtud kemikaalidele tundlike ja laboratoorselt kergesti kasvatavate mikrovetikate kohta. Magevee mikrovetikatest levinuimad on *Pseudokirchneriella subcapitata*, *Chlorella vulgaris*, *Chlamydomonas reinhardtii*, *Chlorella pyrenoidosa* ja *Scenedesmus quadricauda*. [2]

*Pseudokirchneriella subcapitata* tuntud ka kui *Selenastrum capricornutum* ja *Raphidocelis subcapitata* on ökotoksikoloogilistes uuringutes enim kasutatud mikrovetikas [2]. *Pseudokirchneriella subcapitata* on mageveeline planktoni koosseisu kuuluv rohevetikas. See ainurakne vetikas moodustab paljunemise ajal läbipaistva lima abil mõnerakulisi kobarasse kogunenud rühmi. Vetika pikkus on 4,8–10,8 µm ja laius 1,6–4,4 µm. Vetika kaardunud keha teeb tavaliselt ühe pöörde, vanemad vetikad kuni poolteist pööret. [6,7]

## 1.2. Ökotoksikoloogilised testid

Ökotoksikoloogiliste testide läbiviimiseks on välja töötatud standardiseeritud juhised, mis tagavad tulemuste usaldusväärse võrdlemise. *Pseudokirchneriella subcapitata* on testorganismina kasutusel nii Majanduskoostöö ja Arengu Organisatsiooni (OECD) [8] kui ka Rahvusvahelise Standardiseerimise Organisatsiooni (ISO) [9] testides.

OECD testis (nr 201) uuritakse keemilise aine mõju magevee mikrovetikale ja tsüanobakterile. Test viiakse läbi 72 h vältel eksponentsiaalses kasvufaasis testorganismiga, seejuures on tagatud piisav toitainevaru, pidev valgus ja sobiv temperatuur vahemikus 21–24 °C. Muutusi vetika kasvukiiruses (võrrandid 1–2) analüüsitakse 24, 48 ja 72 h möödudes mõne biomassi määramise meetodiga, nagu rakkude loendamine, fluorestsentsi või optilise tiheduse mõõtmine. Koostatakse doosi-toime graafik, millelt saadakse  $EC_{50}$  väärtus – uuritava aine kontsentratsioon, mis pärsib vetika kasvu 50% võrra. Tehakse kolm korduskatset ja igale korduskatsele negatiivne kontroll. [8]

**Võrrand 1.** Kasvukiiruse ( $\mu_{i-j}$ ) arvutamine, kui  $X_i$  ja  $X_j$  on vetika biomassid ajahetkedel  $t_i$  ja  $t_j$

$$\mu_{i-j} = \frac{\ln X_j - \ln X_i}{t_j - t_i}$$

**Võrrand 2.** Kasvukiiruse muutus (%I), kui  $\mu_T$  on vetika kasvukiirus eksponeerituna uuritavale ainele ja  $\mu_K$  on kontrollkatse kasvukiirus

$$\%I = \frac{\mu_K - \mu_T}{\mu_K} \cdot 100\%$$

ISO testis (nr 8692) uuritakse puhtas vees või reovees leiduvate keemiliste ainete ja nende segude mikrovetika kasvu pärssivat mõju. Kasutatakse mageveelisi vetikaid *Desmodesmus subspicatus* ja *Pseudokirchneriella subcapitata*. Üldiselt on ISO 8692 test sarnane OECD 201 ülesehitusega. [9]

### 1.3. Toimeviisid: Verhaari klassifikatsioon

Keemiliste ühendite toksilisuse edukaks modelleerimiseks on oluline, et ühendid oleksid sarnase struktuuri ja toimeviisiga. Ökotoxikoloogias kasutatakse tihti Verhaari toimeviiside klassifikatsiooni skeemi [10], mis jagab orgaanilised ühendid struktuuri alusel nelja klassi: inertsed (1. klass), vähem inertsed (2. klass), reaktsioonivõimelised (3. klass) ja spetsiifilise toimemehhanismiga (4. klass) ühendid. Iga kõrgema Verhaari klassi ühendite toksilisus on eelnevast klassist suurem, samuti suureneb iga klassiga struktuurne heterogeensus. Vastavalt uutele teadmistele ühendi struktuuri ja toimemehhanismi seostest on klassifitseerimise reegleid täiendatud ja kasutusele võetud modifitseeritud Verhaari klassifikatsioon [11]. Hetkel on kõige levinumad Verhaari skeemi rakendusi sisaldavad programmid OECD QSAR Toolbox [12] ja Toxtree [13].

Verhaari 1.–3. klassi kuuluvad ühendid peavalt vastama esmalt kolmele tingimusele: (i) molekul sisaldab vaid C, H, N, O, S ja halogeeni (v.a I) aatomeid, (ii) logaritm oktanooli-vee jaotuskoefitsiendist ( $\log K_{OW}$ ) on vahemikus 0–6 ning (iii) molekulmass ei ületa 600 Daltonit. Neljandasse klassi ei määrata ühendit erinevalt teistest Verhaari klassidest mitte struktuuri, vaid olemasoleva info alusel keemilise ühendi või ühendite grupi toimemehhanismi kohta. Viienda klassi moodustavad ühendid, mida ei suudetud Verhaari skeemi alusel klassifitseerida. [10]

Verhaari esimesse, inertsete ühendite klassi kuuluvad ühendid, mis koosnevad peamiselt C ja H aatomitest, kuid võivad teatud tingimustel sisaldada ka O, N või halogeenide aatomeid. Esimesse klassi kuuluvad näiteks alifaatsed alkoholid ja ketoonid. Verhaari teise, vähem inertsete ühendite klassi kuuluvad head vesiniksideme doonorid, nagu fenoolid ja aniliinid, millel võib olla alküül-, kloro- ja nitrorühmi. Lisaks kuuluvad vähem inertsete ühendite hulka kloro- ja/või

alküülrühmadega püridiinid ja aromaatsed ühe nitrorühmaga ühendid ning primaarsed alküülamiinid. [10]

Inertsed ja vähem inertsed keemilised ühendid on organismile anesteetilise toimega, mida põhjustab ühendite akumulatsioon rakumembraanis [14]. Anesteetilise toimega ühendite aktiivsus on positiivses korrelatsioonis log  $K_{OW}$ -ga, mis kirjeldab ühendi baastaseme toksilisust. Siiski on ainult log  $K_{OW}$  kasutamisel tendents hinnata ühendid tegelikult vähem toksiliseks, sest arvesse ei võeta muid võimalikke elektronstruktuurist tulenevaid toimemehhanisme. [10,15]

Kolmandasse, reaktsioonivõimeliste ühendite klassi kuuluvad elektrofiilsete rühmadega molekulid või metabolismi käigus elektrofiilideks muutuvad molekulid, mis atakeerivad bioloogiliste makromolekulide nukleofiilseid osi [16]. Reaktsioonivõimeline ühend sisaldab kaksik- või kolmiksideme  $\alpha$ -asendis head lahkuvat või polariseeritavat funktsionaalrühma, lämmastik-lämmastik ühe- või mitmekordset sidet, on kolmeatomiline heterotsükkel (epoksiid, asardiin) või aktiveeritud nitril. Samuti loetakse kolmandasse klassi ühendeid, mis sisaldavad teatud funktsionaalrühmi, sealhulgas happe anhüdriidid ja haliidid, keteenid, disulfiidid ning aldehüüdid. Sarnaselt 3. klassi ühenditele on ka 4. klassi ühendid baastasemest kõrgema aktiivsusega, ent toimivad spetsiifilise mehhanismi järgi, näiteks seostuvad retseptoriga. [10]

#### **1.4. QSAR *Pseudokirchneriella subcapitata***

Enamik vahemikus 2006–2023 *Pseudokirchneriella subcapitata* avaldatud QSAR mudeleid kirjeldavad orgaaniliste ainete akuutset toksilisust (tabelid 1 ja 2, põhjalikum tabel 1. lisas). Lisaks on mudeldatud kroonilist toksilisust [17,18], samuti nanoosakeste [19] ja ionvedelike [20] toksilisust. Kasutusala põhiseid mudeleid on tehtud kosmeetika koostisosadele [21], pestitsiididele [22] ning ravimitele [23]. Ravimitest eraldi on modelleeritud antidepressante [24] ja antibiootikume [25]. Lisaks on arendatud liikideüleseid mudeleid, näiteks QSAR mudel rohevetikatele [26], ja tehtud kvantitatiivseid toksilisuse-toksilisuse sõltuvuse (QTTR) mudeleid, mis teisendavad keemilise ühendi toksilisuse ühelt vetikaliigilt teisele [27].

#### **Mudelid sarnase struktuuri või toimeviisiga ühenditele**

Üldiselt saab *Pseudokirchneriella subcapitata* suhtes toksiliste ühendite andmeseeriadele tehtud QSAR mudeleid jagada kahte rühma: (i) mudelid sarnase struktuuri või toimeviisiga ühenditele (tabel 1) ja (ii) mudelid suure struktuurse varieeruvusega ühenditele (tabel 2). Tabelis 1 on näha,

et peaaegu kõik sarnase struktuuri või toimeviisiga ühenditele tehtud mudelid on lineaarsed või multilineaarsed regressioonid (MLR).

Verhaari 1. klassi ehk inertsete ühendite toksilisus on positiivses korrelatsioonis ühendi  $\log K_{OW}$  väärtusega ning enamasti on juba ühe molekulaarse struktuuritunnusega ehk molekulaartunnusega (ingl *molecular descriptor*) mudel heade statistiliste parameetritega ( $R^2 = 0,76-0,95$ ) [28–31]. Mudelites on lisaks  $\log K_{OW}$ -le molekulaartunnuseid, mis sisaldavad informatsiooni molekuli polariseeritavuse (*MATS7i*, *MATS3p*), topoloogilise polaarse pindala (*TPSA*) ja pinna negatiivse potentsiaali ( $V_s$ ) kohta [32]. Nii inertseid kui ka vähem inertseid ühendeid sisaldav mudel sidus toksilisuse molekulmassi (*MW*) ning molekuli tekkeentalpia ja aatomite arvu suhtega ( $DH_f/\#atoms$ ), seejuures suurema tekkeentalpiaga ühendid on ebapüsivamad ja toksilisemad [30].

Tabelis 1 kirjeldavad Verhaari 2. klassi ehk vähem inertsete ühendite struktuuri-aktiivsuse sõltuvusi ka mudelid, mille sisendiks oli mõni vähem inertsete ühendite alla kuuluv ainegrupp, näiteks klorofenoolid, bensoehapped ja aniliinid. Lisaks baastoksilisust kirjeldavale  $\log K_{OW}$ -le on vähem inertsetele ühenditele tehtud mudelites elektroonseid efekte kirjeldavaid ja asendusrühmi loendavaid molekulaartunnuseid. Sage molekulaartunnus vähem inertsete ühendite mudelites on *ELUMO*, seejuures väiksema *ELUMO* väärtusega ühendid on paremad vesiniksideme doonorid ja organismile toksilisemad [30,33]. Sarnast efekti kirjeldab happe dissotsiatsioonikonstant ( $pK_a$ ) klorofenoolidele tehtud mudelis [34].

Keemilise ühendi toksilisus võib sõltuda asendusrühmade iseloomust ja asukohast. Ühendi toksilisust suurendab halogeeniaatomite lisamine. Suurema toksilisusega halogeenitud bensoehapetest on broomasendatud bensoehapped ning meta- ja paraasendustustega bensoehapped. Hüdroksübensoehapete puhul annavad  $\log K_{OW}$  ja hüdroksüülrühmade arv ( $N_{OH}$ ) väga hea korrelatsiooni. [35]

Kui mudeli ennustatud ühendi toksilisus jääb oluliselt alla eksperimentaalsele väärtustele, võib tegu olla reaktsioonivõimelise või spetsiifilise toimemehhanismiga ühendiga. Kui tertsiarsed amiinid on väga heas korrelatsioonis  $\log K_{OW}$ -ga ( $R^2 = 0,92$ , vabaliige =  $-1,81$ ), siis primaarsete ja sekundaarsete amiinid toksilisus on baastasemest kõrgem ning ainult  $\log K_{OW}$ -st ei piisa nende täieliku toksilisuse kirjeldamiseks ( $R^2 = 0,84$ , vabaliige =  $-4,12$ ) [24]. Mudel kolmeteistkümnele antibiootikumile tõi välja, et varieeruvuse antibiootikumide toksilisuses määravad kõige enam elektroonsed efektid (*ELUMO*, 60%), siis stabiilsus (tekkeentalpia,  $H_f$ , 19%) ja läbitavus ( $\log K_{OW}$ ,

12%) [25]. Reaktsioonivõimeliste ja spetsiifilise toimemehhanismiga ühendite aktiivsuse kirjeldamiseks on kasutatud veel elektronegatiivsust ( $GATS1e$ ), polariseeritavust ( $SpMin1\_Bh(p)$ ,  $\alpha$ ), van der Waalsi raadiust ( $R6v$ ) ja ionisatsioonipotentsiaali ( $Mor32i$ ,  $GATS3i$ ) iseloomustavaid molekulaartunnuseid [32].

**Tabel 1.** Mudelid sarnase struktuuri või toimeviisiga ühenditele

Ühendid	$N_{treening} / N_{test}$	Meetod	Molekulaartunnused	$R^2$	$Q^2$	$\sigma$
1. Klorofenoolid [34]	8	LR	$\log P, pK_a$	0,90–0,96	-	-
2. Nitriilid [36]	12	(M)LR	$E_{lumo}, \log K_{ow}$	0,85	0,71–0,78	-
3. Verhaari 1. klass [28]	26	LR	$\log K_{ow}$	0,94	0,932, 0,933	0,325, 0,332
4. Verhaari 2. klass [33]	18–20	(M)LR	$\log K_{ow}, E_{lumo}$	0,750–0,916	0,715– 0,866	0,328– 0,499
5. Bensoehappe derivaadid [35]	8–20	(M)LR	$\log K_{ow}, NOH$	0,921–0,998	0,896– 0,997	0,03–0,23
6. Verhaari 1. ja 2. klass, halogeenitud nitriilid [29]	6-53	(M)LR	$\log K_{ow}, E_{lumo}$	0,766–0,897	0,572 – 0,857	
7. 1°, 2° ja 3° propargüülalkoholid [37]	7–15	(M)LR	$\log K_{ow}, E_{lumo}$	0,76–0,97	0,69–0,94	0,16–0,50
8. Verhaari 1. ja 2. klass [30]	50–87 / 21	(M)LR	$\log K_{ow-mix}, DH_f/\#atoms, MW, E_{lumo}$	0,8532– 0,9469	0,8449– 0,9426	0,0887– 0,1829
9. Aniliinid ja fenoolid [38]	13–45	aug-MIA	-	0,84	0,60	-
10. Verhaari 1. ja 2. klass [31]	29–65	(M)LR	$\log P(1), HF$	0,71–0,86	0,65–0,84	0,35–0,52
11. Antibiootikumid [25]	13	MLR	$E_{lumo}, H_f, \log K_{ow}$	0,91	-	0,52
12. Kõik Verhaari klassid [32]	8–73 / 4–28	(M)LR	$\log K_{ow}, MATS7i, t^*10^3, MATS3p, TPSA(Tot), \bar{V}_s, V_s;min, qc, G3p, Mor32i, V_s;max, GATS1e, SpMin1\_Bh(p), V_s+, R6v, GATS3i, GGI9, DLS\_02, SpMAD\_AEA(dm), Mor10e, Gli, E_homo, HATS1i$	0,613–0,925	0,550– 0,960	-
13. 1°, 2° ja 3° amiinid [24]	4–5	LR	$\log K_{ow}$	0,84–0,92	-	0,15–0,31

Lühendite selgitused:  $R^2$  – determinatsioonikordaja,  $Q^2$  – ristvalideerimiskordaja,  $\sigma$  – standardhälve,  $N_{treening}$  – treeningseeria ühendite arv,  $N_{test}$  – testseeria ühendite arv; põhjalikum tabel on lisas 1

## Mudelid suure struktuurse varieeruvusega ühenditele

Alates 2014. aastast on üha enam mudeldatud struktuurselt heterogeensemaid ja suuremaid, kuni 600 ühendiga andmeseeriaid (tabel 2). Mitmed suure struktuurse varieeruvusega ühenditele tehtud mudelitest on MLR-id, mis on tuletatud vähimruutude meetodil (PLS või OLS) [39,40], ning mitmes MLR mudelis on deskriptorite valmisel kasutatud geneetilisi algoritme (GA ja GFA) [21–

23,41,42]. Samuti on kasutatud masinõppemeetoditest tugivektormasinat (SVM) ja tehislikke närvivõrke (ANN) [39,43] ning ansambelõppemeetoditest otsustusmetsa (EL-DFT) ja otsustuspuude võimendamist (EL-DTB) [44]. Klassifitseerimismudelites (toksiline või mitte) on kasutatud lineaarset diskriminantanalüüsi (LDA) [40].

Mudelid suure struktuurse varieeruvusega ühenditele sisaldavad võrreldes sarnase struktuuri ja toimeviisiga ühenditele tehtud mudelitega rohkem molekulaartunnuseid. Mudelitesse valitud molekulaartunnused seovad keemilise ühendi toksilisuse muuhulgas molekuli suuruse (*MW* ja *CrippenMR*) [40,41,43] ja elektroonsete efektidega (*TopoPSA*, *nHBD<sub>on</sub>*, *nHB<sub>int2</sub>*) [22,44]. Samuti on levinud molekulaartunnus  $\log K_{ow}$ , sealhulgas erinevad  $\log K_{ow}$  versioonid: *ALOGP2*, *MLOGP* ja *XLogP* [23,40–42,44].

