

Tartu Ülikool
Füüsika-keemiateaduskond
Keemilise füüsika instituut

VILLU RUUSMANN

**VESIKESKKONNAS ESINEVA AKUUTSE TOKSILISUSE
KVALITATIIVSED JA KVANTITATIIVSED ASPEKTID ALGLOOMA
Tetrahymena pyriformis NÄITEL**

Magistritöö
Teoreetilise ja arvutikeemia eriala

Juhendaja: UKO MARAN

Tartu 2006

Sisukord

1 Sissejuhatus.....	4
2 Kirjanduse ülevaade.....	5
2.1 Klassifikatsioonipuu.....	5
2.2 Alifaatsed ühendid.....	8
2.3 Aromaatsete ühendid.....	9
3 Materjalid ja meetodid.....	12
3.1 Keemilised ja bioloogilised andmed.....	12
3.2 Kvalitatiivne analüüs.....	13
3.2.1 SOM - iseorganiseeruv närvivõrk.....	14
3.2.2 Algoritm.....	14
3.2.3 Klassid ja klassipiirid.....	17
3.3 Kvantitatiivne analüüs.....	18
3.3.1 Toimeprofili kontseptsioon.....	19
3.3.2 Baassirge toksilisuse kontseptsioon ja liigtoksilisus.....	20
3.3.3 Regressioonianalüüs.....	21
4 Tulemused ja arutelu.....	23
4.1 Alifaatsete ja aromaatsete ühendite andmekomplekt.....	23
4.1.1 Klassifitseerimine iseorganiseeruva närvivõrguga.....	23
4.1.2 Leitud klasside regressioonianalüüs.....	27
4.2 Aromaatsete ühendite laiendatud andmekomplekt.....	32
4.2.1 Klassifitseerimine iseorganiseeruva närvivõrguga.....	33
4.2.2 Leitud klasside regressioonianalüüs.....	36
5 Kokkuvõte.....	39
6 Summary.....	41
7 Kasutatud kirjandus.....	43
8 Lisad.....	47

LÜHENDID

- ADH Alkoholi dehüdrogenaas (*alcohol dehydrogenase*).
- BA Bioloogiline aktiivsus (*biological activity*).
- BMLR Parim multilineaarne regressioon (*best multi-linear regression*).
- EDG Elektronidonoorne (-loovutav) rühm (*electron-donating group*).
- EWG Elektroniaktseptoorne (-tõmbav) rühm (*electron-withdrawing group*).
- FATS Kalade akuutse toksilisuse sündroom (*fish acute toxicity syndrome*).
- HOMO Kõrgeim hõivatud molekularbitaal (*highest occupied molecular orbital*).
- LSER Solvatatsioonenergia lineaarsõltuvus (*linear solvation energy relationship*).
- LUMO Madalaim vaba molekularbitaal (*lowest unoccupied molecular orbital*).
- MOA Toime (*mode of action*).
- NSI Neuronite sarnasusindeks (*neuron similarity index*).
- QSAR Kvantitatiivne struktuuri-aktiivsuse sõltuvus (*quantitative structure-activity relationship*).
- SOM Iseorganiseeruv närvivõrk (*self-organising (feature topology) map*).
- WARU Nõrgast hapest lahutusfaktor (*weak acid respiratory uncoupler*).

1 SISSEJUHATUS

Kvantitatiivsed struktuuri-aktiivsuse sõltuvused on leidnud tee väga erinevate valdkondadesse. Toksikoloogia probleemide lahendamises on see tehnika märgatav olnud alates möödunud sajandi kaheksakümnendate aastate esimesest poolest. Tänapäevaks on vähem või rohkem põhjalikult läbi uuritud enamuse keemilisi klasse. Tehnika väärtusest ja potentsiaalset räägib kõnekalt fakt, et nii Ameerika Ühendriigid [1] kui Euroopa Liit [2, 3] on selle heaks kiitnud oma keskkonnaprogrammides.

Testsüsteemide arendamine ja lihtsustamine on loonud olukorra, kus varasemate kümnetest ühenditest koosnevate andmekomplektide asemel on kättesaadavad sadadest ja tuhandetest ühenditest koosnevad andmekomplektid. Suurenenud andmehulgad on tekitanud vajaduse ja võimaluse keemiliste ühendite ruumi põhjalikumaks uurimiseks, mis omakorda tähendab, et kvalitatiivne hüpe peab toimuma ka mudelite koostamise ja tõlgendamise osas. Andmete esitluskuju peab muutuma inimest asendama tulevatele ekspertüsteemidele kohasemaks, abstraktsioonitase peab tõusma üksikühendite juurest vähemalt ühesuguste omadustega ühendite hulkaeni.

Käesolevas töös uuritakse akuutset toksilisust alglooma *T. pyriformis* näitel. Mehhanismi seisukohast interpreteeritavate ja korduvkasutatavate mudelite tuletamiseks kavandatakse ja realiseeritakse (pool)automaatne kahest osast koosnev töövoog. Kirjanduse alusel koostatud mehhanismide klassifikatsioonipuu jagab alifaatsete ja aromaatsete ühendite toksilisuse mehhanismid nelja põhiklassi. Kvalitatiivse analüüsi käigus suhestatakse ühendid mehhanismide klassifikatsioonipuuga oma füsikokeemiliste omaduste alusel. Selleks kasutatakse kohapeal programmeeritud iseorganiseeruvat närvivõrku, mille tööpõhimõttel peatutakse ka pikemalt. Kvantitatiivse analüüsi käigus teostatakse regressioonanalüüs multilineaarsete mudelite saamiseks kõigile identifitseeritud ja tüpiseeritud toksilisuse mehhanismidele. Mehhanismide klassifikatsioonipuu abil ennustatakse ette ja interpreteeritakse mudelite koosseisu kuuluvad parameetrid.