**Tabel 2.** Mudelid suure struktuurse varieeruvusega ühenditele

Ühendid	$N_{treening} / N_{test}$	Meetod	Molekulaartunnused	$R^2$	$Q^2$	$\sigma$
14. Orgaanilised ained [39]	655 / 218	MLR, PLS, BPANN, SVM, LSSVM, RF, GP		0,798–0,927	0,734–0,830	-
15. Orgaanilised ained [44]	68–425 / 12–65	(EL)-DTB, (EL)-DTF	<i>VP-2, MDEC-22, BCUT<sub>p-II</sub>, WL-2U, WL-3U, PPSA-1, PNSA-1, XlogP, nAtomP, SP-1, MDEC-23, MDEC-33, ATSm5, ATSp5, TopoPSA, nHBD<sub>on</sub></i>	0,905–0,918	-	0,89–0,98
16. Orgaanilised ained [40]	74 / 31	MLR-PLS, LDA	$x^2\chi$ , density, $[\sum a]_p / \sum a$ , <i>Atype_N_74</i> , <i>Atype_C_36</i> , $\log K_{ow}$ , <i>S_aaaC</i> , <i>Atype_C_24</i> , $x^2\chi_{vs}$ , <i>MW</i>	0,774–0,782	0,708–0,782	-
17. Kosmeetika [21]	20 / 10	GFA-PLS	<i>VR3_Dt, GATS6s, ETA_AlphaP, AATS0s</i>	0,885	0,811	-
18. Orgaanilised ained [41]	251 / 83	GA-PLS	<i>CrippenMR, MLOGP, B06[C-N], B05[C-Cl], B02[N-S], nSO2OH, logKow, H-051, F02[N-S], MW, X3A, B10[C-N], F01[C-C], F04[C-S]</i>	0,70–0,72	0,68–0,70	-
19. Ravimid [23]	53 / 16	GA-PLS	<i>F08[O-O], UNIP, B08[C-C], PW4, ALOGP2, F08[O-O], nR=C<sub>s</sub>, F06[C-Cl], CATS2D_04_DD, piPC1, piPC4, piPC8</i>	0,69–0,74	0,59–0,65	-
20. Orgaanilised ained [43]	167–334 / 167	MLR, SVM	<i>CrippenMR, MHYD, F02[N-S], B05[C-Cl], CCCN, R3m, BLTF96</i>	0,455–0,760	-	-
21. Orgaanilised ained [42]	217 / 54	GA-MLR	<i>nar-oCpplus, fnotringNsp3C1B, fClaroC5B, nTRing, XLogP, SubFP96, fsp2Csp3C8B</i>	0,718	0,700	-
22. Orgaanilised ained [45]	113–118 / 52–85	CORAL		0,8341–0,8866	0,8289–0,8826	0,533–0,683
23. Pestitsiidid [22]	163–169 / 70–76	GA-OLS-MLR	<i>MLFER_L, ExtFP300, ASP-4, GATS3m, ExtFP29, minHB<sub>int2</sub>, ExtFP29</i>	0,651–0,665	0,616–0,635	-

Lühendite selgitused:  $R^2$  – determinatsioonikordaja  $Q^2$  – ristvalideerimiskordaja,  $\sigma$  – standardhälve,  $N_{treening}$  – treeningseeria ühendite arv,  $N_{test}$  – testseeria ühendite arv; põhjalikum tabel on lisas 1

## 2. Andmed ja meetodika

### 2.1. Andmete analüüs, korrastamine ja klassifitseerimine

Modelleerimiseks kasutati Kusk et al. avaldatud eksperimentaalsete andmete seeriat [46], mis kirjeldab 426 ühendi pärssivat mõju *Pseudokirchneriella subcapitata* kasvule. Ökotoksikoloogilised testid olid teostatud vastavalt OECD 201 [8] ja ISO 8692 [9] juhistele. Ühendi toksilisust hinnati kontsentratsiooniga, mis pärssis *Pseudokirchneriella subcapitata* kasvu 50% ( $EC_{50}$ ).

Andmeseeria keemilist koostist analüüsiti ja korrastati nii, et see oleks modelleeritav. Korrastamise käigus eemaldati ühendid, millel  $EC_{50}$  väärtus puudub või on märgitud ebatäpseks ning ühendid, mille CAS number ja nimetus ei lange kokku ehk struktuur ei ole üheselt tuvastatav (71 tk). Kõrvale jäeti püsiva laenguga ühendid (29 tk), sest neutraalse vormi puudumine põhjustab erinevuse molekulaartunnuste väärtustes. Kõrvale jäi ka kaaliumdikromaat, mis on ainus anorgaaniline aine andmeseerias. Samuti jäid välja polüfluoritud ühendid (5 tk), kus kõik või enamik vesinikke on asendatud fluoriaatomitega, ja polüklooritud ühendid (4 tk), kus kõik vesinikud on asendatud klooriaatomitega, kuivõrd need käituvad teistest ühenditest erinevalt ja põhjustavad suurt variatsiooni molekulaartunnuste väärtustes. Ühendid, millele  $EC_{50}$  väärtus saadi ekstrapoleerimise teel või mille veeslahustuvus oli väiksem  $EC_{50}$  väärtusest, jäeti sisse, kuid nende käitumist jälgiti modelleerimisel erilise tähelepanuga.

Modelleerimisel kasutati 316 orgaanilist ühendit (lisa 2), mille  $EC_{50}$  väärtused jäid vahemikku 0,00026–38280 mg/l.  $EC_{50}$  kontsentratsioonid teisendati millimolaarseteks (mM) ning võeti neist negatiivne logaritm ( $pEC_{50}$ ). Ühendite kahedimensionaalsed (2D) ja kolmedimensionaalsed (3D) geomeetriad laeti alla SDF-toimikutena PubChemist [47]. Kõik ühendid jagati rakenduses Toxtree 2.6.13 [13] modifitseeritud Verhaari toimeviiside klassifikatsiooni [11] põhjal klassidesse.

### 2.2. Struktuuride kirjeldamine

Modelleerimisel katsetati nii 2D kui ka 3D molekulaartunnustega. 3D molekulaartunnuste arvutamise esimeseks etapiks oli geomeetria optimeerimine. Geomeetria optimeerimiseks kasutati poolempiirilist kvantkeemia arvutuste tarkvara MOPAC 6.0 [48]. Molekulide geomeetria optimeerimise algoritmi piirtingimustest lähtudes nummerdati vajadusel aatomid Z-matriksis

Moldeni [49] abil ümber. Geomeetriad optimeeriti kolmes etapis ja molekulide elektronstruktuuride lähteolekud defineeriti AM1 (Austin Model 1) poolempiirilise parameterisatsiooni meetodiga [50]. Amiidsidemega molekulidele lisati märksõna NOMM, mis jätab arvestamata korrektsiooni, mis takistab amiidsideme rotatsiooni. 3D molekulaartunnused arvutati optimeeritud geomeetriaatele CODESSA PRO tarkvaraga [51].

2D molekulaarsete struktuuritunnuste arvutamiseks genereeriti SDF-toimikutest kanoonilised SMILESid, milleks kasutati OpenBabeli versiooni 3.1.1 [52]. 2D molekulaartunnused arvutati PaDEL-Descriptori tarkvara [53] ning eraldi arvutati juurde *XlogP3* [54]. Molekulaartunnustega tutvumisel kasutati lisaks PaDEL-Descriptori molekulaartunnuste selgitustele ka Todeschini ja Consonni molekulaartunnuste käsiraamatut [55].

### 2.3. Kvantitatiivsete struktuuri-aktiivsuse sõltuvuste tuletamine

QSAR modelleerimisel tuletatakse kvantitatiivne sõltuvus keemiliste ühendite struktuuri kirjeldavate molekulaartunnuste ja eksperimentaalselt saadud bioloogilise aktiivsuse või füüsikalise suuruse vahel. Nagu kirjeldatud kirjanduse ülevaates (peatükk 1.4), siis on modelleerimiseks mitmeid meetodeid, kus erineb ka QSAR mudelite matemaatiline kuju. Enim levinud mudeli tüüp on multilineaarne regressioon (MLR, võrrand 3). [56]

**Võrrand 3.** Multilineaarse regressiooni üldkuju, kus  $y$  modelleeritav suurus ( $pEC_{50}$ ),  $x_1-x_n$  on molekulaartunnuste väärtused ning  $a_0-a_n$  on molekulaartunnustele vastavad koefitsiendid

$$y = a_0 + a_1x_1 + \dots + a_nx_n$$

Selles töös tuletatud QSAR mudelite modelleeritavaks suuruseks on orgaaniliste ühendite *Pseudokirchneriella subcapitata* kasvu pärssiv toime, mida kirjeldab  $pEC_{50}$ . Sõltumatute muutujatena katsetati nii 2D kui ka 3D molekulaartunnuseid. 3D stuktuuride toimikud ja PaDEL-Descriptoriga [53] arvutatud 2D molekulaartunnused teisendati sobivasse vormi ja kanti üle CODESSA PRO-sse [51]. Molekulaartunnuste ja  $pEC_{50}$  väärtuste vahelised seosed esitati multilineaarse regressioonina, mille tuletamiseks kasutati CODESSA PROs rakendatud parima multilineaarse regressiooni meetodit (BMLR) [51,56]. 2D ja 3D mudelite statistiliste parameetrite võrdluses ilmnis, et parema mudeli andsid 2D molekulaartunnused ning seetõttu jätkati modelleerimist 2D molekulaartunnustega.

BMLR on multilineaarse regressiooni tuletamise meetod, mis seisneb astemeliselt regressioonivõrrandisse molekulaartunnuste valimises. BMLR meetod sisaldab endas

molekulaartunnuste analüüsi, mille käigus jäetakse välja puuduva molekulaartunnusega ühendid. Samuti jäetakse välja molekulaartunnused, mis mudeldavate ühendite seas oluliselt ei erine ehk on madala variatsiooniga. Seejärel leitakse kõik ortogonaalsed molekulaartunnuste paarid, mis vastavad determinatsioonikordaja tingimusele  $R_{ij} < R_{min}$  ( $R_{min}^2 = 0,1$ ). Ortogonaalsete molekulaartunnuste paare, mis annavad parima determinatsioonikordaja ( $R^2$ ) väärtusega kaheparameetrilise regressiooni, kasutatakse kõrgemat järku regressioonide tuletamisel. Iga järgmise molekulaartunnuse lisamisel tehakse mitte-kollineaarsuse kontroll. Kui Fisheri kriteeriumi ( $F$ ) või ristvalideerimiskordaja ( $Q^2$ ) väärtus molekulaartunnuse lisamisel ei parane, on lõpptulemuseks eelmise astme parim regressioon. Kui paranevad, jätkatakse parimate  $R^2$  väärtusega regressioonidega. Lõpuks väljastatakse parima  $F$  ja  $Q^2$  väärtustega regressioonid. [56]

## 2.4. Valideerimine

Tuletatud mudelite kvaliteeti hinnati nii mudelisiseste kui ka -välise meetoditega. Mudelisiseseks valideerimiseks kasutati enim levinud statistilisi parameetreid.  $R^2$  (võrrand 4) kirjeldab, kui suur osa sõltumatu muutuja variatsioonist on kirjeldatud sõltuva muutuja variatsiooniga.  $R^2$  väärtused jäävad vahemikku  $0 \leq R^2 \leq 1$  ning suurem  $R^2$  väärtus tähendab tugevamat muutujatevahelist lineaarset sõltuvust. Lisaks kasutati standardhälvet ( $\sigma$ , võrrand 5), et kirjeldada andmepunktide hajuvust aritmeetilisest keskmisest. Fisheri kriteeriumi ja Studenti t-testi abil hinnati molekulaartunnuste olulisust mudelis. [56]

**Võrrandid 4 ja 5.** Võrrand 4 kirjeldab determinatsioonikordaja ( $R^2$ ) ja võrrand 5 standardhälbe ( $\sigma$ ) arvutamist, kus  $n$  on treeningseeria ühendite arv,  $y_{i, eksp}$  ja  $y_{i, arv}$  on  $i$ -nda ühendi eksperimentaalne ja arvutatud  $pEC_{50}$  väärtus ning  $\bar{y}_{eksp}$  ja  $\bar{y}_{arv}$  on eksperimentaalsete ja arvutatud  $pEC_{50}$  väärtuste aritmeetilised keskmised.

$$R^2 = \frac{\sum_{i=1}^n (y_{i, arv} - \bar{y}_{arv})^2}{\sum_{i=1}^n (y_{i, eksp} - \bar{y}_{eksp})^2} \quad (4)$$

$$\sigma = \sqrt{\sum_{i=1}^n \frac{(y_{i, arv} - y_{i, eksp})^2}{n - 1}} \quad (5)$$

Mudeli ennustusvõimet iseloomustab  $Q^2$ , mille arvutamiseks jäetakse mudeli tuletamisel üks andmepunkt välja ning saadud mudeliga ennustatakse välja jäetud andmepunkti väärtus ja jääkliige. Nii tuletatakse kindel arv mudeleid ning arvutatakse jääkliikmete ruutude summa ( $PRESS$ ) (võrrand 6), mille kaudu saab arvutada  $Q^2$  väärtuse (võrrand 7). [56]

**Võrrandid 6 ja 7.** Võrrand 6 kirjeldab jääkliikmete ruutude summa (*PRESS*) ja võrrand 7 ristvalideerimiskordaja ( $Q^2$ ) arvutamist, kus  $y_{i; eksp}$  ja  $y_{i; arv}$  on *i*-nda ühendi eksperimentaalne ja arvutatud  $pEC_{50}$  väärtus ning  $\bar{y}_{eksp}$  on eksperimentaalsete  $pEC_{50}$  väärtuste aritmeetiline keskmine.

$$PRESS = \sum_{i=1}^n (y_{i; arv} - y_{i; eksp})^2 \quad (6) \quad Q^2 = 1 - \frac{PRESS}{\sum_{i=1}^n (y_{i; eksp} - \bar{y}_{eksp})^2} \quad (7)$$

Mudeliväliseks valideerimiseks moodustati igale mudelile valideerimisseeria. Selleks järjestati ühendid  $pEC_{50}$  kasvu järjekorras, loeti viieks ja iga teine (Verhaari 1.–2. ja 5. klass) või kolmas (Verhaari 3. klass ja kogu andmeseeria) ühend võeti valideerimisseeriasse. Saadud mudelitega ennustati valideerimisseeria ühendite  $pEC_{50}$  väärtusi.

## 2.5. Rakenduspiirkonna hindamine

Esiteks on Verhaari klassidele tehtud mudelite rakenduspiirkonnad määratud modelleerimisel kasutatud ühendite Verhaari klassiga. See tähendab, usaldusväärseks keemilise ühendi  $pEC_{50}$  väärtuse arvutamiseks peab esmalt määrama ühendi Verhaari klassi.

Teiseks määrati mudelite rakenduspiirkonnad nn Williamsi graafikuga. Graafiku *x*-teljel on ühendite omapära väärtused, mis on määratud mütsi maatriksi ( $H = X(X'X)^{-1}X'$ , kus *X* on treeningseeria ühendite molekulaartunnuste maatriks) peadiagonaali elementidega ( $h_{ii}$ ).  $h_{ii}$  on ühendi molekulaartunnuse väärtuse kaugus kõigi ühendite keskmisest tasemest. Kui ühendi omapära on kriitilisest omapärast (võrrand 9) suurem, ei kuulu ühend enam mudeli rakenduspiirkonda ja ennustatud tulemus ei pruugi olla usaldusväärne. Graafiku *y*-teljel on standardiseeritud ennustuserinevused (võrrand 10). Standardiseeritud ennustuserinevuse kriitiliseks väärtuseks võeti  $3\sigma$  ja  $-3\sigma$ , sest normaaljaotuse empiirilise reegli järgi peaks vahemikku  $[\mu - 3\sigma ; \mu + 3\sigma]$  jääma peaaegu kõik, 99% katsepunktidest. Seega võib ühend mudeli rakenduspiirkonnast välja jääda suure omapära, ebatäpse ennustuse või mõlema eelneva põhjuse tõttu. [57,58]

**Võrrand 9.** Kriitilise omapära ( $h^*$ ) väärtuse arvutamine, kus *k* on mudeli molekulaartunnuste arv ja *n* on treeningseeria ühendite arv

$$h^* = \frac{3(k+1)}{n}$$

**Võrrand 10.** Standardiseeritud ennustuserinevuse arvutamine, kus  $y_{eksp}$  ja  $y_{arv}$  on vastavalt eksperimentaalne ja arvutuslikult leitud tulemus ( $pEC_{50}$ ) ning  $\sigma$  on treeningseeria standardhälve

$$\text{standardiseeritud ennustuserinevus} = \frac{y_{eksp} - y_{arv}}{\sigma}$$

### 3. Tulemused

QSAR mudelid *Pseudokirchneriella subcapitata* kasvu pärssivatele ühenditele tuletati nii kogu andmeseeriale kui ka Verhaari klassidele eraldi. Verhaari 4. klassi mudeli analüüsist loobuti, kuna 4. klassi kuulus andmeseeriast vaid kolm ühendit. Keemilise ühendi struktuuri ja aktiivsuse vaheliste seoste kirjeldamiseks analüüsiti mudelitesse valitud molekulaartunnuseid. Lisaks hinnati mudelite rakenduspiirkondi ning analüüsiti kõrvalekaldujaid.

Sobiva molekulaartunnuste arvu valimisel jälgiti  $R^2$  ja  $Q^2$  väärtuste muutust molekulaartunnuste lisamisel ja reeglina lõpetati molekulaartunnuste lisamine, kui  $R^2$  väärtus paranes alla 0,02 ühiku. Lisaks jälgiti molekulaartunnuste valimisel, et mudelil poleks palju suure ennustuserinevusega ühendeid. Kõigist mudelitest jäeti välja molekulaartunnused, mille väärtused ühel või mitmel ühendil erinesid oluliselt ülejäänud ühendite keskmisest tasemest.

#### 3.1. Mudel inertsetele ühenditele

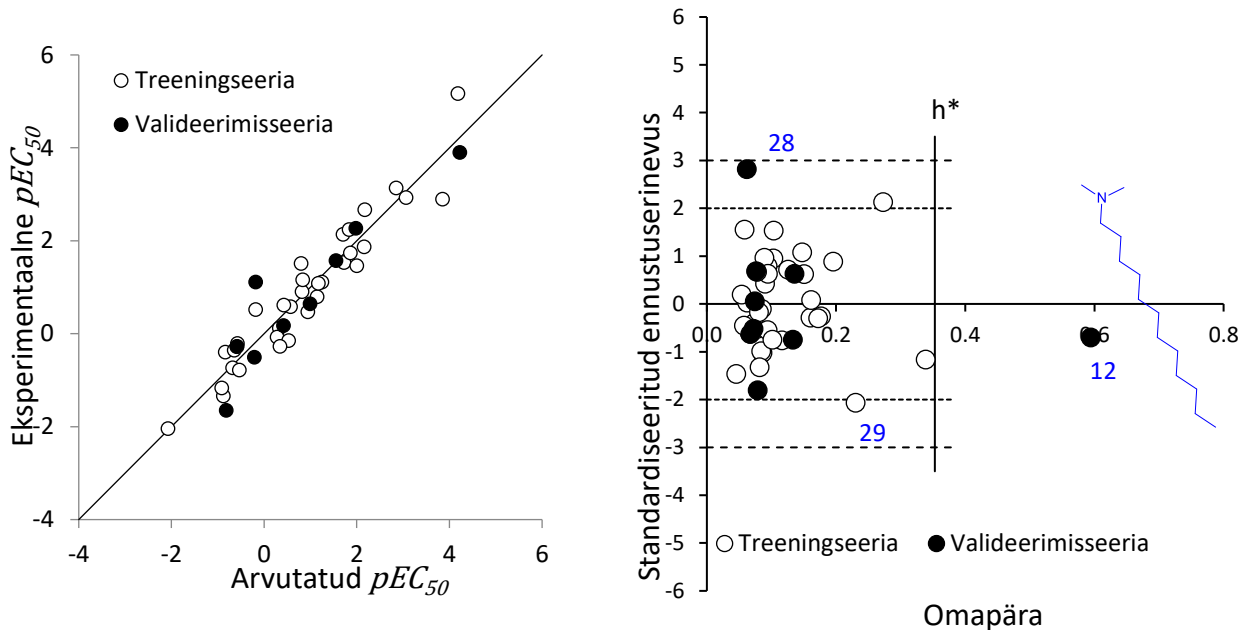
Verhaari esimesse, inertsete ühendite klassi kuulus andmeseeriast 43 ühendit, millest 34 määrati treeningseeriasse ning 9 valideerimisseeriasse. Kuna tegemist on Verhaari klassidest kõige homogeensemaga, kirjeldab juba ühe molekulaartunnusega mudel keemilise ühendi seost toksilisusega võrdlemisi hästi ( $R^2 = 0,60$ ;  $Q^2 = 0,54$ ;  $F = 47,19$ ;  $\sigma^2 = 0,94$ ). Kui üldiselt on inertsed ühendid heas korrelatsioonis  $\log K_{ow}$ -ga, siis selles töös andis parima ühe molekulaartunnusega mudeli *BCUT<sub>p-11</sub>* (ingl *nhigh lowest polarizability weighted BCUTS*), mis on  $pEC_{50}$ -ga positiivses korrelatsioonis. *BCUT<sub>p-11</sub>* sisaldab infot aatomitevahelise seotuse ja aatomite omaduste kohta, antud juhul aatomite polariseeritavuse kohta. *BCUT<sub>p-11</sub>* väärtused on kõrged suure molekulmassiga polüaromaatsetel ühenditel 3,4-bensopüreenil (nr 26) ja naftaleenil (nr 40) ning kõige madalam kompaktsel atsetoonil (nr 34). Madalad *BCUT<sub>p-11</sub>* väärtused on ka halogeenitud karboksüülhapetel (nr-d 10, 22, 32).

Edasi analüüsiti kuni viie molekulaartunnusega mudeleid ning jälgides muutust  $R^2$  kasvus, on sobiv molekulaartunnuste arv kaks (graafik L3.1). Kuna kahe molekulaartunnuse seas oli aatomite elektrotopoloogist olekut kirjeldav molekulaartunnus *ndssC* (ingl *count of atom-type E-State: =C<*), millel puudus suurele osale modelleeritavatest molekulidest kirjeldus, valiti lõppmudelisse kolm molekulaartunnust (tabel 3, graafik 1). Kõik kolm valitud molekulaartunnust on  $pEC_{50}$ -ga positiivses korrelatsioonis.

**Tabel 3.** Mudel Verhaari 1. klassi ühendite ja selle statistilised parameetrid  
 $N = 34$ ;  $n = 3$ ;  $R^2 = 0,91$ ;  $Q^2 = 0,88$ ;  $F = 106,76$ ;  $\sigma^2 = 0,21$

Molekulaartunnus	Koefitsient	$\sigma$	t-test
Vabaliige	-5,3238	0,3952	-13,4721
<i>BCUTp-1l</i>	0,9659	0,0764	12,6458
<i>SpDiam_Dzi</i>	0,0401	0,0038	10,4653
<i>VE3_Dzv</i>	0,0252	0,0051	4,9868

Lühendite selgitused:  $N$  – treeningseeria ühendite arv,  $n$  – molekulaartunnuste arv,  $R^2$  – determinatsioonikordaja,  $Q^2$  – ristvalideerimiskordaja,  $F$  – Fisheri kriteerium,  $\sigma$  ja  $\sigma^2$  – standardhälve ja dispersioon



**Graafik 1 ja 2.** Graafik 1 (vasakul) kirjeldab seost Verhaari 1. klassi ühendite eksperimentaalsete ja mudeliga ennustatud  $pEC_{50}$  väärtuse vahel ning graafik 2 (paremal) on mudeli rakenduspiirkonda kirjeldav Williamsi graafik, kus  $h^*$  tähistab kriitilist omapära

Studenti t-testi järgi jäi lõppmudeli puhul olulisimaks molekulaartunnuseks parima ühe muutujaga mudeli andnud *BCUTp-1l*. Tähtsuselt järgmine molekulaartunnus *SpDiam\_Dzi* (ingl *spectral diameter from Barysz matrix / weighted by first ionization potential*) sisaldab infot nii molekuli suuruse kui ka ionisatsioonipotentsiaali kohta. *SpDiam\_Dzi* väärtused on madalad andmeseeria kõige väiksematele molekulidel: atsetoon (nr 34), kloroform (nr 35) ja diklorometaan (nr 37), mis viitab jällegi suuremate molekulide kõrgemale toksilisusele. Kolmas molekulaartunnus *VE3\_Dzv* (ingl *logarithmic coefficient sum of the last eigenvector from Barysz matrix / weighted by van der Waals volumes*) on samuti seotud suurusega, täpsemalt van der Waalsi ruumalaga.

Mudeli rakenduspiirkonna analüüsil tuvastati valideerimisseerias üks kriitilisest omapärast suurema omapäraga ühend, N,N-dimetüüldodetsüülamiin (nr 12) (graafik 2). Omapära järgi rakenduspiiridest välja jääva ühendi põhjuseks on struktuuri erinevus teistest andmeseeria ühenditest ja see avaldub ka molekulaartunnuste analüüsis. Nii ilmnebki, et N,N-dimetüüldodetsüülamiini suure omapära põhjuseks on võrreldes teiste ühenditega suurem  $VE3_{DzV}$  väärtus. Samas on antud andmepunkt väikese standardiseeritud ennustuserinevusega ja seega on struktuursele erinevusele vaatamata ennustus korrektne. See viitab ka mudeli heale ennustusvõimele. Suurema ennustuserinevusega on allüületüüleeter (nr 28) ja andmeseeria keskmisest oluliselt suurema molekulmassiga benso(a)antratseen (nr 29).

### 3.2. Mudel vähem inertsetele ühenditele

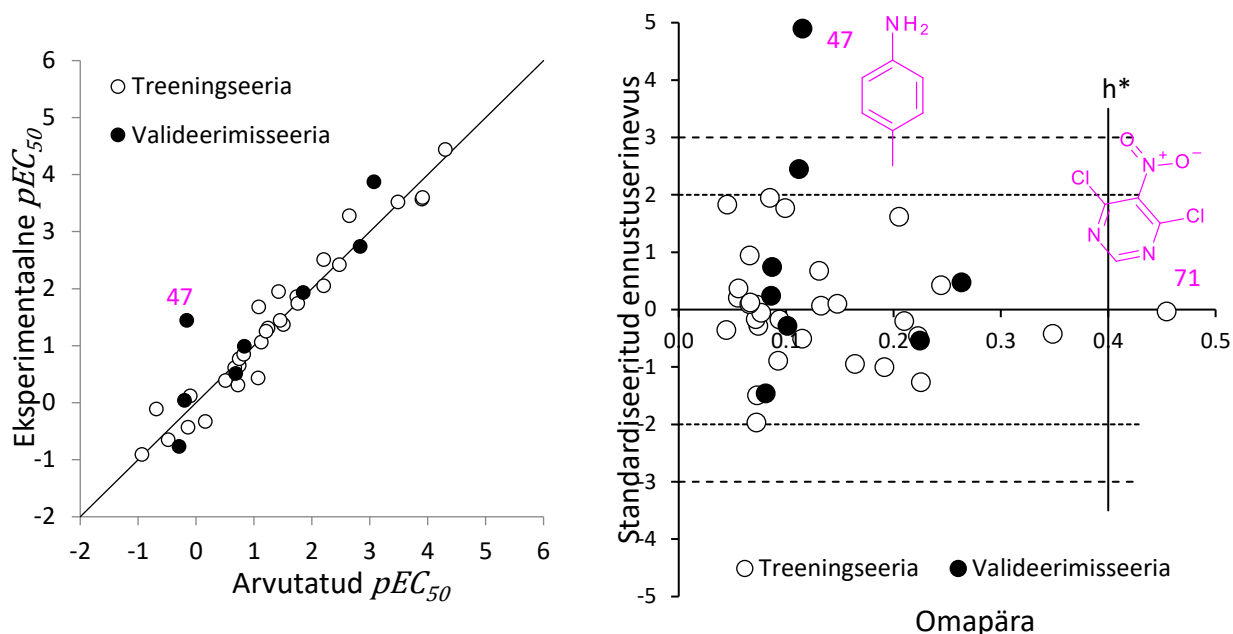
Verhaari teise, vähem inertsete ühendite klassi kuulub andmeseeriast 38 ühendit, millest 30 määrati treeningseeriasse ja 8 valideerimisseeriasse. Parim ühe muutujaga mudel saadi molekulaartunnusega  $ZMIC1$  ( $R^2 = 0,73$ ;  $Q^2 = 0,691$ ;  $F = 75,779$ ;  $\sigma^2 = 0,531$ ), seejuures ilma kõrvalekaldujateta.  $ZMIIC1$  (ingl *Z-modified information content index, neighborhood symmetry of 1-order*) on toksilisusega positiivses korrelatsioonis ja kirjeldab molekulis 1. järku naaberaatomite sümmeetriat. See tähendab, et vaadatakse sarnasust aatomist ühe sideme kaugusel ja loetakse kokku üht tüüpi fragmentide arv molekulis. Edasi analüüsiti kuni kuueparameetrilisi võrrandeid ning lõppmudelisse valiti kolm molekulaartunnust (graafik L3.2). Kõik kolm on  $pEC_{50}$ -ga positiivses korrelatsioonis (tabel 4, graafik 3).

**Tabel 4.** Mudel Verhaari 2. klassi ühenditele ja selle statistilised parameetrid

$N = 30$ ;  $n = 3$ ;  $R^2 = 0,95$ ;  $Q^2 = 0,94$ ;  $F = 163,46$ ;  $\sigma^2 = 0,11$

Molekulaartunnus	Koefitsient	$\sigma$	t-test
Vabaliige	-4,0631	0,5787	-7,0211
$C2SP3$	0,3638	0,0188	19,3414
$AATS1p$	2,1665	0,3494	6,1998
$piPC7$	0,2461	0,0415	5,9308

Lühendite selgitused:  $N$  – treeningseeria ühendite arv,  $n$  – molekulaartunnuste arv,  $R^2$  – determinatsioonikordaja,  $Q^2$  – ristvalideerimiskordaja,  $F$  – Fisheri kriteerium,  $\sigma$  ja  $\sigma^2$  – standardhälve ja dispersioon



**Graafik 3 ja 4.** Graafik 3 (vasakul) kirjeldab seost Verhaari 2. klassi ühendite eksperimentaalsete ja mudeliga ennustatud  $pEC_{50}$  väärtuse vahel ning graafik 4 (paremal) on mudeli rakenduspiirkonda kirjeldav Williamsi graafik, kus  $h^*$  tähistab kriitilist omapära

Studenti t-testi järgi hinnati olulisimaks molekulaartunnuseks  $C2SP3$  (ingl *singly bound carbon bound to two other carbons*).  $C2SP3$  väärtused on kõrgemad just suure süsinikskeletiga ühendite puhul ning kuna modelleeriti orgaanilisi ühendeid, siis on  $C2SP3$  seotud peamiselt molekuli suurusega. Molekulaartunnuse puudusena ei kirjelda see ühendeid, milles ühekordsed süsinik-süsinik sidemed puuduvad.

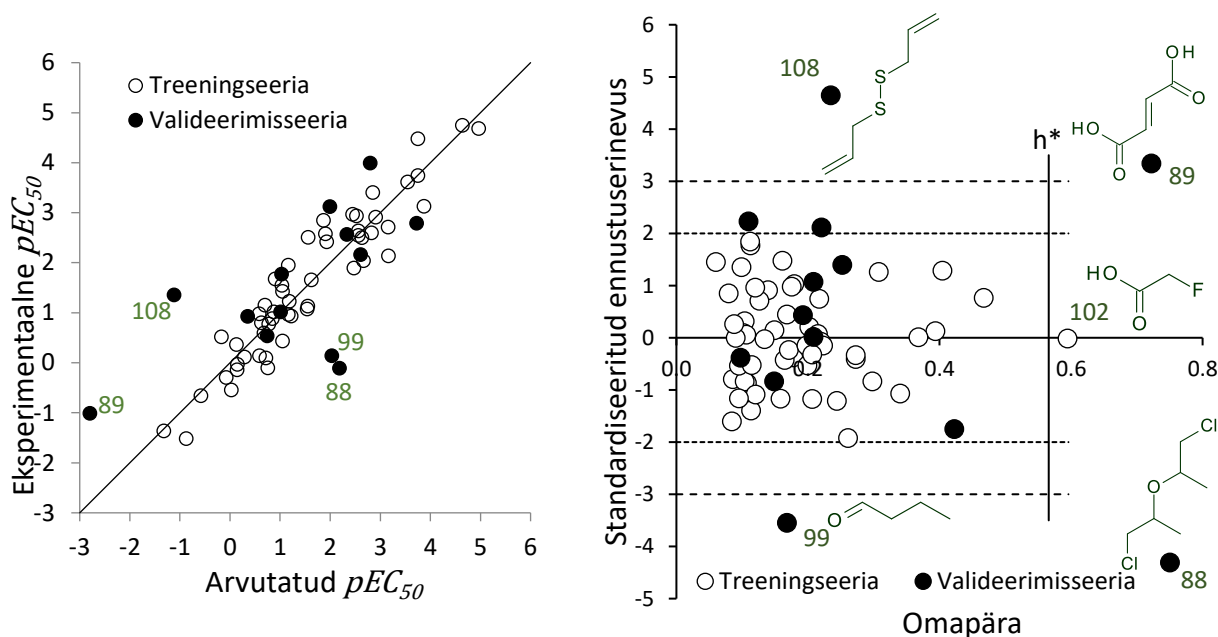
Studenti t-testi järgi on ülejäänud kahe molekulaartunnuse panus ühendi toksilisuse kirjeldamisel peaaegu võrdne. Need molekulaartunnused on  $AATS1p$  (ingl *average Broto-Moreau autocorrelation – lag 1 / weighted by polarizabilities*), mis iseloomustab topoloogilises distantsil üks olevate aatomite polartiseeritavust, ja  $piPC7$  (ingl *conventional bond order ID number of order 7 ( $\ln(1+x)$ )*), mis näitab selgelt suuremate ühendite kõrgemat toksilisust.

Mudeli rakenduspiirkonda hinnati Williamsi graafikuga (graafik 4), mille põhjal tuvastati, et kuigi valideerimisühend 4-metüülaniliin (nr 47) kuulub omapära järgi rakenduspiirkonda, pole sellele tehtud ennustus täpne. Ilmselt on põhjus mudelisse valitud molekulaartunnustes, sest  $C2SP3$  ja  $piPC7$  väärtused 4-metüülaniliinile on nullid, mis tähendab, et kumbki molekulaartunnus ühendi

toksilisust ei kirjelda. Rakendusala piiridest jäi välja ka 4,6-dikloro-5-nitropürimidiin (nr 71), kuid selle  $pEC_{50}$  väärtus on mudeliga täpselt kirjeldatud.

### 3.3. Mudel reaktsioonivõimelistele ühenditele

Verhaari kolmandasse, reaktsioonivõimeliste ühendite klassi kuulus 66 ühendit, millest 53 määrati treeningseeriasse ja 13 valideerimisseeriasse. Erinevalt Verhaari 1. ja 2. klassist polnud selle andmeseeria jaoks üht domineerivat molekulaartunnust ( $R^2 > 0,5$ ). Kokku analüüsiti kuni 13 molekulaartunnusega võrrandeid. Graafikul L3.3 on näha, et pärast esimest nn käänukohta kasvavad mudelite  $R^2$  ja  $Q^2$  väärtused ühtlaselt ning seetõttu jätkati molekulaartunnuste lisamist kuni järgmise käänukohani, mis jäi üheksa molekulaartunnuse juurde. Lõppmudelisse valitigi üheksa molekulaartunnust (tabel 5, graafik 5), mis muudab molekulaartunnuste lahtiseletamise keerulisemaks.