2 KIRJANDUSE ÜLEVAADE

Toksilisust uuriv eksperiment ei anna üldjuhul otsest informatsiooni toksilisuse mehhanismi kohta. Hulkrakseid organisme uuritakse individipõhiselt. BA mõõtmisprotokollis märgitakse MOA iseloomustavad käitumuslikud ja füsioloogilised sümptomid. Heaks näiteks on McKim jkt. [4] töö, kus kalade respiratoorse ja kardiovaskulaarse aktiivsuse alusel defineeriti mitu FATS: (a) narkoos, (b) hingamisahela regulatsiooni häiritus, (c) lõpusmembraanide ärritus, (d) kesknärvisüsteemi tegevuse häiritus.

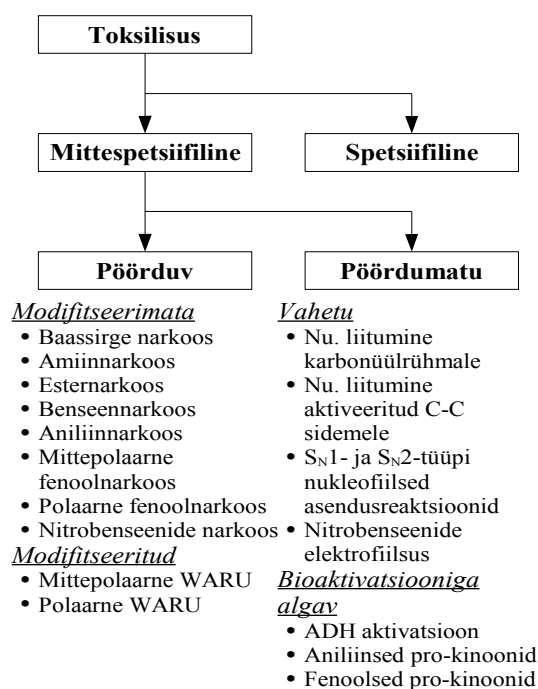
Ainurakseid organisme uuritakse populatsioonipõhiselt. *T. pyriformis* korral peetakse üheks parimaks mõõtmiste lõpp-punktiks (*endpoint*) populatsiooni kasvu aeglustumist saasteaine subletaalse kontsentratsiooniga keskkonnas [5]. Mõõtesuuruseks on sellisel juhul mõni populatsiooni suurusest sõltuv suurus (toitainete tarbimine, jääkainete väljutamine, lahuse optilised omadused), mitte populatsiooni enda koosseis (elusate ja surnud rakkude suhe). Ühe saasteaine näiteks viie erineva kontsentratsiooni juures saadud tulemused teisendatakse (enamasti interpoleerides) ühe kokkuleppelise efektiivkontsentratsiooni peale. Tänu rakuprotsesside universaalsusele peaksid MOA olema organismide vahel vabalt ülekantavad. Kalade ja *T. pyriformis* ühilduvust on põhjalikult kajastatud [6-9].

MOA võib eneses sisaldada ühte või mitut mehhanismi. Näiteks lõpusmembraanide ärritust tuleneb saasteaine reageerimisest sealsete proteiinidega. Et aminohappejääkides sisalduvad sulfhüdrüül-, amino- ja hüdroksüülrühmad on tugevad nukleofiilid, võib saasteaine olla nii nõrgem kui tugevam elektrofiil ja reageerida nii asendus- kui liitumisreaktsioon mehhanismi järgi. Iga selliselt identifitseeritud sündmuse jada kuidas hulk teatud ühisjooni omav saasteaineid interakteerub kindla bioloogilise substraadiga nimetatakse toksilisuse mehhanismiks, ehk lihtsalt mehhanismiks.

2.1 Klassifikatsioonipuu

Kasutatud mehhanismide klassifikatsioon on hierarhiline (Joonis 1). Selle kõige kõrgemal tasemel jagatakse mehhanismid mittespetsiifilisteks ja spetsiifilisteks. Spetsiifilisust võib defineerida kui olukorda kus saasteaine näitab üles soodumust avaldada toimet läbi ühe mehhanismi paljude võimalike mehhanismide hulgast. Spetsiifiliste mehhanismidega tuleb arvestada eeskätt kõrgemate organismide puhul, sest nende funktsioneerimine sõltub suuremast arvust biokeemilistest protsessidest. Heaks näiteks on pestitsiididena kasutatavad fosfororgaanilised ühendid, mis toimivad alküülivate

agentidna. Nende MOA (kesk-)närvisüsteemiga varustatud organismide korral on halvatus (*central nervous system - CNS - seizure*), sest inhibeeritakse närvisignaali edasikandes osalevat ensüümi atsetüülkoliini esteraasi, ja kõigi ülejäänud organismide korral ärritus [10]. Tänu oma lihtsale ehitusele jäävad algloomade sh *T. pyriformis* jaoks spetsiifilise toksilisuse mehhanismid ebaoluliseks.



Joonis 1. Klassifikatsioonipuu.

Järgmisel tasemel jagatakse mehhanismid pöörduvateks ja pöördumatuteks. Pöörduvust võib defineerida kui sümptomite leevenemist nende ilmutumisele lähedase kiirusega kui organism asetatakse saastatud keskkonnast tagasi puhtasse keskkonda. Pöörduvates mehhanismides on saasteaine retseptoriga seotud füüsikaliste interaktsioonide kaudu nagu näiteks van der Waalsi jõud, vesiniksidemed jne. Pöördumatutes mehhanismides moodustuvad saasteaine ja retseptori vahel juba uued kovalentsed sidemed. Vesikeskkonnas esineva akuutse toksilisuse seisukohast on elektrofiilsus nukleofiilsusest palju kordi olulisem. Elektrofiilsed saasteained on vesikeskkonnas parema püsivusega (puudub oht enneaegselt neutraliseerumiseks protoneerimisreaktsioonidel) ning nende "substraadibaas" on laiem.