**Graafik 5 ja 6.** Graafik 5 (vasakul) kirjeldab seost Verhaari 3. klassi ühendite eksperimentaalsete ja mudeliga ennustatud  $pEC_{50}$  väärtuse vahel ning graafik 6 (paremal) on mudeli rakenduspiirkonda kirjeldav Williamsi graafik, kus  $h^*$  on kriitiline omapära

**Tabel 5.** Mudel Verhaari 3. klassi ühenditele ja selle statistilised parameetrid $N = 53; n = 9; R^2 = 0,89; Q^2 = 0,84; F = 39,46; \sigma^2 = 0,29$ 

Molekulaartunnus	Koefitsient	$\sigma$	t-test
Vabaliige	-7,7859	1,0406	-7,4825
TIC5	0,0155	0,0014	10,9677
AATSC1c	240,3230	23,1773	10,3689
GATS3s	1,9757	0,2043	9,6705
meanI	2,4056	0,2652	9,0712
BCUTc-1h	14,2598	1,9726	7,2291
GATS3p	-1,9715	0,3598	-5,4795
VE2_Dze	49,9129	9,9162	5,0335
ETA_BetaP_ns_d	13,3113	2,9233	4,5535
MATS2m	-2,7785	0,6608	0,4539

Lühendite selgitused:  $N$  – treeningseeria ühendite arv,  $n$  – molekulaartunnuste arv,  $R^2$  – determinatsioonikordaja,  $Q^2$  – ristvalideerimiskordaja,  $F$  – Fisheri kriteerium,  $\sigma$  ja  $\sigma^2$  – standardhälve ja dispersioon

Studenti t-testi järgi on mudeli olulisim molekulaartunnus TIC5 (ingl *total information content index (neighborhood symmetry of 5-order)*), mis kirjeldab molekulis naaberaatomite viiendat järku sümmeetriat ehk siis sarnaste viiesidemeliste fragmentide hulka. Suuruse kohta sisaldab infot ka mudeli kõige väiksema osakaaluga molekulaartunnus MATS2m (ingl *Moran autocorrelation - lag 2 / weighted by mass*).

Kuna 3. klassi ühendid on võrreldes 1. ja 2. klassi ühenditega reaktsioonivõimelisemad ja andmeseeria koosneb üksteisest erinevalt käituvatest molekulidest, on mudelis tähtsal kohal elektroonseid efekte kirjeldavad molekulaartunnused. AATSC1c (ingl *average centered Broto-Moreau autocorrelation - lag 1 / weighted by charges*) sisaldab infot topoloogilisel distantstil üks olevate aatomite laengu, GATS3s (ingl *Geary autocorrelation - lag 3 / weighted by I-state*) topoloogilisel distantstil kolm olevate aatomite ionisatsiooni ja GATS3p (ingl *Geary autocorrelation - lag 3 / weighted by polarizabilities*) topoloogilisel distantstil kolm olevate aatomite polariseeritavuse kohta. Samuti kirjeldavad elektroonseid efekte BCUTc-1h (ingl *nlow highest partial charge weighted BCUTS*) ja VE2\_Dze (ingl *average coefficient sum of the last eigenvector from Barysz matrix / weighted by Sanderson electronegativities*), mis sisaldavad infot vastavalt osalaengute ja elektronegatiivsuse kohta.

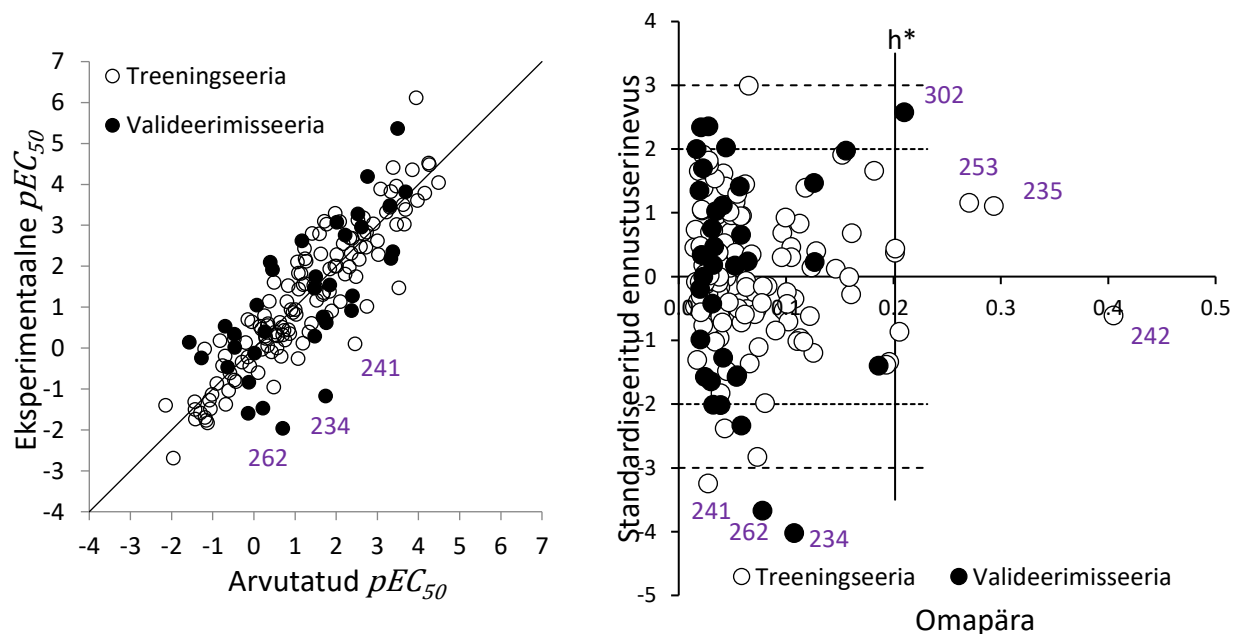
Elektroonseid efekte kirjeldavad veel mudelisse valitud meanI (ingl *mean intrinsic state values I*), mis võtab arvesse heteroatomite sidemejärgusid ja suurt  $\pi$ -elektronide arvu ning kasutab selleks aatomite valents- ja  $\sigma$ -elektronide arvu ning peakvantarvu. ETA\_BetaP\_ns\_d (ingl *measure of lone electrons entering into resonance relative to molecular size*) kirjeldab üksikuid vabu elektronpaare, mis osalevad resonantsefektides, ja seob selle molekuli suurusega. ETA\_BetaP\_ns\_d on

toksilisusega positiivses korrelatsioonis, mis viitab jällegi suuremate molekulide kõrgemale toksilisusele.

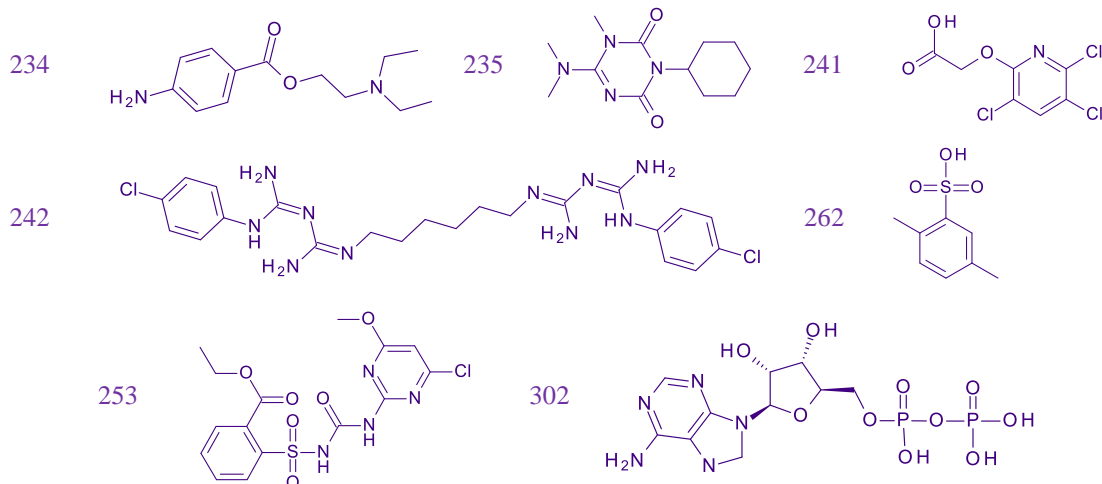
Mudeli rakenduspiirkonna analüüsil tuvastati nii omapära kui ka ennustusvea järgi kaks kõrvalekaldujat: 2,2'-diklorodiisopropüüleeter (nr 88) ja fumaarhape (nr 89) (graafik 6). Kuigi omapära järgi rakenduspiirkonnas, polnud mudeli ennustus butanaalile (nr 99) ja diallüüldisulfiidile (nr 108) täpne ega lubatud ennustuserinevuse piirides. Omapära poolest jääb rakenduspiirkonnast välja ka 2-fluoroetaanhape (nr 102), mis eristub teistest, kuna sisaldab fluoriaatomit. Kõik kõrvalekaldujad on võrdlemisi väikesed molekulid.

### 3.4. Mudel liigitamisele mittealluvatele ühenditele

Verhaari 5. klassi kuuluvad ühendid, mida ei suudetud 1.–4. klassi määrata, ja seega on tegemist võrdlemisi heterogeense klassiga. Viiendasse klassi kuulus andmeseeriast 166 ühendit, millest treeningseeriasse määrati 131 ja valideerimisseeriasse 35 ühendit. Graafikul L3.4 on näha, et pärast esimest käänukohta olid  $R^2$  ja  $Q^2$  väärtused madalad ning jätkati mudeldamist kuni  $R^2$  väärtus kasvas veel ligi 0,02 ühikut. Lõppmudelisse valiti kaheksa molekulaartunnust (tabel 6, graafik 7).



**Graafik 7 ja 8.** Graafik 7 (vasakul) kirjeldab seost Verhaari 5. klassi ühendite eksperimentaalsete ja mudeliga ennustatud  $pEC_{50}$  väärtuse vahel ning graafik 8 (paremal) on mudeli rakenduspiirkonda kirjeldav Williamsi graafik, kus  $h^*$  on kriitiline omapära. Hälbijate struktuurid on toodud joonisel 1.



Joonis 1. Verhaari 5. klassile tehtud mudeli hälbijate struktuurid

Tabel 6. Mudel Verhaari 5. klassi ühenditele ja selle statistilised parameetrid  
 $N = 131$ ;  $n = 8$ ;  $R^2 = 0,78$ ;  $Q^2 = 0,75$ ;  $F = 53,47$ ;  $\sigma^2 = 0,67$

Molekulaartunnus	Koefitsient	$\sigma$	t-test
Vabaliige	-1,6372	0,5825	-2,8106
<i>nHBAcc</i>	-0,4692	0,0390	-12,0170
<i>SpMax5_Bhe</i>	2,1828	0,3334	6,5471
<i>CrippenMR</i>	0,0503	0,0078	6,4688
<i>BCUTw-1h</i>	0,0225	0,0041	5,5238
<i>SpMin2_Bhm</i>	-2,9290	0,5463	-5,3614
<i>MDEN-22</i>	0,3931	0,0815	4,8261
<i>ATSC4i</i>	0,0363	0,0089	4,0821
<i>ATSC8p</i>	-0,1656	0,0542	-3,0557

Lühendite selgitused:  $N$  – treeningseeria ühendite arv,  $n$  – molekulaartunnuste arv,  $R^2$  – determinatsioonikordaja,  $Q^2$  – ristvalideerimiskordaja,  $F$  – Fisheri kriteerium,  $\sigma$  ja  $\sigma^2$  – standardhälve ja dispersioon

Studenti t-testi järgi on esimesel positsioonil molekulaartunnus *nHBAcc*, mis loendab molekulis vesiniksideme aktseptoreid ja näitab vesiniksidemete olulisust toksilisuse mehhanismides. Olulisuselt teine molekulaartunnus on elektronegatiivsust kirjeldav *SpMax5\_Bhe* (ingl *largest absolute eigenvalue of Burden modified matrix - n 5 / weighted by relative Sanderson electronegativities*). Elektroonseid efekte kirjeldavad ka *ATSC4i* (ingl *centered Broto-Moreau autocorrelation - lag 4 / weighted by first ionization potential*), mis on seotud topoloogilises distantstil neli olevate aatomite ionisatsioonipotentsiaaliga, ja *ATSC8p* (ingl *centered Broto-Moreau autocorrelation - lag 8 / weighted by polarizabilities*), mis on seotud topoloogilisel distantstil kaheksa olevate aatomite polariseeritavusega.

Olulisuselt kolmas molekulaartunnus on *CrippenMR* (ingl *Crippen's molar refractivity*), mis on seotud molekuli suurusega. Kuna *CrippenMR* on  $pEC_{50}$ -ga positiivselt korreleeritud, viitab see

suuremate molekulide kõrgemale toksilisusele. Suurust kirjeldavad veel *BCUTw-1h* (ingl *nlow highest atom weighted BCUTS*), mis kombineerib infot aatomitevahelise seotuse ja aatommasside kohta, ning *SpMin2\_Bhm* (ingl *smallest absolute eigenvalue of Burden modified matrix - n 2 / weighted by relative mass*). Molekulaartunnus *MDEN-22* (ingl *molecular distance edge between all secondary nitrogens*) on ühest küljest seotud sekundaarsete lämmastike olemasoluga molekulis, ent teisest küljest sisaldab samuti infot molekuli suuruse kohta.

Mudeli rakenduspiirkonna analüüsil leiti valideerimisseeria ühendite seas kaks tugevat kõrvalekaldujat, 2,5-dimetüülenseensulfoonhappe (nr 262) ja 2-(dietüülamino)etüül-4-aminobensoaat (nr 234) – mõlemal on väga madalad *ATSC4i* väärtused. Treeningseeria ühendite seas on üks kõrvalekalduja, 3,5,6-trikloro-2-püridiloksüetaanhape (nr 241), millel on peaaegu kõik vesinikud asendatud klooriaatomitega. Samas on see võrdlemisi lähedal  $-3\sigma$  joonele. Lisaks on veel mõned ühendid, mis ei kuulu rakenduspiirkonda omapära järgi, ent neile tehtud ennustused jäävad lubatud ennustusvea piiridesse.

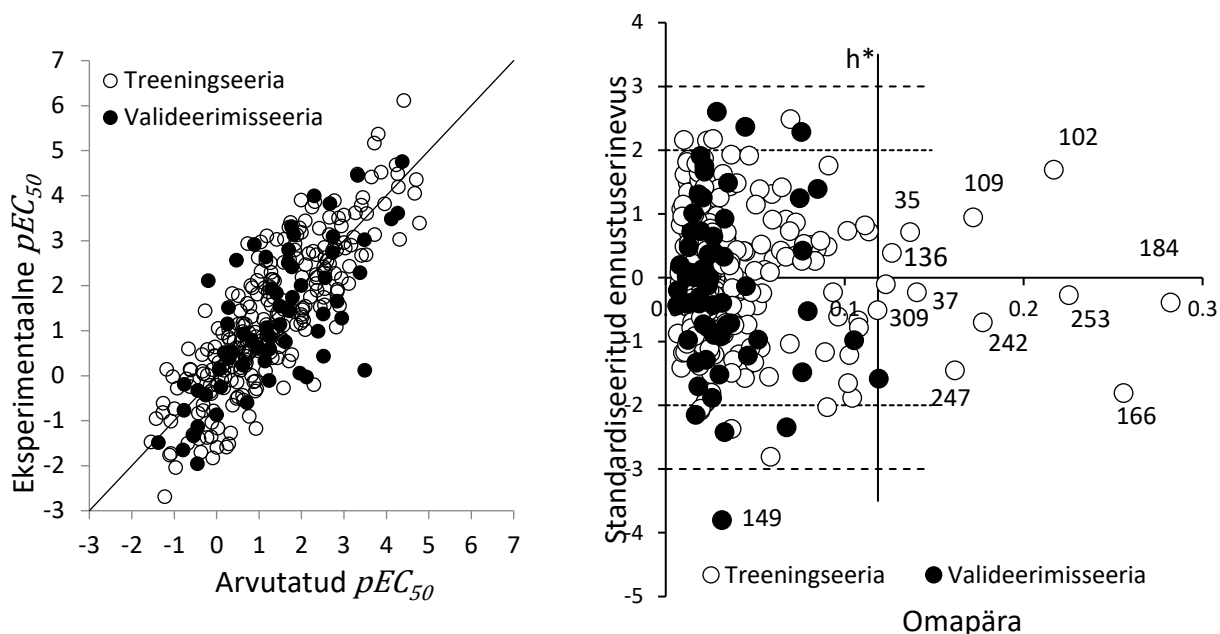
### 3.5. Mudel kogu andmeseeriale

Kokku oli modelleeritavas andmeseerias 316 ühendit, millest 253 kuulusid treeningseeriasse ning 63 valideerimisseeriasse. Sobiv molekulaartunnuste arv valiti  $R^2$  väärtuse kasvu järgi (graafik L3.5), mis üheksanda molekulaartunnuse lisamisel paranes ligi 0,02 ühikut. Seega valiti lõppmudelisse üheksa molekulaartunnust (graafik 9, tabel 7).

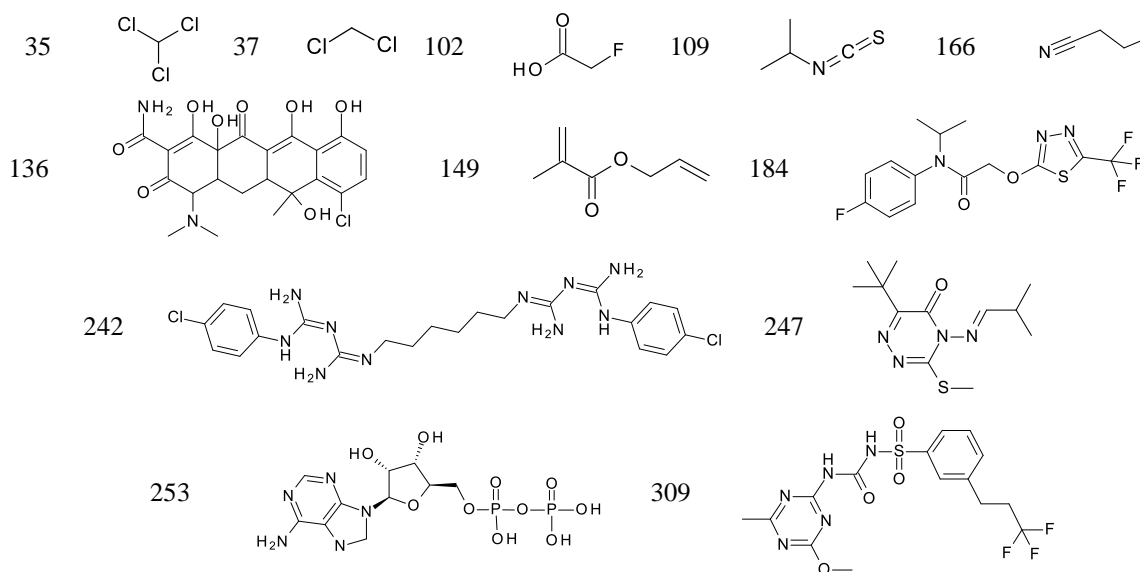
**Tabel 7.** Mudel kogu andmeseeriale ja selle statistilised parameetrid  
 $N = 253; n = 9; R^2 = 0,64; Q^2 = 0,60; F = 47,91; \sigma^2 = 0,96$

Molekulaartunnus	Koefitsient	$\sigma$	t-test
Vabaliige	-1,5496	0,6439	-2,4068
<i>CrippenMR</i>	0,0614	0,0039	15,7905
<i>MLFER_BO</i>	-0,8713	0,0997	-8,7362
<i>ASP-3</i>	5,3085	1,0320	5,1438
<i>ATSC6v</i>	-0,000857	0,000189	-4,5229
<i>MATS5v</i>	-0,9937	0,2490	-3,9910
<i>AATSC0i</i>	0,7967	0,2024	3,9353
<i>ATSC8p</i>	-0,1889	0,0504	-3,7453
<i>GATS5c</i>	-0,4555	0,1241	-3,6698
<i>SpMax1_Bhs</i>	-0,3699	0,1174	-3,1503

Lühendite selgitused:  $N$  – treeningseeria ühendite arv,  $n$  – molekulaartunnuste arv,  $R^2$  – determinatsioonikordaja,  $Q^2$  – ristvalideerimiskordaja,  $F$  – Fisheri kriteerium,  $\sigma$  ja  $\sigma^2$  – standardhälve ja dispersioon



**Graafik 9 ja 10.** Graafik 9 (vasakul) kirjeldab seost kogu andmeseeriale tehtud mudeli eksperimentaalsete ja ennustatud  $pEC_{50}$  väärtuse vahel ning graafik 10 (paremal) on mudeli rakenduspiirkonda kirjeldav Williamsi graafik, kus  $h^*$  on kriitiline omapära. Hälbijate struktuurid on toodud joonisel 2.



**Joonis 2.** Kogu andmeseeriale tehtud mudeli hälbijate struktuurid

Studenti t-testi järgi on olulisim molekulaartunnus Crippeni molaarne refraktsioon (*CrippenMR*), mis on seotud molekuli suurusega. *CrippenMR* on  $pEC_{50}$ -ga positiivses sõltuvuses ja viitab suuremate molekulide kõrgemale toksilisusele. Samuti on suurusega seotud *ATSC6v* (ingl *centered*

*Broto-Moreau autocorrelation - lag 6 / weighted by van der Waals volumes*) ja *MATS5v* (ingl *Moran autocorrelation - lag 5 / weighted by van der Waals volumes*), mille järgi mõjutab ühendi toksilisust topoloogilisel distantstil kuus või viis olevate aatomite van der Waalsi ruumala. Kaudselt võib olla suurusega seotud ka *ASP-3* (ingl *average simple path, order 3*), mis loendab kolmeaatomilisi fragmente.

Tähtsuselt teine molekulaartunnus on *MLFER\_BO* (ingl *overall or summation solute hydrogen bond basicity*), mis kirjeldab vesiniksideme aluselisust ehk siis molekuli võimet prootonit siduda. Olles *pEC<sub>50</sub>*-ga negatiivses korrelatsioonis, tähendab see, et head vesiniksideme doonorid võivad olla toksilisemad. See on kooskõlas ka 5. klassi mudelisse valitud molekulaartunnusega *nHBAcc* ning ka varem on *Pseudokirchneriella subcapitatale* tehtud QSAR mudelites saadud hea korrelatsioon sarnast efekti kirjeldava *ELUMO*-ga.

Lisaks on mudelis veel elektroonseid efekte kirjeldavad molekulaartunnused, nagu aatomite ionisatsiooni kirjeldavad *AATSC0i* (ingl *average centered Broto-Moreau autocorrelation - lag 0 / weighted by first ionization potential*) ja *SpMax1\_Bhs* (ingl *largest absolute eigenvalue of Burden modified matrix - n 1 / weighted by relative I-state*). *GATS5c* (ingl *Geary autocorrelation - lag 5 / weighted by charges*) sisaldab infot topoloogilisel distantstil viis olevate aatomite laengute kohta. *ATSC8p* (ingl *centered Broto-Moreau autocorrelation - lag 8 / weighted by polarizabilities*) on seotud topoloogilisel distantstil kaheksa olevate aatomite polariseeritavusega ning kuna polariseeritavus on seotud aatomi suurusega, sisaldab *ATSC8p* infot ka molekuli suuruse kohta.

Mudeli rakenduspiirkonna analüüsil (graafik 10) ilmnes valideerimisseeria ühendite seas üks  $-3\sigma$  piirist kõrvalekalduja, allüülmetakrülaat (nr 149), mille kõrvalekaldumisel puudub hetkel põhjendatud seletus. Samuti oli treeningseeria ühendite seas neid, mis hälbisid omapära poolest, kuid neile tehtud ennustused jäid lubatud ennustusvea piiridesse. Üks osa omapära järgi hälbijatest on väikese molekulmassiga ühendid, mis sisaldavad halogeeniaatomeid, tsüno- või isotiotsünaadirühma. Teine osa hälbijatest on keerulisema struktuuriga ühendid. Kogu andmeseeriale tehtud mudeli hälbijatest kolm (nr-d 102, 253, 242) olid hälbijad ka teistes valminud mudelites.

## Kokkuvõte

Bakalaureusetöö eesmärk oli uurida ja kirjeldada keemilise aine struktuuri ja vetika *Pseudokirchneriella subcapitata* kasvu pärssiva mõju vahelisi seoseid ning koostada neid seoseid kirjeldavad kvantitatiivsed struktuuri-aktiivsuse sõltuvused. Kokku tuletati viis mõõdukate kuni väga heade statistiliste parameetrite ja ennustusvõimega QSAR mudelit, millest neli koostati Verhaari klassidele ja üks kogu andmeseeriale.

Struktuurilt homogeensete inertsete ja vähem inertsete ühendite toksilisuse kirjeldamiseks piisas mõlemas mudelis kolmest molekulaartunnusest. Saadi heade statistiliste parameetrite ja ennustusvõimega mudelid (vastavalt  $R^2 = 0,91$ ;  $Q^2 = 0,88$  ja  $R^2 = 0,95$ ;  $Q^2 = 0,94$ ), seejuures vaid üksikute kõrvalekaldujatega. Nende mudelite molekulaartunnused kirjeldavad peamiselt molekuli suuruse ning elektroonsete efektide seost ühendi toksilisusega.

Struktuurilt heterogeensemata ja erinevate toimemehhanismidega ühendite modelleerimisel tuli toksilisuse kirjeldamiseks kasutada rohkem molekulaartunnuseid. Reaktsioonivõimelistele ühendite tehtud mudel sisaldas üheksat molekulaartunnust, millest seitse on seotud erinevate elektroonsete efektidega. Lisaks on esindatud suurust ja topoloogilist ülesehitust iseloomustavad molekulaartunnused. Mudelil on võrdlemisi head statistilised parameetrid ( $R^2 = 0,89$ ;  $Q^2 = 0,84$ ), kuid valideerimisseeria hulgas esines suure ennustuserinevusega ühendeid.

Liigitamisele mittealluvate ühendite toksilisuse modelleerimisel saadi mõõdukalt heade statistiliste parameetrite ja ennustusvõimega mudel ( $R^2 = 0,78$ ;  $Q^2 = 0,75$ ), millel esines mitmeid kõrvalekaldujaid nii treening- kui ka valideerimisseeria ühendite hulgas. Mudeli kaheksa molekulaartunnust kirjeldasid ühendi toksilisuse seost molekuli suuruse ja elektroonsete efektidega, samuti täheldati seost toksilisuse ja vesiniksideme aktseptorite arvu vahel.

Tendentsi, et head vesiniksideme doonorid võivad olla toksilisemad, kinnitas kogu andmeseeriale tehtud üheksa molekulaartunnusega mudel. Kogu andmeseeriale tehtud mudeli statistilised parameetrid jäid küll oodatust madalamaks ( $R^2 = 0,64$ ;  $Q^2 = 0,60$ ), ent samas tuvastati valideerimisseerias vaid üks ennustuserinevuse järgi hälbija. See on niivõrd suure andmeseeria modelleerimisel hea tulemus ning väärrib täiendavat modelleerimist mõne teise masinõppemeetodiga.

## Summary

The aim of the bachelor thesis was to study the relationships between chemical structure and compound's potential to inhibit the growth of algae *Pseudokirchneriella subcapitata*. Altogether five QSAR models with proper to excellent statistical parameters and predicting power were presented, including four models for Verhaar classes and one model for all the modeled compounds.

Modeling structurally homogeneous datasets such as inert and less inert compounds required three molecular descriptors to have excellent statistical parameters and predicting power (respectively  $R^2 = 0,91$ ;  $Q^2 = 0,88$  and  $R^2 = 0,95$ ;  $Q^2 = 0,94$ ). Also, only a few outliers were detected. The selected descriptors encoded information about the molecular size and electronic effects.

As the structural diversity and variety in mechanisms of action increase, developing a model with good predicting power gets more challenging; therefore, more descriptors were required to have suitable statistical parameters. The model for reactive compounds had nine descriptors, seven describing electronic effects. Also, descriptors encoding information about the molecular size and topological structure were selected. The model had relatively good statistical parameters and predicting power ( $R^2 = 0,89$ ;  $Q^2 = 0,84$ ), although few outliers among validation series were spotted.

The model for the compounds not classified into the first four Verhaar classes had eight descriptors and proper statistical parameters ( $R^2 = 0,78$ ;  $Q^2 = 0,75$ ). However, again, some outliers were detected also in the training set. In addition to molecular size and other electronic effects descriptors, one descriptor described the relationship between toxicity and the number of hydrogen bond acceptors in the molecule.

The same tendency that good hydrogen bond donors are more toxic than poor ones occurred in the model for all the modeled compounds. Although the statistical parameters for this nine descriptor model were lower than expected ( $R^2 = 0,64$ ;  $Q^2 = 0,60$ ), there was only one outlier in the validation set with a more significant prediction error than allowed. Considering the complexity of the data set, it was a good result, and it is worth studying the data set with more advanced machine learning methods.

## Kasutatud kirjandus

1. United Nations Environment Annual Report 2019. United Nations Environment Programme: 2020.
2. Ceschin, S.; Bellini, A.; Scalici, M. Aquatic Plants and Ecotoxicological Assessment in Freshwater Ecosystems: A Review. *Environ. Sci. Pollut. Res.* **2021**, *28* (5), 4975–4988.
3. Transforming our world: the 2030 Agenda for Sustainable Development. UN General Assembly: 2015.
4. Regulation concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). European Parliament and Council: 2006.
5. Richarz, A. N.; Neagu, D. *Big Data in Predictive Toxicology*, 1st ed.; Royal Society of Chemistry: London, 2019, p. 186–187.
6. Aruoja, V., 2011. Algae Pseudokirchneriella Subcapitata In Environmental Hazard Evaluation Of Chemicals And Synthetic Nanoparticles. Estonian University of Life Sciences.
7. Nygaard, G., Komárek, J., Kristiansen, J., Skulberg, O.M. Taxonomic designations of the bioassay alga NIVA-CHL 1 (*Selenastrum capricornutum*) and some related strains. *Opera Bot.* **1986**, *90*, 1–46.
8. OECD. Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Organisation for Economic Co-operation and Development: Paris, 2011.
9. ISO 8692 Water Quality — Fresh Water Algal Growth Inhibition Test with Unicellular Green Algae. The International Organization for Standardization: Geneva, 2012.
10. Verhaar, H. J. M.; van Leeuwen, C. J.; Hermens, J. L. M. Classifying Environmental Pollutants. *Chemosphere* **1992**, *25* (4), 471–491.
11. Enoch, S. J.; Hewitt, M.; Cronin, M. T. D.; Azam, S.; Madden, J. C. Classification of Chemicals According to Mechanism of Aquatic Toxicity: An Evaluation of the Implementation of the Verhaar Scheme in Toxtree. *Chemosphere* **2008**, *73* (3), 243–248.
12. Dimitrov, S. D.; Diderich, R.; Sobanski, T.; Pavlov, T. S.; Chankov, G. V.; Chapkanov, A. S.; Karakolev, Y. H.; Temelkov, S. G.; Vasilev, R. A.; Gerova, K. D.; Kuseva, C. D.; Todorova, N. D.; Mehmed, A. M.; Rasenberg, M.; Mekenyan, O. G. QSAR Toolbox - Workflow and Major Functionalities. *SAR QSAR Environ. Res.* **2016**, *27* (3), 203–219.
13. Patlewicz, G.; Jeliaskova, N.; Safford, R. J.; Worth, A. P.; Aleksiev, B. An Evaluation of the Implementation of the Cramer Classification Scheme in the Toxtree Software. *SAR QSAR Environ. Res.* **2008**, *19* (5–6), 495–524.
14. Roth, H. S., Membrane and cellular actions of anesthetic agents. *Federation Proceedings.* **1980**, *39* (5), 1595–1599.
15. Ellison, C. M.; Madden, J. C.; Cronin, M. T. D.; Enoch, S. J. Investigation of the Verhaar Scheme for Predicting Acute Aquatic Toxicity: Improving Predictions Obtained from Toxtree Ver. 2.6. *Chemosphere* **2015**, *139*, 146–154.

16. Lipnick, R. L. Outliers: their origin and use in the classification of molecular mechanisms of toxicity. *The Science of the Total Environment*. **1991**, 109/110, 131–153.
17. Bragin, G. E.; Parkerton, T. F.; Redman, A. D.; Letinksi, D. J.; Butler, J. D.; Paumen, M. L.; Sutherland, C. A.; Knarr, T. M.; Comber, M.; den Haan, K. Chronic Toxicity of Selected Polycyclic Aromatic Hydrocarbons to Algae and Crustaceans Using Passive Dosing. *Environ. Toxicol. Chem.* **2016**, 35 (12), 2948–2957.
18. Ding, F.; Wang, Z.; Yang, X.; Shi, L.; Liu, J.; Chen, G. Development of Classification Models for Predicting Chronic Toxicity of Chemicals to *Daphnia Magna* and *Pseudokirchneriella Subcapitata*. *SAR QSAR Environ. Res.* **2019**, 30 (1), 39–50.
19. Concu, R.; Kleandrova, V. V.; Speck-Planche, A.; Cordeiro, M. N. D. S. Probing the Toxicity of Nanoparticles: A Unified in Silico Machine Learning Model Based on Perturbation Theory. *Nanotoxicology* **2017**, 11 (7), 891–906.
20. Peric, B.; Sierra, J.; Marti, E.; Cruanas, R.; Antonia Garau, M. Quantitative Structure-Activity Relationship (QSAR) Prediction of (Eco)Toxicity of Short Aliphatic Protic Ionic Liquids. *Ecotoxicol. Environ. Saf.* **2015**, 115, 257–262.
21. Khan, K.; Roy, K. Ecotoxicological Modelling of Cosmetics for Aquatic Organisms: A QSTR Approach. *SAR QSAR Environ. Res.* **2017**, 28 (7), 567–594.
22. Banjare, P.; Singh, J.; Papa, E.; Roy, P. P. Aquatic Toxicity Prediction of Diverse Pesticides on Two Algal Species Using QSTR Modeling Approach. *Environ. Sci. Pollut. Res.* **2023**, 30 (4), 10599–10612.
23. Khan, K.; Benfenati, E.; Roy, K. Consensus QSAR Modeling of Toxicity of Pharmaceuticals to Different Aquatic Organisms: Ranking and Prioritization of the DrugBank Database Compounds. *Ecotoxicol. Environ. Saf.* **2019**, 168, 287–297.
24. Minguez, L.; Bureau, R.; Halm-Lemeille, M.-P. Joint Effects of Nine Antidepressants on *Raphidocelis Subcapitata* and *Skeletonema Marinoi*: A Matter of Amine Functional Groups. *Aquat. Toxicol.* **2018**, 196, 117–123.
25. Fu, L.; Huang, T.; Wang, S.; Wang, X.; Su, L.; Li, C.; Zhao, Y. Toxicity of 13 Different Antibiotics towards Freshwater Green Algae *Pseudokirchneriella Subcapitata* and Their Modes of Action. *Chemosphere* **2017**, 168, 217–222.
26. Onlu, S.; Sacan, M. T. An in Silico Algal Toxicity Model with a Wide Applicability Potential for Industrial Chemicals and Pharmaceuticals. *Environ. Toxicol. Chem.* **2017**, 36 (4), 1012–1019.
27. Tugcu, G.; Ertürk, M. D.; Saçan, M. T. On the Aquatic Toxicity of Substituted Phenols to *Chlorella Vulgaris*: QSTR with an Extended Novel Data Set and Interspecies Models. *J. Hazard. Mater.* **2017**, 339, 122–130.
28. Hsieh, S.-H.; Hsu, C.-H.; Tsai, D.-Y.; Chen, C.-Y. Quantitative Structure-Activity Relationships for Toxicity of Nonpolar Narcotic Chemicals to *Pseudokirchneriella Subcapitata*. *Environ. Toxicol. Chem.* **2006**, 25 (11), 2920–2926.
29. Chen, C. Y.; Wang, Y. J.; Yang, C. F. Estimating Low-Toxic-Effect Concentrations in Closed-System Algal Toxicity Tests. *Ecotoxicol. Environ. Saf.* **2009**, 72 (5), 1514–1522.