Ühendi potentsiaalset reaktsioonivõimet saab hinnata piiorbitaalide energiatega

järgi. Elektroni loovutamine (nukleofiilsus) toimub kõrgeimalt hõivatud molekuloorbitaalilt, elektroni vastuvõtmine (elektrofiilsus) toimub madalaimal vabal molekuloorbitaalil. Ühendi tegelik reaktsioonivõime sõltub lisaks piirorbitaalide lokaliseeritusest reaktsioonisentri(te)le ja viimaste steerilisest varjestatusest. Sellist tüüpi omaduste jaoks on heade ja universaalsete deskriptorite leidmine väga keeruline.

Pöörduvate mehhanismide all eristatakse modifitseerimata ja modifitseeritud pöörduvaid mehhanisme. Modifitseeritus tähendab saasteaine “katalüüsivat” lisatoimet. Tüüpiline stsenaarium on selline, kus ühend muudab oma viibimisega (arvesse tulevad nii aktiivne kui passiivne transport) biomembraani läbitavaks ka teistele lahuse komponentidele. Üks paremini uuritud mehhanisme on nõrkade hapete interferents mitokondrites toimuva makroenergeetiliste ühendite sünteesi regulatsiooniga. Terada [11] järgi hävitavad WARU agendid mitokondrite sise- ja välissektsioonide vahel loomulikes tingimustes esineva prootonitegradiendi nn. prootonite süstiktranspordi (*proton shuttling*) mehhanismi järgi.

Pöördumatute mehhanismide all eristatakse vahetuid ja bioaktivatsiooniga algavaid mehhanisme. Bioaktivatsioon tähendab keemiliselt mitteaktiivse ühendi muutumist keemiliselt aktiivseks tänu metabolismiradades läbitehtavatele muudatustele. Kui saaduseks on elektrofiilne agent siis nimetatakse neid tihti Lipnick jkt [12] järgi ka proelektrofiilseteks mehhanismideks. Aktiveeruvate struktuuri(fragmenti)de *ad hoc* identifitseerimiseks puuduvad usaldusväärsed meetodid. Kuna nende isoaktiivsed kontsentratsioonid on sarnastest kuid mitteaktiveeruvatest struktuuri(fragmenti)dest kuni mitu suurusjärku madalamad, siis nad tõusevad ilmekate statistiliste hälbijatena esile kvantitatiivse analüüsi käigus [13]. Bioaktivatsiooniga algavate pöördumatute mehhanismide ülekandmisel ühelt organismilt teisele tuleb kindlasti arvesse võtta võimalikke erinevusi nende ensümaatilises koostises.

Tabel 1 summeerib nelja mehhanismitüübi omadused: sihtkoha, interaktsioontüübi ja sõltuvuse passiivsest transpordist.

Tabel 1. Mehhanismide põhitüübid.

<i>Mehhanism</i>	<i>Sihtkoht</i>	<i>Interaktsiooni tüüp</i>	<i>Sõltuvus passiivsest transpordist</i>
Pöörduv ja modifitseerimata	Rakumembraan	Füüsikaline	-
Pööduv ja modifitseeritud	Rakusisemus	Keemiline ja/või füüsikaline	täielik
Pöördumatu	Rakumembraan ja/või rakusisemus	Keemiline	osaline
Pöördumatu, aga bioaktivatsiooniga algav	Rakusisemus	Keemiline	täielik

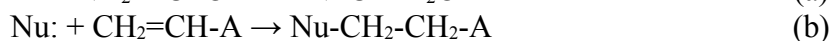
2.2 Alifaatsed ühendid

Modifitseerimata pöörduvatest mehhanismidest on tuntuimad **baassirge narkoos** [14, 15], **amiinnarkoos** [6, 16] ja **esternarkoos** [6, 17]. Võrreldes alkoholidega peetakse amiinide eripäraks nende kõrget nukleofiilsust ning estrite eripäraks nende kerget ensümaatilist transformeeritavust.

Baassirge narkoosi alkoholidest eraldi uuritakse α -alkeenalkohole ja α -alküünalkohole, sest ensüüm ADH oksüdeerib kordse sideme kõrval asuvad hüdroksüülrühmad karbonüülrühmadeks [18]. Primaarsetest ja sekundaarsetest alkoholidest tekivad vastavalt α,β -küllastumata aldehüüdid ja ketoonid, mis on tugevad elektrofiilid. Tertsiaarsed alkoholid ei aktiveeru ja jäävad kirjeldatuks baassirge narkoosi mudeliga, sest neis puudub metaboliseeruv vesinik. Kirjandusest on teada, et algloomad sh *T. pyriformis* jäävad ADH ja esteraaside aktiivsuse poolest kaladele alla.