30. Aruoja, V.; Moosus, M.; Kahru, A.; Sihtmae, M.; Maran, U. Measurement of Baseline Toxicity and QSAR Analysis of 50 Non-Polar and 58 Polar Narcotic Chemicals for the Alga *Pseudokirchneriella Subcapitata*. *Chemosphere* **2014**, *96*, 23–32.
31. Furuhashi, A.; Hasunuma, K.; Hayashi, T. I.; Tatarazako, N. Predicting Algal Growth Inhibition Toxicity: Three-Step Strategy Using Structural and Physicochemical Properties. *Sar Qsar Environ. Res.* **2016**, *27* (5), 343–362.
32. Bakire, S.; Yang, X.; Ma, G.; Wei, X.; Yu, H.; Chen, J.; Lin, H. Developing Predictive Models for Toxicity of Organic Chemicals to Green Algae Based on Mode of Action. *Chemosphere* **2018**, *190*, 463–470.
33. Chen, C.-Y.; Ko, C.-W.; Lee, P.-I. Toxicity of Substituted Anilines to *Pseudokirchneriella Subcapitata* and Quantitative Structure-Activity Relationship Analysis for Polar Narcotics. *Environ. Toxicol. Chem.* **2007**, *26* (6), 1158–1164.
34. Chen, C.-Y.; Lin, J.-H. Toxicity of Chlorophenols to *Pseudokirchneriella Subcapitata* under Air-Tight Test Environment. *Chemosphere* **2006**, *62* (4), 503–509.
35. Lee, P. Y.; Chen, C. Y. Toxicity and Quantitative Structure–Activity Relationships of Benzoic Acids to *Pseudokirchneriella Subcapitata*. *J. Hazard. Mater.* **2009**, *165* (1–3), 156–161.
36. Huang, C.-P.; Wang, Y.-J.; Chen, C.-Y. Toxicity and Quantitative Structure–Activity Relationships of Nitriles Based on *Pseudokirchneriella Subcapitata*. *Ecotoxicol. Environ. Saf.* **2007**, *67* (3), 439–446.
37. Chen, C. Y.; Kuo, K.-L.; Fan, J.-W. Toxicity of Propargylic Alcohols on Green Alga—*Pseudokirchneriella Subcapitata*. *J. Environ. Monit.* **2012**, *14* (1), 181–186.
38. Guimaraes, M. C.; da Mota, E. G.; Silva, D. G.; Freitas, M. P. Aug-MIA-QSPR Modelling of the Toxicities of Anilines and Phenols to *Vibrio Fischeri* and *Pseudokirchneriella Subcapitata*. *Chemom. Intell. Lab. Syst.* **2014**, *134*, 53–57.
39. Jin, X.; Jin, M.; Sheng, L. Three Dimensional Quantitative Structure-Toxicity Relationship Modeling and Prediction of Acute Toxicity for Organic Contaminants to Algae. *Comput. Biol. Med.* **2014**, *51*, 205–213.
40. Pramanik, S.; Roy, K. Predictive Modeling of Chemical Toxicity towards *Pseudokirchneriella Subcapitata* Using Regression and Classification Based Approaches. *Ecotoxicol. Environ. Saf.* **2014**, *101*, 184–190.
41. Khan, K.; Roy, K. Ecotoxicological QSAR Modelling of Organic Chemicals against *Pseudokirchneriella Subcapitata* Using Consensus Predictions Approach. *SAR QSAR Environ. Res.* **2019**, *30* (9), 665–681.
42. Masand, V. H.; Zaki, M. E. A.; Al-Hussain, S. A.; Ben Ghorbal, A.; Akasapu, S.; Lewaa, I.; Ghosh, A.; Jawarkar, R. D. Identification of Concealed Structural Alerts Using QSTR Modeling for *Pseudokirchneriella Subcapitata*. *Aquat. Toxicol.* **2021**, *239*, 105962.
43. Yu, X. Quantitative Structure-Toxicity Relationships of Organic Chemicals against *Pseudokirchneriella Subcapitata*. *Aquat. Toxicol.* **2020**, *224*, 10549637.
44. Singh, K. P.; Gupta, S.; Kumar, A.; Mohan, D. Multispecies QSAR Modeling for Predicting the Aquatic Toxicity of Diverse Organic Chemicals for Regulatory Toxicology. *Chem. Res. Toxicol.* **2014**, *27* (5), 741–753.

45. Lotfi, S.; Ahmadi, S.; Kumar, P. Ecotoxicological Prediction of Organic Chemicals toward *Pseudokirchneriella Subcapitata* by Monte Carlo Approach. *RSC Adv.* **2022**, *12* (38), 24988–24997.
46. Kusk, K. O.; Christensen, A. M.; Nyholm, N. Algal Growth Inhibition Test Results of 425 Organic Chemical Substances. *Chemosphere* **2018**, *204*, 405–412.
47. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. PubChem 2023 update. *Nucleic Acids Res.*, **2023**, *51* (D1), D1373–D1380.
48. Stewart, J. J. P. MOPAC: A Semiempirical Molecular Orbital Program. *J. Comput. Aided Mol. Des.* **1990**, *4* (1), 1–103.
49. Schaftenaar, G.; Noordik, J. H. Molden: A Pre- and Post-Processing Program for Molecular and Electronic Structures. *J. Comput. Aided Mol. Des.* **2000**, *14* (2), 123–134.
50. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902–3909.
51. Katritzky, A. R., Petruhkin, R., Petruhkina, I., Lomaka, A., Tatham, D. B., Karelson, M. Codessa PRO: Comprehensive Descriptors for Structural and Statistical Analysis. University of Florida: 2001.
52. O’Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, *3* (1), 33.
53. Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32* (7), 1466–1474.
54. Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model with Knowledge, *J. Chem. Inf. Model.* **2007**, *47*, 2140–2148.
55. Todeschini, R.; Consonni, V. *Molecular Descriptors for Cheminformatics, Volume I: Alphabetical Listing*, 2nd ed.; VCH: Weinheim, Germany, 2009.
56. Karelson, M. *Molecular Descriptors in QSAR/QSPR*. Wiley- Interscience publication: 2000, p. 391–394, 396–399.
57. Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *Alternatives to Laboratory Animals.* **2005**, *33* (5), 445–459.
58. Moosus, M.; Maran, U. Quantitative structure–activity relationship analysis of acute toxicity of diverse chemicals to *Daphnia magna* with whole molecule descriptors. *SAR and QSAR in Environmental Research.* **2011**, *22* (7–8), 757–774.

# Lisad

## Lisa 1. *Pseudokirchneriella subcapitata* tehtud mudelid sarnase struktuuri ja toimeviisiga ühenditele (1–13) ja suure struktuurse varieeruvusega ühenditele (14–23)

Ühendid	$N_{greenin} / N_{test}$	Meetod	Võrrand	$R^2$	$Q^2$	$\sigma$
1. Klorofenoolid [34]	8	LR	$\log(1/EC50)_{DO} = 1,3014\log P + 0,8299$ $\log(1/EC50)_{DO} = -0,7557pK_a + 10,568$ $\log(1/EC50)_{GR} = 1,1576\log P + 1,3304$ $\log(1/EC50)_{GR} = -0,7122pK_a + 10,348$	0,93 0,90 0,94 0,96	- - - -	- - - -
2. Nitriilid [36]	12	(M)LR	$\log(1/EC50)_{DO} = -51,8E_{lumo} - 1,70$ $\log(1/EC50)_{GR} = -53,7E_{lumo} - 1,91$ $\log(1/EC50)_{DO} = -1,71 - 0,075\log K_{ow} - 52,96E_{lumo}$ $\log(1/EC50)_{GR} = -1,92 - 0,034\log K_{ow} - 54,23E_{lumo}$	0,85 0,85 0,85 0,85	0,78 0,77 0,71 0,71	- - - -
3. Verhaari 1. klass [28]	26	LR	$\log(1/EC50)_{DO} = 0,978\log K_{ow} - 1,83$ $\log(1/EC50)_{GR} = 0,974\log K_{ow} - 1,95$	0,94 0,94	0,932 0,933	0,332 0,325
4. Verhaari 2. klass [33]	19 20 19 18 19 18	(M)LR	$\log(1/EC50)_{DO} = 1,136\log K_{ow} - 1,771$ $\log(1/EC50)_{GR} = 0,946\log K_{ow} - 1,100$ $\log(1/EC50)_{DO} = -2,572E_{lumo} + 1,410$ $\log(1/EC50)_{GR} = -2,491E_{lumo} + 1,486$ $\log(1/EC50)_{DO} = -0,24 - 1,331E_{lumo} + 0,588\log K_{ow}$ $\log(1/EC50)_{GR} = 0,81 - 1,97E_{lumo} + 0,24\log K_{ow}$	0,885 0,750 0,886 0,879 0,916 0,880	0,829 0,715 0,833 0,821 0,866 0,817	0,372 0,499 0,370 0,368 0,328 0,371
5. Bensoehappe derivaadid [35]	8 8 20 20	(M)LR	$\log(1/EC50)_{FY} = 0,973\log K_{ow} - 1,169$ $\log(1/EC50)_{GR} = 1,005\log K_{ow} - 1,717$ $\log(1/EC50)_{FY} = 1,124\log K_{ow} + 0,0266(NoH)^4 - 1,569$ $\log(1/EC50)_{GR} = 1,111\log K_{ow} + 0,0267(NoH)^4 - 1,997$	0,986 0,998 0,921 0,965	0,943 0,997 0,896 0,955	0,080 0,030 0,230 0,147
6. Verhaari 1. klass Verhaari 2. klass Halogeenuid nitriilid [29]	18 53 15 24 6 6	(M)LR	$\log(1/NOEC)_{BP} = 0,8624\log K_{ow} - 1,0633$ $\log(1/EC10)_{BP} = 0,9349\log K_{ow} - 1,0380$ $\log(1/NOEC)_{BP} = 0,8786\log K_{ow} - 0,5927E_{lumo} + 0,0188$ $\log(1/EC10)_{BP} = 0,9599\log K_{ow} - 1,3208E_{lumo} - 0,360$ $\log(1/NOEC)_{BP} = 2,2283\log K_{ow} - 4,717E_{lumo} - 0,639$ $\log(1/EC10)_{BP} = 2,2886\log K_{ow} - 4,8790E_{lumo} - 0,742$	0,897 0,859 0,876 0,766 0,876 0,850	0,857 0,845 0,572 0,595 -	- - - -
7. 1° propargüülalkoholid 2° propargüülalkoholid 3° propargüülalkoholid [37]	15 7 8	(M)LR	$\log(1/EC50)_{FY} = 0,57\log K_{ow} - 0,37$ $\log(1/EC50)_{GR} = 0,6\log K_{ow} - 0,9$ $\log(1/EC50)_{FY} = 2,20\log K_{ow} - 1,10E_{lumo} - 1,56\log K_{ow}E_{lumo} + 3,40$ $\log(1/EC50)_{GR} = 0,0014\log K_{ow} - 2,10E_{lumo} - 0,024\log K_{ow}E_{lumo} + 4,00$ $\log(1/EC50)_{FY} = 1,025\log K_{ow} - 1,46$ $\log(1/EC50)_{GR} = 1,01\log K_{ow} - 1,96$	0,90 0,76 0,93 0,85 0,95 0,97	0,87 0,69 -	0,25 0,50 0,22 0,25 0,20 0,16
8. Verhaari 1. klass Verhaari 1. ja 2. klass [30]	50 87 / 21	(M)LR	$\log(1/EC50)_{GR} = 0,8763\log K_{ow-mix} - 1,799$ $\log(1/EC50)_{GR} = 0,5088\log K_{ow-mix} + 0,1344DHf/\#atoms + 0,009526MW - 1,654$ $\log(1/EC50)_{GR} = 0,7486\log K_{ow-mix} - 0,2302E_{lumo} - 1,059$	0,9469 0,9149	0,9426 0,9061	0,0887 0,1073
9. Aniliinid ja fenoolid [38]	13–45	aug-MIA		0,84	0,60	-
10. Verhaari 1. klass Verhaari 2. klass [31]	65 29	(M)LR	$\log(1/EC50) = 1,01\log P(1) - 1,99$ $\log(1/EC50) = 0,95\log P(1) - 0,96$ $\log(1/EC50) = 0,86\log P(1) + 0,016HF - 1,03$	0,76 0,71 0,86	0,73 0,65 0,84	0,52 0,49 0,35
11. Antibiootikumid [25]	13	MLR	$\log(1/EC50) = -0,663E_{lumo} - 0,00766Hf + 0,295\log K_{ow} + 2,96$	0,91	-	0,52
12. Verhaari 1. klass Verhaari 2. klass Verhaari 3. klass Verhaari 3. klass Verhaari 4. klass Verhaari 5. klass Verhaari 5. klass [32]	67 / 28 61 / 25 10 / 5 9 / 4 8 / 4 73 / 26 20 / 8	(M)LR	$\log(1/EC50) = 0,694\log K_{ow} + 0,581MATS7i - 1,828t*10^3 - -0,829MATS3p + 0,009TPSA(Tot) + 72,141Vs + 1,589$ $\log(1/EC50) = 0,544\log K_{ow} + 14,632Vs;min + 1,813qc + 0,426G3p + 0,007\alpha - 0,410Mor32i + 0,001Vs;max + 1,072$ $\log(1/EC50) = 4,438GATS1e + 8,769SpMin1_Bh(p) - 20,714Vs + -15,278$ $\log(1/EC50) = 0,823\log K_{ow} + 1,876$ $\log(1/EC50) = 102,518R6v - 2,478GATS3i + 9,369$ $\log(1/EC50) = 0,009\alpha + 0,282\log K_{ow} - 6,951GGI9 + 3,289DLS_02 - 2,851SpMAD_AEA(dm) - 0,997Mor10e - 1,429Gli + 5,379$ $\log(1/EC50) = 20,719E_{lumo} - 4826t*10^3 - 0,012Vs;max - -4,668HATS1i + 14,009$	0,829 0,827 0,891 0,798 0,925 0,613 0,818	0,880 0,815 0,960 0,782 0,910 0,550 0,723	- - - - - -
13. 1° ja 2° amiinid 3° amiinid [32]	5 4	LR	$\log(1/EC50) = -0,45\log K_{ow} - 4,12$ $\log(1/EC50) = -0,79\log K_{ow} - 1,81$	0,84 0,92	- -	0,15 0,31

## Lisa 1 jätk

Ühendid	$N_{treening}/N_{test}$	Meetod	Võrrand	$R^2$	$Q^2$	$\sigma$
14. Orgaanilised ained [39]	655 / 218	MLR		0,813	0,756	-
		PLS		0,798	0,760	-
		BPANN		0,927	0,734	-
		SVM		0,854	0,825	-
		LSSVM		0,839	0,774	-
		RF		0,827	0,772	-
GP		0,872	0,830	-	-	
15. Orgaanilised ained [44]	68–425 / 12–65	EL-DTB	VP-2, MDEC-22, BCUTp-1I, WL-2U, WL-3U, PPSA-1, PNSA-1, XlogP,	-	-	-
		EL-DTF	nAtomP	-	-	-
16. Orgaanilised ained [40]	74 / 31	DTB	SP-1, MDEC-23, MDEC-33, ATSm5, ATSp5, TopoPSA, nHBDon,	0,918	-	0,98
		DTF	XlogP	0,905	-	0,89
		MLR-PLS	$pEC_{50} = -3,25199 + 0,40376x^2\chi + 3,44722density - 2,86449[\sum\alpha]_p/\sum\alpha + +0,65341Atype\_N\_74 + 0,81062Atype\_C\_36$	0,774	0,708	-
MLR-PLS	$pEC_{50} = -3,209 + 0,340x^2\chi + 0,098logKow + 3,434density - -2,977[\sum\alpha]_p/\sum\alpha + 0,669Atype\_N\_74 + 0,978Atype\_C\_36$	0,782	0,705	-		
LDA	$DF = 13,0420 - 1,9488 S\_aaaC + 1,1885Atype\_C\_24 - 4,1799x^2\chi_v + +0,1964MW$	-	-	-		
17. Kosmeetika [21]	20 / 10	GFA-PLS	$pEC_{50} = -12,2563 - 0,1086 * < 16,53 - VR3\_Dt > +2,1959 * < GATS6s - 0,393536 > +32,9320 * ETA\_AlphaP + 0,1996 * AATS0s$	0,885	0,811	-
18. Orgaanilised ained [41]	251 / 83	GA-PLS	$pEC_{50} = 3,67917 + 0,02888CrippenMR + 0,39725MLOGP + +0,60323B06[C-N] + 0,78557B05[C-Cl] + 1,16506B02[N-S] - -1,46683nSO2OH$	0,72	0,70	-
			$pEC_{50} = 3,90575 + 0,02754CrippenMR + 0,2973logKow + +0,67324B06[C-N] + 0,83114B05[C-Cl] + 1,12005B02[N-S] - -1,05917nSO2OH$	0,72	0,70	-
			$pEC_{50} = 3,90595 + 0,03373CrippenMR + 0,28608LogKow + +0,96872B05[C-Cl] + 1,07463B02[N-S] - 0,20679H-051 - -1,23943nSO2OH$	0,72	0,70	-
			$pEC_{50} = 3,943 + 0,02707CrippenMR + 0,30156LogKow + +0,67215B06[C-N] + 0,81936B05[C-Cl] + 0,61348F02[N-S] - -1,15045nSO2OH$	0,71	0,70	-
			$pEC_{50} = 3,92235 + 0,03364CrippenMR + 0,28524LogKow + +0,96588B05[C-Cl] + 0,65962F02[N-S] - 0,20618H-051 - -1,23579nSO2OH$	0,71	0,70	-
			19. Ravimid [23]	53 / 16	GA-PLS	$pEC_{50} = -0,16674 + 0,56807F08[O-O] - 0,01791UNIP + 1,62082B08[C-C] + 7,88502PW4 + 0,02285ALOGP2$
			$pEC_{50} = -0,23950 + 0,28416F08[O-O] + 8,97627PW4 - 0,50834nR=Cs + +0,37153F06[C-Cl] + 0,73990CATS2D\_04\_DD$	0,69	0,62	-
			$pEC_{50} = -1,51547 + 0,53696F08[O-O] + 0,96070piPC1 - 0,01877UNIP + 1,09609B08[C-C] + 0,33749F06[C-Cl]$	0,73	0,63	-
			$pEC_{50} = -0,31362 + 0,54673F08[O-O] - 0,01643UNIP + 1,07024B08[C-C] + 0,38995piPC4 + 0,32394F06[C-Cl]$	0,74	0,65	-
			$pEC_{50} = -0,691319 + 0,515657F08[OO] + 0,530632piPC1 + +0,243608piPC8 - 0,015819UNIP + 0,334861F06[C-Cl]$	0,71	0,63	-
20. Orgaanilised ained [43]	334	MLR	$pEC_{10} = 1,861 + 0,055CrippenMR$	0,455	-	-
			$pEC_{10} = 1,273 + 0,042CrippenMR - 0,418MHYD$	0,636	-	-
			$pEC_{10} = 1,266 + 0,041CrippenMR - 0,412MHYD + 0,908F02[N-S]$	0,681	-	-
			$pEC_{10} = 1,332 + 0,038CrippenMR - 0,403MHYD + 0,970F02[N-S] + +0,927B05[C-Cl]$	0,707	-	-
			$pEC_{10} = 1,733 + 0,026CrippenMR - 0,388MHYD + 0,930F02[N-S] + +0,950B05[C-Cl] + 0,435CCCN$	0,728	-	-
			$pEC_{10} = 1,579 + 0,027CrippenMR - 0,370MHYD + 0,927F02[N-S] + +0,768B05[C-Cl] + 0,497CCCN + 1,017R3m$	0,751	-	-
	167 / 167	SVM		0,75–0,76	-	-
21. Orgaanilised ained [42]	217 / 54	GA-MLR	$pEC_{50} = 3,298 + 0,186nar-oCpcplus + 0,479fnotringNsp3C1B + +1,565fClaroC5B + 0,104nTRing + 0,326XLogP + 1,029SubFP96 - -2,439sp2Csp3C8B$	0,718	0,700	0,615
22. Orgaanilised ained [45]	113–118 / 52–85	CORAL		0,8341–0,8866	0,8289–0,8826	0,533–0,683
23. Pestitsiidid [22]	163–169 / 70–76	GA-OLS-MLR	$pEC_{50} = 6,532 + 0,285MLFER\_L - 1,335ExtFP300 - 11,058ASP-4 - -1,533GATS3m - 0,762ExtFP29 - 0,148minHBint2$	0,651	0,616	-
			$pEC_{50} = 6,277 + 0,294MLFER\_L - 1,389ExtFP300 - 10,451ASP-4 - -0,851ExtFP - 1,409Gats3m - 0,143minHBint2$	0,655	0,623	-
			$pEC_{50} = 6,267 + 0,308MLFER\_L - 1,457ExtFP300 - 11,153ASP-4 - -1,433GATS3m - 0,776ExtFP29 - 0,131minHBint2$	0,665	0,635	-