Pöördumatutest mehhanismidest on olulisemad **nukleofiilide liitumised karbonüülrühmale** [19, 20, 21] ja **aktiveeritud C-C kordsele sidemele** [22] ning **nukleofiilsed asendusreaktsioonid** [23, 24]. Nukleofiilidena figureerivad põhiliselt proteiinide ja nukleiinhapete vabad sulfhüdrüül-, amino- ja hüdroksüülrühmad [25]. Nukleofiilide liitumist karbonüülrühmale näitlikustab amiinide reaktsioon aldehüüdidega (Skeem 1: a), mille saaduseks on imiinid ehk nn. Schiffi alused. Karbonüülühendite reaktsioonivõime kahaneb reas $H_2CO \rightarrow RHC=O \rightarrow R_2C=O$ elektroonsete ja steeriliste efektide koosmõju tõttu [26]. Nukleofiilide liitumist aktiveeritud C-C kordsetele sidemele nimetatakse tihti Michaeli-tüüpi reaktsioonideks analoogia tõttu Michaeli reaktsioonidega. Elektrofiilsustsenter paikneb aktiveeriva rühma suhtes β -süsinikul (Skeem 1: b). Aktiveerivaks rühmaks võivad olla erinevad EWG nagu näiteks nitrorühm, karbonüül-

rühm, esterrühm jt. karboksüülrühma derivaadid. Teoreetiliselt võimaldavad α,β -küllastumata karbonüülühendid $C=C-CO$ lisaks eelnevatele liitumist ka 1,4-asendites [26]. Nukeofiilsetes asendusreaktsioonides vahetatakse substraadi nukleofiilne rühm agendi nukleofiilse rühma vastu (Skeem 1: c). Täpsemalt eristatakse kahte alatüüpi. S_N1 -tüüpi reaktsioonimehhanismis on atakeerivaks osakeseks agendi dissotsieerumisel tekkinud karbokatioon, S_N2 -tüüpi reaktsioonimehhanismis agent ise [26]. Dissotsiatsiooni tõenäosus sõltub rohkem põhiahela võimest stabiliseerida positiivset laengut kui lahkuva rühma headusest. Halogeenderivaatidest võivad S_N1 -tüüpi reaktsioonimehhanismi järgi toimida need, mille struktuuris asub halogeeniatomi suhtes α -asendis kergestipolariseeritav funktsionaalrühm.



a - Nukleofiili liitumine karbonüülrühmale (Schiffi aluse teke); b - Nukleofiili liitumine aktiveeritud kordsele sidemele; c - Nukleofiilne asendusreaktsioon.

Skeem 1. Reaktsioonimehhanismid.

2.3 Aromaatset ühendid

Modifitseerimata pöörduvad mehhanismid sõltuvad kõigi keemiliste klasside jaoks jaotustasakaalust mittepolaarse ja polaarise faasi vahel. Benseenides puudub juhtiv funktsionaalrühm mis eelistaks üht faasi teisele. Kuna **benseennarkoos** on oma omadustelt väga lähedane baassirge narkoosile, siis segaandmekomplektide modelleerimisel on nad tihti kokku võetud üheks nn. mittepolaarseks narkoosiks (*nonpolar narcosis*) [27]. Aniliinide aminorühm mõjutab jaotustasakaalu vesiniksidemete moodustamise kaudu. Erinevalt alifaatsetest amiinidest eksisteerivad aromaatsed amiinid vesilahuses neutraalsete osakeste mitte aniliiniumkatioonidena, sest lämmastiku jagamata elektronipaar on delokaliseeritud tuumale. Fenoolide hüdroksüülrühm mõjutab jaotustasakaalu vesiniksidemete moodustamise ja happelise dissotsiatsiooni kaudu.

Aine dissotsiatsioonimäära erinevate keskkonnatingimuste juures saab hinnata Henderson-Hasselbalchi võrrandist kui on teada tema (esimest järku-) happelise dissotsiatsiooni koefitsiendi pK_a väärtus. Kvalitatiivse pildi seisukohast piisab kui eristada kolme klassi: (a) fenoolid, mille pK_a on vähemalt 9 kuni 10 on füsioloogilistes tingimustes peaaegu täielikult dissotsieerumata, (b) fenoolid, mille pK_a on vahemikus 6 kuni 8 on

osaliselt dissotsieerunud, (c) fenoolid, mille pK_a ei ületa 4 kuni 5 on peaaegu täielikult dissotsieerunud. Sellisel tasemel otsuste tegemiseks võib pK_a vabalt asendada ühe-kahe lihtsamini kättesaadava deskriptoriga. Päril hästi töötab laengujaotuse deskriptoritest koostatud paar, milles katioonseks komponendiks on aktiivseima vesinikuaatomi laeng ja anioonseks komponendiks ülejäänud molekuli laeng. Käesolevas töös seostatakse need kolm klassi vastavalt **mittepolaarse fenoolnarkoosi**, **polaarse fenoolnarkoosi** ja **WARU** mehhanismidega.

Aniliinnarkoosi [28] ja **mittepolaarse fenoolnarkoosi** [29, 30] üheparameetrilised $\text{Log}K_{ow}$ põhised mudelid on omavahel väga sarnased. Kokkuvõetult moodustavad nad mittepolaarsele narkoosile vastanduva nn. polaarse narkoosi (*polar narcosis*) [27, 31]. Teadusharus on mittepolaarse ja polaarse narkoosi ühis- ja eriosade üle aktiivselt diskuteeritud. Kahjuks puudub siiaaani konsensus, et kas mehhanismid on üksteisest tõeliselt erinevad (molekulaarne tase, näiteks rohkemate ja tugevamate vesiniksidemete teke) või on tegemist pelgalt kasutatavate deskriptorite ebaadekvaatsusega [32].

Mittepolaarse fenoolnarkoosi mehhanismi järgi toimivad ühendid moodustavad ühe suurima ja keemiliselt mitmekesisema andmekomplekti. Seniste üheparameetriliste mudelite statistiline kvaliteet paraneb märgatavalt abstraheriva suunitlusega elektroonsete parameetrite lisamisega [33].

Polaarse fenoolnarkoosi jaoks pole jaotuskoeffitsientide põhjal võimalik tuletada üheparameetrilist mudelit, mis oleks eelnevatega võrreldava statistilise kvaliteediga. Klassi tüüpilised esindajad mononitrofenoolid toimivad mittevesikeskkonnas (naha sensibiliseerimise eksperiment) nn. pehme elektrofiilsuse (*soft electrophilicity*) mehhanismi järgi [34]. Teadusharus valitseva seisukoha järgi kehtib sama edukalt ka vesikeskkonnas. Kedagi ei paista seejuures häirivat loogiline vastuolu, et mononitrofenoolid esindavad pöördumatut mehhanismi aga nendest märksa elektronivaegsemad di- ja trinitrofenoolid modifitseeritud pöörduvat mehhanismi. Käesolevas töös seletatakse sirgest hälbumist jaotuskoeffitsientide suutmatusega koondada erineval määral ioniseerunud (so. happeliselt dissotsieerunud) ühendeid. On ju hästi teada, et laetud osakeste nn. bioainesse ülevõtmine (*biouptake*) on võrreldes neutraalsete osakestega tugevasti pärssitud.

Mahukate nitrobenseenide andmekomplektide jaoks saadakse paremaid tulemusi mononitrobenseene ja dinitrobenseene eraldi modelleerides [35, 36, 37]. Kuna mono-

nitrobenseenide korral on isoefektiivsed kontsentratsioonid tunduvalt suuremad kui dinitrobenseenide korral, siis võib rääkida vastavalt **nitrobenseenide narkoosi** ja **nitrobenseenide elektrofiilsuse** mehhanismidest. Suuremat selgust aromaatsete ühendite elektrofiilse reaktsioonivõime kohta vesikeskkonnas oodatakse teise andmekomplekti analüüsist, mis sisaldab rohkesti mono- ja dinitrofenoolide ning mono- ja dinitrobenseene.

Bioaktivatsiooniga algavatest pöördumatutest mehhanismidest pakub huvi amino-fenoolide ja benseendiolide oksüdatsioon semikinoonideks kuni kinoonideks [38]. Kuna lämmastiku- ja hapnikuaatomite keemia on erinev siis defineeritakse ettevaatavalt vastavalt **aniliinsete-** ja **fenoolsete pro-kinoonide** mehhanismid. Aktivatsioon kulgeb mitmes etapis, mistõttu on väga raske hinnata milline on vaheproduktide ja lõpp-produkti panus kogutoimesse. Ennustatavaks lõpp-produktiks olevat (benso)kinooni peetakse Michaeli-tüüpi aktseptoriks [39].

3 MATERJALID JA MEETODID

3.1 Keemilised ja bioloogilised andmed

Mõlemad andmekomplektid pärinevad Schultzi [5] hallatavast ja arendatavast andmebaasist TETRATOX. Selle andmebaasi väärtusteks on suur maht (erialases kirjanduses küll publitseeritud ainult osalisena) ning andmete esinduslikkus ja ühtlane kõrge kvaliteet.

Eksperimentaalne suurus $p(IGC_{50})$ on negatiivne logaritm *T. pyriformis* populatsiooni kasvu 50% võrra aeglustavast kontsentratsioonist (*population growth impairment concentration*). Kõik kontsentratsioonid on teisendatud molaarseteks. Eksperiment on teostatud 40-tunni jooksul staatilises režiimis, keskkonna pH on algushetkel puhverdatud 7,35 juures.

Struktuuride töötlemine teostati kahes etapis. Alguses kaardistati kogu konformatsiooniruum kasutades MacroModel tarkvaras [40] realiseeritud Monte Carlo mitme-miinumumi (*Monte Carlo multiple minimum*) meetodit [41] ja MMFF94s jõuvälja [42]. Otsitavate konformeeride maksimaalne arv määrati raskete aatomite vaheliste üksiksidemete arvu korrutamisel tuhandega. Geomeetrilise ja/või optilise isomeeria korral rakendati vajalikke kitsendusi, et tagada stereokeskmete konfiguratsiooni säilimine. Globaalsele miinumumile ja selle lähematele (kuni 5 kJ) lokaalsetele miinumumidele vastavad konformeerid talletatakse. Lõpus viidi läbi kvantkeemilised arvutused kasutades MOPAC tarkvaras [43] realiseeritud poolempiirilist AM1 Hamiltoniaani [44]. Iga talletatud konformeer tegi läbi täiendava geomeetria optimeerimise (GNORM=0.00 NOMM) enne omaduste arvutamist (1SCF VECTORS BONDS PI POLAR ENPART PRECISE NOMM). Kvalitatiivses analüüsis vajatakse struktuuri esitamiseks kõiki tema konformeeere, kvantitatiivses analüüsis aga ainult globaalsele miinumumile vastavat konformeeeri.

Kasutades CODESSA tarkvara [45] arvutati suurem hulk koostislikke (38 nimetust), topoloogilisi (38 nimetust), geomeetrilisi (12 nimetust), elektrostaatilisi (83 nimetust) ja kvantkeemilisi (443) deskriptoreid. Deskriptorite arvutuseeskirjad ja tähendus on pikemalt lahti seletatud mujal [46]. Kasutades CLogP tarkvara [47] arvutati hüdrofoobsust kirjeldav parameeter logaritm jaotuskoefitsiendist 1-oktanooli ja vee vahel $LogK_{ow}$. Viimasel ajal on päevakorda tõusnud teema, et biomembraanide realistlikumaks modelleerimiseks tuleks 1-oktanool asendada mõne lihtsama lipiidiga, näiteks L-dimüris-

toüülfosfatidüülkoliiniga (*L-dimyristoyl phosphatidylcholine* - DMPC). Kuna jaotuskoefitsiendi $\text{Log}K_{DMPC}$ arvutamiseks puudub usaldusväärne algoritm ja andmebaasid on lünklikud siis jäi see võimalus hetkel lähemalt uurimata.