Lühendite selgitused:  $R^2$  – determinatsioonikordaja,  $Q^2$  – ristvalideerimiskordaja,  $\sigma$  – standardhälve,  $N_{treening}$  – treeningseeria ühendite arv,  $N_{test}$  – testseeria ühendite arv. Lühendid *DO* – lahustunud hapniku toodang, *GR* – kasvukiirus, *FY* – lõplik saagis, *BP* – biopopulatsioon viitavad kasutatud meetodile ühendi  $EC_{50}$  määramisel. Kasutatud meetoditest ja molekulaartunnustest saab lugeda originaalartiklitest

## Lisa 2. Modelleerimisel kasutatud keemiliste ühendite loend

Lühendite selgitused:  $pEC_{50;eksp}$  – eksperimentaalselt saadud ühendi  $EC_{50}$  väärtus,  $pEC_{50;arv}$  – Verhaari klassile tehtud mudeliga ennustatud  $EC_{50}$  väärtus,  $pEC_{50;suur mudel}$  – kogu andmeseeriale tehtud mudeliga ennustatud  $EC_{50}$  väärtus; tärniga (\*) on tähistatud ühendid, mis kuulusid kogu andmeseeria modelleerimisel valideerimiseeriasse

Nr	Nimetus	CAS	$pEC_{50;eksp}$	$pEC_{50;arv}$	$pEC_{50;suur mudel}$
<b>Verhaari 1. klass</b>					
<b>Treeningseeria</b>					
1	4-bromodifenüüleeter	101-55-3	2.93	3.06	1.85
2	Difenüüleeter	101-84-8	2.14	1.70	1.33
3	Tribütüülamiin	102-82-9	0.47	0.94	1.77
6	Metüülpropüülketoon	107-87-9	-1.34	-0.89	-0.54*
7	N,N,N',N'-tetrametüül-1,3-propaandiamiin	110-95-2	0.58	0.57	-0.22
8	Diisobütüülamiin	110-96-3	0.91	0.82	1.18
9	N,N,N',N'-tetrametüül-1,6-heksaandiamiin	111-18-2	1.53	1.72	0.79
10	5-kloropentaanhape	1119-46-6	-0.21	-0.58	0.02
11	Ditsüklopropüülketoon	1121-37-5	-0.73	-0.68	-0.99
13	1-bromodekaan	112-29-8	2.67	2.17	3.41
15	1,2-bis(4-metoksüfenüül)etaan-1,2-dioon	1226-42-2	1.46	2.00	2.02
19	4,4'-dibromodifenüüleeter	2050-47-7	3.14	2.85	1.82*
20	4-fenüül-2-butanool	2550-26-7	0.61	0.42	0.63
21	4-bromobutaanhape	2623-87-2	-0.40	-0.84	-0.13
22	6-bromoheksaanhape	4224-70-8	0.52	-0.19	0.64
23	1,8-dibromooktaan	4549-32-0	1.87	2.15	2.84
24	1,4-(N-dimetüülamino)-2-buteen	4559-79-9	1.51	0.79	0.27*
25	7-tridekanool	462-18-0	2.24	1.83	2.34
26	3,4-bensopüreen	50-32-8	5.17	4.18	3.72
27	2-atsetüül-5-norborneen	5063-03-6	0.12	0.33	0.17
29	Benso(a)antratseen	56-55-3	2.90	3.85	3.16
30	2-heksanool	591-78-6	-0.78	-0.54	-0.08
31	Tripentüülamiin	621-77-2	1.74	1.86	2.53
32	4-klorobutaanhape	627-00-9	-1.17	-0.92	0.03
34	Atsetoon	67-64-1	-2.04	-2.08	-0.97
35	Kloroform	67-66-3	-0.07	0.28	-0.71
36	2-adamantanool	700-58-3	-0.15	0.53	0.48
37	Diklorometaan	75-09-2	-0.36	-0.65	-0.16
38	2-metüül-2-propanool	75-65-0	-0.28	-0.59	-0.93
40	Naftaleen	91-20-3	1.11	1.25	0.64
41	2,4-dimetoksübensoehape	91-52-1	-0.27	0.34	1.57
42	1,2-dimetüülbenseen	95-47-6	0.80	1.14	0.82*
43	1,2,4-trimetüülbenseen	95-63-6	1.09	1.17	1.39
44	1,2,4,5-tetrametüülbenseen	95-93-2	1.16	0.83	1.98
<b>Valideerimiseeria</b>					
4	4-(4-metoksüfenüül)-2-butanool	104-20-1	0,64	0,99	1,49
5	3-oktanool	106-68-3	0,18	0,42	0,57
12	N,N-dimeüüldodetsüülamiin	112-18-5	3,90	4,22	1,99
14	Trikloroetanool	115-20-8	-0,51	-0,22	0,47
16	4-heptanool	123-19-3	-0,27	-0,59	0,16
18	Diheksüülamiin	143-16-8	2,27	1,98	2,78
28	Allüületüüleeter	557-31-3	1,11	-0,19	1,57
33	1,6-dibromoheksaan	629-03-8	1,57	1,55	2,54
39	2-butanool	78-92-2	-1,65	-0,82	-0,79*

Nr	Nimetus	CAS	<i>pEC</i> <sub>50;eksp</sub>	<i>pEC</i> <sub>50;arv</sub>	<i>pEC</i> <sub>50;suur mudel</sub>
<b>Verhaari 2. klass</b>					
<b>Treeningseeria</b>					
46	2-etüülheksüülamiin	104-75-6	0,65	0,74	1,09
48	Propüülamiin	107-10-8	-0,11	-0,69	-0,55
49	2-propeen-1-amiin	107-11-9	-0,91	-0,93	0,77
50	3-metüülbutüülamiin	107-85-7	-0,65	-0,48	-0,32
53	Heksüülamiin	111-26-2	0,61	0,66	0,46
54	Heptüülamiin	111-68-2	1,31	1,24	0,83
55	Oktüülamiin	111-86-4	1,86	1,74	1,21
56	Nonüülamiin	112-20-9	2,51	2,20	1,59
57	N-etüül-naftaleen-1-amiin	118-44-5	1,37	1,51	2,51*
58	2-aminoetüüldiisopropüülamiin	121-05-1	-0,43	-0,14	-0,26
59	2-kloro-4-nitroaniliin	121-87-9	1,06	1,13	1,17*
60	Difenüülamiin	122-39-4	1,95	1,42	1,13
61	2-heptüülamiin	123-82-0	0,43	1,07	0,70
62	N1-(3-aminopropüül)butaan-1,4-diamiin	124-20-9	0,39	0,51	0,30*
63	N-dodetsüülamiin	124-22-1	3,52	3,49	2,66
64	Tetradetsüülamiin	2016-42-2	4,44	4,30	3,33*
65	Detsüülamiin	2016-57-1	3,28	2,64	1,95
67	Tridetsüülamiin	2869-34-3	3,57	3,90	3,00
68	4-heksüülaniliin	33228-45-4	2,42	2,48	2,52
70	4-detsüülaniliin	37529-30-9	3,60	3,91	3,53
71	4,6-dikloro-5-nitropürimidiin	4316-93-2	1,44	1,45	1,71*
72	Pentaan-1,5-diamiin	462-94-2	-0,33	0,16	-0,45*
73	6-metüül-2-heptüülamiin	543-82-8	0,77	0,74	1,04
74	Dietüülstilbestrool	56-53-1	2,05	2,20	3,12
75	Aniliin	62-53-3	0,12	-0,10	-0,62
77	3,4,5-trikloroaniliin	634-91-3	1,74	1,76	1,34
78	N,N'-bis(3-aminopropüül)-1,4-butaandiamiin	71-44-3	1,68	1,08	1,21
80	2-nitroaniliin	88-74-4	0,31	0,72	0,59
81	2-kloroaniliin	95-51-2	0,85	0,82	0,28
82	3,4-dikloroaniliin	95-76-1	1,25	1,21	1,11
<b>Valideerimiseseeria</b>					
45	4-nitroaniliin	100-01-6	0,51	0,69	0,18*
47	4-metüülaniliin	106-49-0	1,44	-0,16	0,55
51	Butüülamiin	109-73-9	0,04	-0,20	-0,31
52	1,4-diaminobutaan	110-60-1	-0,77	-0,29	-0,77*
66	2,6-diisopropüülaniliin	24544-04-5	0,99	0,83	2,39*
69	4-heptüülaniliin	37529-27-4	2,74	2,83	2,73*
76	2,3,4-trikloroaniliin	634-67-3	1,93	1,85	1,30*
79	N-undetsüülamiin	7307-55-3	3,87	3,07	2,31
<b>Verhaari 3. klass</b>					
<b>Treeningseeria</b>					
83	Asobenseen	103-33-3	2,04	2,67	2,34
84	1,2-epoksü-5-hekseen	10353-53-4	0,37	0,13	0,51
85	Fenüülsotiotsünaat	103-72-0	2,64	2,56	1,16*
86	Tsinnaamüülalkohol	104-54-1	0,10	0,72	-0,33
87	Tertbutüülkloroatsetaat	107-59-5	2,51	1,56	1,27
90	Heptanaal	111-71-7	0,80	0,63	0,31
91	Dekanoüülkloriid	112-13-0	1,08	1,54	2,82
92	Tsükloheksüülsotiotsünaat	1122-82-3	2,42	1,93	1,77*
93	2-(2-butoksüetoksü)etüülsotiotsünaat	112-56-1	1,23	1,18	2,05
94	Metüüldikloroatsetaat	116-54-1	0,98	0,58	0,62

Nr	Nimetus	CAS	<i>pEC</i> <sub>50;eksp</sub>	<i>pEC</i> <sub>50;arv</sub>	<i>pEC</i> <sub>50;suur mudel</sub>
96	2-hüdroksu-1,2-bis(4-metoksüfenüül)etaan-1-oon	119-52-8	1,13	1,56	1,87
98	4-metoksübensaldehüüd	123-11-5	0,14	0,59	0,06*
100	4-nitrobensüültsünaat	13287-49-5	3,40	2,85	1,74
101	Fenasopüridiin	136-40-3	2,58	1,90	1,72
102	2-fluoroetaanhape	144-49-0	3,75	3,75	2,25
103	Etoksükarbonüülsotiotsünaat	16182-04-0	0,52	-0,17	1,26
104	1-penteen-3-oon	1629-58-9	2,97	2,45	1,04
105	1-hekseen-3-oon	1629-60-3	2,55	2,56	1,16
106	2,2-dikloropentaanhape	18240-68-1	-0,30	-0,08	0,94
109	Isopropüülsotiotsinnaamat	2253-73-8	2,14	3,16	1,30
110	N-(butoksümetüül)-2-kloro-N-(2,6-dietüülfenüül)atsetamiid	23184-66-9	4,75	4,64	4,38*
111	2-(dimetüülamino)etüülakrülaad	2439-35-2	1,95	1,16	0,84
112	Diisopropüül asodikarboksülaad	2446-83-5	-0,03	0,15	2,12*
114	Tsinnamüülkloriid	2687-12-9	0,97	1,16	0,91
115	2-kloropropaanhape	28554-00-9	-1,52	-0,87	0,28
116	Allüülkloroetanaad	2916-14-5	2,60	2,82	1,89
117	3-(5-metüül-2-furüül)butanaal	31704-80-0	0,88	0,84	1,23*
118	2,4-dimetokspfentüülsotiotsünaat	33904-03-9	2,71	3,15	2,17
119	Atsetokloor	34256-82-1	4,48	3,75	3,31
121	Triadimefon	43121-43-3	1,43	1,05	1,81
123	2-kloro-1,2-difenüületanoon	447-31-4	1,66	1,62	2,84*
124	Metolakloor	51218-45-2	3,13	3,87	3,72
125	Pretilakloor	51218-49-6	4,68	4,97	4,23
126	Bensoüülsotiotsünaat	532-55-8	-0,10	0,75	0,34
130	Kloramfenikool	56-75-7	1,90	2,47	2,37
131	Allüülsotiotsünaat	57-06-7	2,85	1,87	2,54
132	Ksantureenhape	59-00-7	0,93	1,22	0,64*
133	Butüülsotiotsünaat	592-82-5	2,51	2,63	1,69*
134	Fenarimool	60168-88-9	1,55	1,04	2,89
135	N,N-difenüülformamiid	607-00-1	0,78	0,78	1,57
136	Klorotetratsükliin	64-72-2	3,61	3,54	3,27
137	Heksanaal	66-25-1	0,43	1,05	-0,04
138	1-naftaldehüüd	66-77-3	1,15	0,70	0,26*
139	2-naftaldehüüd	66-99-9	1,02	0,88	0,28
140	4-fenoksübensaldehüüd	67-36-7	1,68	0,90	1,55
141	2,2-dikloroatsetamiid	683-72-7	-0,55	0,02	0,63
143	N,N-dibutüülformamiid	761-65-9	-0,14	0,14	0,87
144	Tert-butüülformaad	762-75-4	-1,36	-1,32	-0,13
145	Metüülvinüülketoon	78-94-4	2,91	2,90	0,89*
146	2-propeenamiid	79-06-1	-0,66	-0,58	0,38
147	2-kloroetaanhape	79-11-8	2,93	2,53	0,73
148	2,5-dimetoksübensaldehüüd	93-02-7	0,60	0,68	1,25*
149	Allüülmetakrülaad	96-05-9	0,12	0,29	3,49*
<b>Valideerimisseeeria</b>					
88	2,2'-diklorodiisopropüüleeter	108-60-1	-0,11	2,19	1,23*
89	Fumaarhape	110-17-8	-1,01	-2,80	-1,08
97	2-etüülheksanaal	123-05-7	0,54	0,74	0,69
99	Butanaal	123-72-8	0,14	2,03	-0,63
107	Allüülsinnaamat	1866-31-5	2,16	2,61	2,94
108	Diallüüldisulfiid	2179-57-9	1,36	-1,12	1,46
113	3-(Dietüülamino)propüülsotiotsünaat	2626-52-0	2,56	2,33	0,47*
120	3-kloro-2-butanoon	4091-39-8	1,77	1,03	0,51
122	Tsinnamüülbromiid	4392-24-9	1,02	1,01	1,42