3.2 Kvalitatiivne analüüs

Paljude kirjeldatud mehhanismide kvalitatiivseks määramiseks piisab ühe või mitme iseloomuliku struktuurse fragmendi ehk nn. toksikofoori (*toxicophore*) äratundmisest [48, 49]. Toksikofooride tehnika on oluliselt edukam alifaatsete ühendite korral kui aromaatssete ühendite korral, sest küllastatud süsivesinikahel on võrreldes konjugeeritud kordsete sidemetega süsivesinikahelaga väheefektiivne elektronitiheduse ümberjagaja. Alifaatsete ühendite korral piisab kui üles otsida ainult kõige tugevam toksikofoor. Sellise tööga saavad päris edukalt hakkama mitmesugused ekspertsüsteemid mis üritavad leida kokkulangevusi uuritava ühendi 2D struktuurivalemi ja oma sisemise fragmendiandmebaasi vahel. Otsing võib toimuda tekstiotsinguna, sest fragmendi lähem ja kaugem keemiline naabruskond arvesse ei tule. Erinevalt alifaatsetest ühenditest tuleb aromaatssete ühendite korral üles otsida ja järjestada kõik toksikofoorid. Järjestamist võimaldavate reeglite defineerimine on väga keeruline kui mitte ütelda võimatu, sest arvestada tuleb rikkaliku struktuurse ja positsioonilise isomeeriaga. Schultz jkt. [50] on defineerinud viiele fenoolide MOA-le vastavad toksikofoorid ja rakendanud need edukalt ligi kahesajaliikmelise andmekomplekti peal.

Toksikofooride tehnika puuduseks tuleb pidada väga tagasihoidlikku ennustus- ja üldistusvõimet. Piisab ühe uue funktsionaalrühma lisandumisest ja kõik senikogutud teadmised muutuvad kaheldavateks. Puuduvad objektiivsed kriteeriumid toksikofooride järjestamiseks ja kombineerimiseks. Schultz jkt. [50] on astunud esimesed sammud uurimaks toksikofooride tõlkimist füsikokeemiliste omaduste keelde. Neid saatnud edu on sellele rajale meelitanud hulgaliselt teisigi uurijaid. Täna on sama või paarikümne punkti võrra suurem fenoolide andmekomplekt läbi uuritud mitmesuguste statistiliste klassifikatsioonimeetoditega, millest mõningatena võib nimetada diskriminantanalüüsi [51-54], vähimosaruutude analüüsi [55], otsustuspuud [56], närvivõrke [57]. Statistilised klassifikatsioonimeetodid on struktuursete klassifikatsioonimeetodite puudustest vabad. Pikemaajalisel kokkupuutel selgub aga fakt, et nad ei suuda probleeme lahendada loovalt ega avardada senist teadmisteruumi, sest neis ei eksisteeri põhjus-tagajärg seoseid. Parimate statistiliste klassifikatsioonimeetodite täpsus ulatub 90 kuni 95 protsendini.

Valestiklassifitseeritud punktide uurimine jätab paraku liigagi tihti mulje, et süüdlaseks pole mitte ühend (ja tema füsikokeemilised omadused) või meetod, vaid mehhanism, mida üritatakse reprodutseerida.

3.2.1 SOM - iseorganiseeruv närvivõrk

Juhendatud õppimise (*supervised learning*) meetodid (siia alla kuuluvad kõik eelnevas lõigus nimetatud statistilised klassifikatsioonimeetodid) põhinevad eeldusel, et juhendaja valduses on ette olemas täielik komplekt õigeid või õigeks peetavaid vastuseid. Agent parandab oma tulemusi juhendajalt laekuva tagasiside põhjal, et kui hästi rahuldab ennustatud tulemus oodatud tulemusele seatud kriteeriume. Juhendamata õppimise (*unsupervised learning*) meetodid sellist eeldust ei oma. Agent parandab oma tulemusi iseseisva "arutluse" najal. Juhendatud õppimine sobib oma iseloomult paremini kvantitatiivseks analüüsiks, juhendamata õppimine aga kvalitatiivseks analüüsiks.

Klassifikatsiooniprobleemi lahendamiseks sobivad hästi iseorganiseeruvad närvivõrgud, mida nende väljatöötaja ja populariseerija järgi tuntakse paremini Kohoneni närvivõrkudena [58]. SOM teostab andmepunktide üleviimist n -mõõtmelisest sisendruumist m -mõõtmelisse väljundruumi säilitades nende omaduste topoloogia. Enamikus rakendustes on sooviks mõõtmelisust vähendada (so lihtsustada), seega $n > m$. Omaduste topoloogia kirjeldab andmepunktide jaotumist ruumis nende omaduste alusel. Andmepunktid mis on üksteisele lähedased sisendruumis saavad niiviisi paiknema ka väljundruumis. SOM peetakse üheks paremaks mudeliks selle kohta kuidas töötavad bioloogilised närvivõrgud (sh inimaju).

Sisendruumi disain sõltub lahendatavast ülesandest. Käesolevas töös saavad omadustvektori (*feature vector*) elementideks mehhanismide klassifikatsioonipuus (Joonis 1) otsuste tegemiseks kasutatud deskriptorid. Kirjanduses võib leida omadustvektoreid mille pikkus ulatub mõnest mitmekümne ja isegi mitmesajani. Väljundruumi disain sõltub tulemuste analüüsimeetodist. Klassipiiride visualiseerimise ja tõlgendamise poolest on kõige mugavamad ühe-, kahe- ja kolmemõõtmelised ruumid kus neuronid on asetatud ruudukujuliselt. Kui tööd tehakse automatiseeritult mitte käsitsi, tulevad kõne alla ka kõrgemamõõtmelised ruumid kus neuronid on asetatud heksagonaalselt jne.

3.2.2 Algoritm

Käesolevas töös programmeeritud SOM algoritm (modifitseeritud Zupan jkt. [59] järgi) on esitatav alljärgneva pseudokoodina (Algoritm 1).