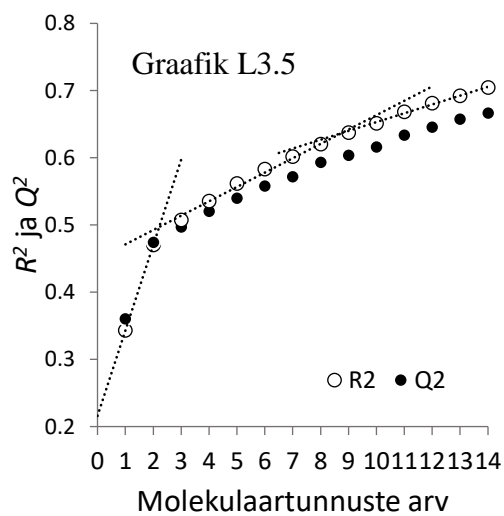
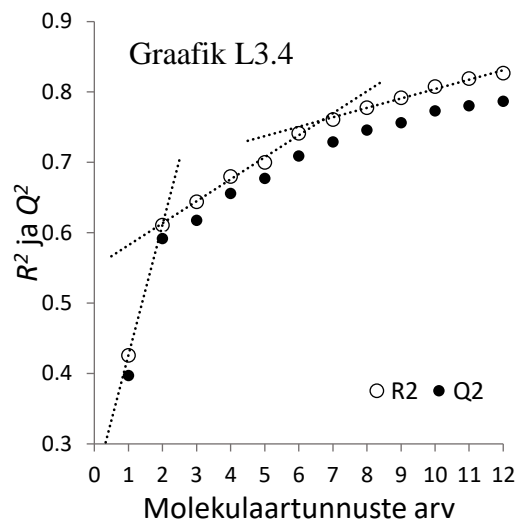
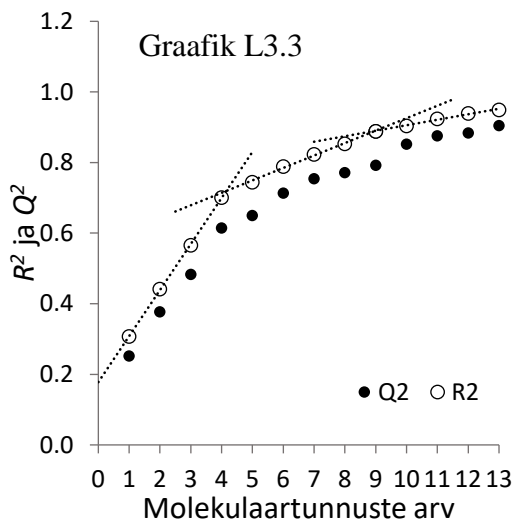
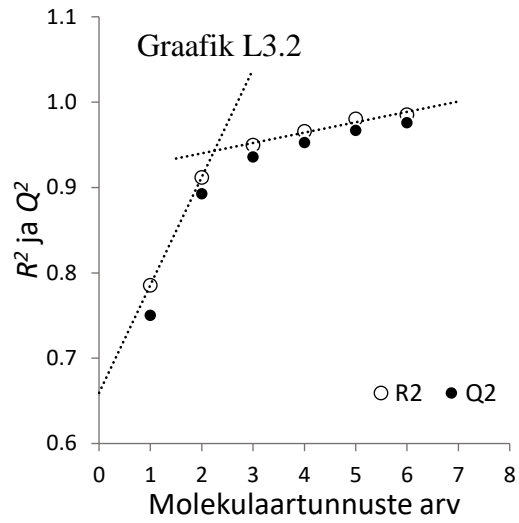
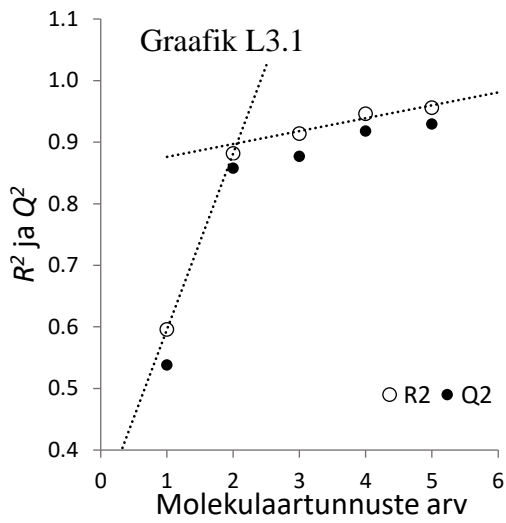
Nr	Nimetus	CAS	<i>pEC</i> <sub>50;eksp</sub>	<i>pEC</i> <sub>50;arv</sub>	<i>pEC</i> <sub>50;suur mudel</sub>
127	1,5-difenüülkarbasoon	538-62-5	2,79	3,72	2,83
128	1-naftüülisotiotsünaat	551-06-4	3,99	2,80	2,30*
129	N-(2-kloroetüül)dibensüülamiin	55-43-6	0,93	0,35	1,28
142	4-atsidofenüülisotiotsünaat	74261-65-7	3,12	1,99	1,89
<b>Verhaari 4. klass</b>					
150	Malatioon	121-75-5	1,14	-	0,68
152	2,4,6-triklorofenüülisotsünaat	22134-07-2	3,31	-	1,76*
153	Diklofentioon	97-17-6	2,89	-	2,79
<b>Verhaari 5. klass</b>					
<b>Treeningseeria</b>					
154	Bensonitriil	100-47-0	-0,10	0,42	0,12
156	4-(4-klorofenoksü)aniliin	101-79-1	2,55	2,17	2,60
157	Ditsükloheksüülamiin	101-83-7	1,52	0,83	2,45
159	1-kloro-4-jodo-butaan	10297-05-9	1,14	2,10	1,18
160	N-(-methüületüül)bensüülamiin	102-97-6	1,53	1,45	1,71
161	Tsinnamüülatsetaat	103-54-8	1,13	0,80	1,49*
162	Atsetaniliid	103-84-4	-0,34	-0,28	0,91
164	1-jodopropaan	107-08-4	0,75	1,76	1,62*
165	3-aminopropaanhape	107-95-9	-0,02	-1,20	-1,04
166	N-butüronitriil	109-74-0	-1,28	-1,08	0,33
167	Dipropüülsulfiid	111-47-7	0,32	0,70	1,14*
168	Trietüleetetramiin	112-24-3	1,44	1,09	-0,27
172	2,3,5,6-tetrakloronitrobenseen	117-18-0	2,47	2,72	2,09
173	2,3-dikloro-1,4-naftokinoon	117-80-6	3,03	2,90	1,60
174	Guanosiin	118-00-3	-0,96	0,48	-1,42
177	N,N-dimetüülaniliin	121-69-7	0,30	0,57	0,05
178	Simasiin	122-34-9	2,68	2,39	2,35
179	Kloorbromuroon	13360-45-7	3,82	3,33	2,67*
180	Triklorometiasiid	133-67-5	-0,20	0,66	2,29
181	Fenoksasiin	135-67-1	2,12	1,27	0,96
182	Propasiin	139-40-2	2,78	2,68	2,73
183	4'-(kloroatsetüül)atsetaniliid	140-49-8	2,79	1,59	1,84
184	Flufenatseet	142459-58-3	4,36	3,85	4,70
186	1-dekaantiool	143-10-2	2,69	2,33	3,52
188	4-atsetüülbensonitriil	1443-80-7	0,18	-0,82	0,13
189	Sulfapüridiin	144-83-2	0,12	1,19	0,56
190	4-aminobensoehape	150-13-0	-0,78	-0,47	-0,32
191	4-(4-aminofenüül)butaanhape	15118-60-2	-0,43	-0,10	0,61
192	Metoksüpropasiin	1610-18-0	2,83	2,74	3,03
194	N-metüüladenosiin	1867-73-8	2,11	1,06	-0,20*
195	4-(dipropüülamino)-3,5-dinitrobenseensulfoonamiid	19044-88-3	3,02	3,47	3,49*
196	Atrasiin	1912-24-9	3,13	2,53	2,65
197	Trietasiin	1912-26-1	1,02	2,75	2,09
198	4-kloro-3,5-dinitrobensonitriil	1930-72-9	2,48	2,59	2,66
200	2,4-dikloro-1-naftool	2050-76-2	3,18	2,66	2,00
201	Metribusiin	21087-64-9	3,79	4,15	3,41
202	Dimetametriin	22936-75-0	4,52	4,25	3,88
203	1-aminopentadekaan	2570-26-5	4,41	3,39	3,65
204	2-fenoksüaniliin	2688-84-8	1,35	1,70	2,00
205	3-aminometüül-3,5,5-trimetüülsükloheksüülamiin	2855-13-2	1,98	2,32	1,25
207	2-amino-5-metüülbensoehape	2941-78-8	3,10	1,71	1,18
208	L-karosiin	305-84-0	0,62	0,63	0,98*
209	Metobromuroon	3060-89-7	2,62	3,00	2,55

Nr	Nimetus	CAS	<i>pEC</i> <sub>50;eksp</sub>	<i>pEC</i> <sub>50;arv</sub>	<i>pEC</i> <sub>50;suur mudel</sub>
210	Tsükloheksaanmetüülamiin	3218-02-8	0,94	0,90	0,14
211	2-ketoglutaarhape	328-50-7	-0,82	-0,44	-1,27
212	N,N-dietüüldodekaanamiid	3352-87-2	3,32	3,22	2,94
215	2,3,5,6-tetrakloroaniliin	3481-20-7	2,32	2,37	1,05
216	5-bromo-2,4-dikloropürimidiin	36082-50-5	2,80	1,41	1,69*
219	4-(heksüüloksü)aniliin	39905-57-2	1,93	1,84	2,55
220	Penoksaliin	40487-42-1	3,39	3,69	4,78
221	1-aminoheptadekaan	4200-95-7	3,61	3,98	4,26*
222	Metüül-5-(2,4-diklorofenoksü)-2-nitrobensoaat	42576-02-3	6,11	3,94	4,41
224	6-metüüladeniin	443-72-1	0,51	0,69	0,76
226	N-metüüldioktüülamiin	4455-26-9	3,50	3,62	2,87
229	4-fluorobensoehape	456-22-4	0,23	-0,47	0,68
230	Asoksübenseen	495-48-7	2,17	2,56	2,55*
231	1,3-ditiaan	505-23-7	-0,60	0,10	0,71*
232	Alloksaan	50-71-5	-1,13	-1,02	-0,45*
235	Heksasinoon	51235-04-2	3,88	3,08	2,77
236	DL-penitsillamiin	52-66-4	0,79	0,26	0,60
237	2-metüül-4,6-dinitrofenool	534-52-1	2,18	1,24	2,36
239	Tsüklobütüülfenüülketoon	5407-98-7	1,16	1,53	0,62
240	Dibütüülsulfiid	544-40-1	1,30	1,67	2,01
241	Triklopüür	55335-06-3	0,11	2,46	1,16
242	Kloroheksidiin	55-56-1	4,05	4,49	4,67
243	3-jodo-1-propeen	556-56-9	2,43	1,23	3,13
244	4-metoksüfenetüülamiin	55-81-2	0,39	0,52	1,43
245	2,4-diamino-6-hüdroksüpürimidiin	56-06-4	-0,61	-0,58	-1,26
246	4-aminobutaanhape	56-12-2	-0,19	-0,70	-0,76*
247	Isometiosiin	57052-04-7	3,03	3,65	4,31
248	4-fluoro-1-naftoehape	573-03-5	0,32	0,51	1,71
249	Kofeiin	58-08-2	-0,25	1,07	0,09*
250	5-metüültsütosiin	58366-64-6	-1,38	-0,68	-0,23
251	3,5-dinitrofenool	586-11-8	2,30	1,63	1,28
253	Adenosiin-5'-difosfaat	58-64-0	0,70	-0,14	0,94
254	4-aminopürimidiin	591-54-8	-1,04	-0,61	-0,12
255	2-hüdroksükinioliin	59-31-4	-0,43	-0,76	-0,26*
256	3-amino-2-naftoehape	5959-52-4	0,83	1,02	1,53
258	Tribütüriin	60-01-5	0,43	0,73	2,51*
259	Propikonasool	60207-90-1	1,47	3,52	1,58
260	6-aminoheksaanhape	60-32-2	-1,83	-1,13	-0,10
261	Floretiin	60-82-2	0,91	1,89	1,32
263	Fenüülmetüülnitrosamiin	614-00-6	0,53	0,15	0,36*
264	Zoksasolamiin	61-80-3	0,87	0,95	0,93
265	Metüül-4-aminobensoaat	619-45-4	0,23	0,29	0,62*
266	Benzamide, 4-nitro-	619-80-7	2,00	1,96	0,80
267	Adrenaloon	62-13-5	1,83	1,08	1,41*
269	N-(4-etoksüfenüül)atsetamiid	62-44-2	0,05	0,34	1,96*
270	1-fenüüluurea	64-10-8	-0,03	0,19	0,20
271	2-jodoetaanhape	64-69-7	1,57	1,25	1,28
272	Klorosulfuroon	64902-72-3	3,09	2,08	2,74*
273	Püridoksaal	65-22-5	0,45	0,84	-0,05
274	2'-desoksüadenosiin-5'-monofosfaat	653-63-4	0,13	-0,20	-0,10
275	Salitsüülamiid	65-45-2	0,20	0,75	0,36
276	Küromasiin	66215-27-8	0,55	0,33	0,98
277	Uratsill	66-22-8	-1,70	-1,19	-0,37
280	Dimetüülsulfoksiid	67-68-5	-2,69	-1,96	-1,22

Nr	Nimetus	CAS	<i>pEC</i> <sub>50;eksp</sub>	<i>pEC</i> <sub>50;arv</sub>	<i>pEC</i> <sub>50;suur mudel</sub>
281	N,N-dietüülätsetamiid	685-91-6	-1,31	-1,44	-0,56
283	Propaniil	709-98-8	3,44	3,31	2,08
284	L-histidiin	71-00-1	-0,87	-0,90	-0,004*
285	Prometriin	7287-19-6	3,96	3,46	3,44
286	6-aminopuriin	73-24-5	0,64	-0,06	-0,04
288	2-amino-6-hüdroksüpuriiin	73-40-5	0,46	0,19	-0,34
289	2,3-naftaleendiamiin	771-97-1	1,55	1,20	1,48*
290	Metüüliteensulfuroon	79277-27-3	2,29	3,03	3,37*
291	N-metüül-N-nitroso-p-tolueensulfoonamiid	80-11-5	1,83	1,14	0,36
292	1,3-dimetüül-2-imidasolidinooon	80-73-9	-1,50	-1,43	-0,68
293	Naftaleen-1,5-disulfoonhape	81-04-9	-1,73	-1,42	-1,08
294	1-amino-4-bromo-2-metüülantrakinoon	81-50-5	4,48	4,26	4,27
295	2-klorotsükloheksanoon	822-87-7	1,14	0,38	0,51
296	4-nitro-N-fenüülaniliin	836-30-6	2,00	2,00	1,99*
297	Riboflaviin	83-88-5	1,50	1,50	2,04
299	Salitsüülaniliid	87-17-2	2,29	2,01	2,54
300	Inositool	87-89-8	-1,49	-1,06	-1,37*
301	2,5-diaminobenseensulfoonhape	88-45-9	0,59	0,34	-0,66
303	N-fenüülantraniihape	91-40-7	1,38	1,84	1,43
304	N,N-dietüültsükloheksaanamiin	91-65-6	0,35	0,87	0,68
305	N,N-dietüülaniliin	91-66-7	1,61	0,49	0,73
306	6-dimetüülaminopuriin	938-55-6	0,60	1,40	0,99
307	Fenüülbensoaat	93-99-2	1,74	2,48	1,79*
308	Etüül-4-aminobensoaat	94-09-7	0,39	1,34	1,31
310	Propüülparabeen	94-13-3	3,03	1,76	1,42
312	Butüülparabeen	94-26-8	3,29	1,97	1,71
313	5-kloro-2-metoksüaniliin	95-03-4	0,94	1,00	0,77
314	2,4-diaminotolueen	95-80-7	0,00	0,56	0,82
315	Tsüklopentanool	96-41-3	-1,77	-1,16	-1,11
316	Allantoiin	97-59-6	-1,40	-2,14	-0,42
318	Atsetofenoon	98-86-2	-0,22	-0,15	-0,15
319	2-kloro-3',4'-dihüdroksüatsetofenoon	99-40-1	1,81	2,21	0,96
320	N-atsetüül-L-asparagiinhape	997-55-7	0,29	0,28	0,23
321	Metüülparabeen	99-76-3	-1,59	-1,30	0,22
<b>Valideerimisseeria</b>					
155	Anilasiin	101-05-3	2,19	3,33	2,01
158	2-(dibutüülamino)etanool	102-81-8	0,29	1,48	0,70
163	4-dodetsüülaniliin	104-42-7	3,48	3,31	4,11*
169	2-atsetüülpüridiin	1122-62-9	-0,83	-0,12	-0,43
170	Undetsüleenhape	112-38-9	1,74	1,50	2,21
175	3,3'-dimetoksübensidiin	119-90-4	1,27	2,40	2,95*
176	Etüülparabeen	120-47-8	2,62	1,17	0,98
185	Oktüül-4-aminobensoaat	14309-41-2	2,96	2,62	3,00
187	Heksadetsüülamiin	143-27-1	3,81	3,69	3,96
193	Monolinuroon	1746-81-2	2,76	2,22	2,28
199	2,6-diklorobensamiid	2008-58-4	0,62	1,75	0,52
206	1,3,5-triasiin	290-87-9	-0,47	-0,63	0,50
213	2-(1-tsükloheksenüül)etüülamiin	3399-73-3	0,34	-0,47	0,37
214	Isoproturoon	34123-59-6	3,27	2,53	2,42
217	3-atsetüül-2,5-diklorotiofeen	36157-40-1	1,46	1,47	1,68
218	Tsitulliin	372-75-8	-1,60	-0,14	0,00
223	2-fenüületaantiool	4410-99-5	2,10	0,40	1,37
225	4-tsükloheksüülbutaanhape	4441-63-8	0,53	-0,70	0,77
227	2,5-dihüdroksüfenüületaanhape	451-13-8	1,05	0,07	0,03
228	Etüülbutüülnitrosamiin	4549-44-4	0,40	0,26	0,26

Nr	Nimetus	CAS	<i>pEC</i> <sub>50;eksp</sub>	<i>pEC</i> <sub>50;arv</sub>	<i>pEC</i> <sub>50;suur mudel</sub>
233	Atsetüülsalitsüülhape	50-78-2	-0,13	0,01	0,91
234	2-(dietüülamino)etüül-4-aminobensoaat	51-05-8	-1,17	1,74	0,93
238	2-atsetüülaminofluoreen	53-96-3	1,92	0,45	2,85
252	Inosiin	58-63-9	-1,47	0,22	-1,55
257	4-nitrosomorfoliin	59-89-2	0,14	-1,57	-1,17
262	2,5-dimetüülbenseensulfoonhape	609-54-1	-1,96	0,70	-0,45*
268	4-nitrobensoehape	62-23-7	0,76	1,68	0,41
279	Fenpropidiin	67306-00-7	2,36	3,37	2,56
282	2,6,8-trioksüpuriiin	69-93-2	0,01	-0,46	-0,22
287	Mefenatseet	73250-68-7	3,08	2,01	3,22
298	Difenüülnitrosamiin	86-30-6	0,92	2,37	1,19
302	Etüüklorimuroon	90982-32-4	5,36	3,49	3,80
309	Prosulfuroon	94125-34-5	4,19	2,76	4,28
311	Butüül-4-aminobensoaat	94-25-7	1,54	1,84	2,09
317	p-etüülbenseensulfoonhape	98-69-1	-0,25	-1,27	-0,63

**Lisa 3. Graafikud, mis kirjeldavad  $R^2$  ja  $Q^2$  muutust molekulaartunnuste lisamisel**



## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Iris Tamsalu,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Kvantitatiivsed struktuuri-aktiivsuse sõltuvused *Pseudokirchneriella subcapitata* kasvu pärssivatele ühenditele“, mille juhendaja on Uko Maran, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Iris Tamsalu*

**24.05.2023**