- a: Initseeri sisendruum.
- b: Initseeri väljundruum.
- c: Korda treeningvea koondumiseni:
- d: Korda iga ühendi jaoks andmekomplektis:
- e: Vali signaaliks juhusliku konformeerimise omadustevektor.
- f: Leia võitjaneuron.
- g: Korriga võitjaneuroni ja tema naaberneuronite kaaluvektoreid.
- h: Visualiseeri väljundruum.

Algoritm 1. SOM.

Sisendruumi initseerimisel (Algoritm 1: a) itereeritakse üle kõigi ühendite üle kõigi konformeeride ja koostatakse vastavad omadustevektorid. Omaduste jaoks registreeritakse miinimum- ja maksimumväärtus ning teostatakse normaliseerimine. Omaduse väärtuse suund, ehk see mida tuleks pidada miinimumiks ja mida maksimumiks, on kokkuleppeline. Mõningaid omadusi tasub kontrastuse parandamiseks eelnevalt transformeerida. Näiteks elektrostaatilisid deskriptoreid (vabad osalaengud) on füüsiliselt põhjendatud ruutu tõsta. Treenimisel kaetud omaduste väärtustevahemik peaks olema võimalikult lai ja ühtlane.

Väljundruumi initseerimisel (Algoritm 1: b) pannakse paika kahemõõtmeline massiiv neuroneid. Neuronite kaaluvektorite (*weight vectors*) pikkus on võrdne omadustevektorite pikkusega ja nende elementide algväärtused seatakse juhuarvudeks. Massiivi optimaalse küljepikkuse määramiseks puudub formaalne eeskiri. Ka selleks, et õppida paremini tundma andmekomplekti iseloomu, tasub teha mitu prooviklassifitseerimist erinevate küljepikkuste juures. Kokku testiti seitset konfiguratsiooni: 10x10, 15x15, 20x20, 25x25, 30x30, 35x35 ja 40x40. Parimad tulemused saadi keskmiste (20 kuni 30) suuruste juures. Väikesed närvivõrgud pakivad klassid liiga tihedasti kokku, mis põhjustab arvestatava detailikao. Esineb oht, et vähemaliikmelised klassid neelatakse alla või surutakse maha suuremaliikmeliste klasside poolt vastavalt sellele kas nad erinevad üksteisest veidi vähem või rohkem. Suured närvivõrgud kalduvad teise äärmusesse. Nende puhul kipub klasside sisse ja vahele jääma liiga palju tühja ruumi, mis põhjustab piirjoonte hägustumist.

Närvivõrgu treenimine (Algoritm 1: c - g) toimub iteratiivselt kuni treeningvea koondumiseni. Iga treeningtsükli (*epoch*) jooksul esitatakse iga ühendit ühe juhuslikult valitud konformeeriga (Algoritm 1: e). Uudne mitmekonformeerimise lähenemine peaks

võrreldes tavaline ühekonformeeri-lähenedamisega võimaldama närvivõrgule üle kanda arvestataval hulgal peidetud informatsiooni mis seostub konformatsioonimuutuste tõenäosusega ja 3D deskriptorite vastuvõtlikkusega sellele. Lisaks sooviti kontrollida, et kas võib esineda olukord, et ühe ja sama ühendi erinevad konformeerid toimivad erineva mehhanismi järgi. Viiestsajast ühendist koosneva andmekomplekti treeningvea koondumiseks kulub ligikaudu sada treeningtsükli. Arvutuskulu skaleerub ruutfunktsioonina küljepikkusest ja jääb keskmise lauaarvuti võimsuse juures mõne sekundi kuni mõne minuti vahemikku.

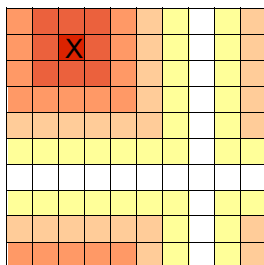
SOM põhineb võistlerval õppimisel (*competitive learning*), mis tähendab seda, et sisestatud signaali töötlemiseks valitakse ainult üks neuron paljudest. Võitja väljaselgitamiseks (Algoritm 1: f) võrreldakse signaali s moodustavat omadustevektorit X_s ($x_{s1}, x_{s2}, \dots, x_{sm}$) kõigi neuronite j kaaluvektoritega W_j ($w_{j1}, w_{j2}, \dots, w_{jm}$). Võitjaks osutub neuron mille puhul on X_s ja W_j vaheline eukleidiline kaugus kõige väiksem, ehk nad on kõige sarnasemad (Võrrand 1). Kui ühendi jaoks on omadus i määramata, näiteks kui tegemist on aatomitüübipõhise deskriptoriga ja struktuuris puudub vastav aatomitüüp, siis tähistatakse omadustevektoris vastava elemendi väärtus kokkuleppelise konstandiga NaN (*not-a-number*).

$$winner \leftarrow \min \left\{ \sum_{i=1}^m (x_{si} - w_{ji})^2 \right\} \quad (1)$$

Õppimise sisuks (Algoritm 1: g) on korrigeerida võitjaneuroni ja tema naaberneuronite kaaluvektoreid nii, et signaali s korduval sisestamisel jääks võitja täpselt samaks aga viga oleks vähenenud. Naaberneuronite kaasahaaramise eesmärgiks on piirkonda laiemalt välja kujundada. Esimene närvivõrku sisestatud signaal võib võitja leida suvalisest kohast (mäletatavasti on kaaluvektorite elemendid algväärtustatud juhuarvudega). Korrektsiooni tulemusena muutub võitjaneuron ja tema naaberneuronid üksteisele sarnasemaks. Teine närvivõrku sisestatud signaal ei oma enam sellist otsustusvabadust. Kui ta on esimese signaaliga vähem või rohkem sarnane siis ta leiab võitja juba korrastumas piirkonnast, kui sellest täiesti erinev, siis kusagilt mujalt. Mida rohkem signaale närvivõrku sisestatud on seda korrapärasemaks see muutub. Hea korrektsioonifunktsioon võtab seda arvesse ja kahandab oma mõju tugevust ja ulatust.

Tõenäoliselt vajab selgitamist neuroni naabruskonna mõiste. SOM nõuab, et kõigi neuronite naabruskond oleks samaväärne nii naaberneuronite arvu kui paigutuse osas.

Pind ei tohi loogilisel tasemel diskrimineerida ääreneuronite ja mitte-ääreneuronite vahel. Kahemõõtmelisest massiivist saab soovitud omadustega pinna (toroidi) kujundada kui ühenda omavahel vastavalt ülemine ja alumine serv ning vasakpoolne ja parempoolne serv (Joonis 2).



Joonis 2. Neuron $\{2, 1\}$ ja tema 1. - 4. järku naabruskonnad.

Naabruskonnad on esimest-, teist- jne järku. Ruudukujulise asetuse korral on i ndat järku naabruskonnas asuvate neuronite arv n arvutatav Võrrandiga 2 (Joonis 2).

$$n = 4 + 4 \cdot (2 \cdot i - 1) = 8 \cdot i; i \geq 1 \quad (2)$$

Hoolimata unikaalsest topoloogiast on kõik valminud närvivõrgud ühesuguse klasside suhestatusega (vajaduse korral on tulemusi võimalik reprodutseerida kui kasutada etteantud algväärtusega (pseudo-)juhuarvude generaatorit).

Väljundruumi visualiseerimisel (Algoritm 1: h) genereeritakse graafikud mille alusel toimub klasside ja klassipiiride identifitseerimine. Kõige informatiivsemaks tuleb pidada nn. kaalukaarte (*weight distribution maps*), mis on väljundruumi "ristlõikeid" sisendruumi mõõtmete (so. omaduste) järgi. Alusandmeteks olevad kaaluvektorite elementide väärtused tõlgitakse must-valgesse värviskaalasse, kus heledamad toonid tähistavad suuremaid väärtusi ja tumedamad toonid väiksemaid väärtusi.

3.2.3 Klassid ja klassipiirid

Klassi identifitseerimiseks otsitakse ühe või mitme kaalukaardi pealt üles teda iseloomustav(ad) omadus(ed). Kõige tulemuslikumaks osutus järjekord kus keerulisemad (defineeritud silmatorkavamate ja rohkemate omaduse järgi) mehhanismid leitakse enne lihtsamaid mehhanisme. Palju lisateavet pakkus ka neuronite sisunimekirjade vaatamine kui valminud närvivõrk täideti kõigi andmekomplekti kuulunud ühendite kõigi konformeridega.

Klassipiiride identifitseerimiseks otsitakse üles tähendusrikkad tühjad neuronid

või, mis veel parem, nendest moodustunud ahelad. Tühi neuron on neuron mis ei osutunud valminud närvivõrgu täitmisel võitjaks ühegi sisestatud signaali suhtes. Ta on seda tähendusrikkam, mida suurem on tõenäosus, et ta jääb tühjaks ka treeningandme-komplekti asendamisel test- ja tööandmekomplektidega. Heaks indikaatoriks on NSI, mis arvutatakse uuritava neuroni W_x kaaluvektori keskmise eukleidilise kaugusena tema 1. järku naaberneuronite kaaluvektoritega W_j (Võrrand 3).

$$NSI = \frac{1}{8} \cdot \sum_{j=1}^8 \sum_{i=0}^m (w_{xi} - w_{ji})^2 \quad (3)$$

Toodud arvutuseeskiri rõhutab enam suurt muudatust ühe omaduse lõikes kui väikeseid muudatusi mitme omaduse lõikes. Nn. sarnasuskaardi (*similarity map*) koostamiseks tõlgitakse NSI väärtused must-valgesse skaalasse, kus heledamad toonid tähistavad suuremat sarnasust ja tumedamad toonid väiksemat sarnasust.

Kaks klassi on üksteisest seda erinevamad mida pidevam ja tumedamate tühjade neuronitega nendevaheline piirjoon on märgitud. Piirjoone silmapaistvus kasvab liikudes mööda paare: (a) modifitseerimata pöörduvad ja modifitseeritud pöörduvad mehhanismid, (b) modifitseerimata pöörduvad ja bioaktivatsiooniga algavad pöördumatud mehhanismid, (c) modifitseerimata pöörduvad ja pöördumatud mehhanismid. Esimeses paaris on üleminek pidev, mõnikord lausa ilma tühjade neuroniteta. Kunstlikuks jääva piirjoone lähikonnas asuvate ühendite kohta oleks korrektsem väita, et nad sisaldavad ühepalju ($x\%$) üht ja teistpalju ($100\% - x\%$) teist mehhanismi kui, et nad on üks või teine. Modelleerimisel võib saada ühte viisi häid (sageli hoopis ühte viisi halbu) tulemusi mõlema mehhanismi mudelitega. Teises paaris on üleminek katkendlikum. Kolmandas paaris on üleminek suisa järsk.

Kahe silma vahele ei tohi jätta märkust ahelate moodustamise kohta. Suurema klassi hulka kuuluv erandlik ühend võib nautida vabadust endale haarata spetsiaalset neuron. Kui eristav omadus on diskreetse iseloomuga (näiteks funktsionaalrühm, mida ei saa asendada teiste funktsionaalrühmadega), siis moodustub selle neuroni ümber hästi äratuntav tühjade neuronite võru. Kvalitatiivse analüüsi tulemusi on lihtne kommu-keerida tabuleeritult.

3.3 Kvantitatiivne analüüs

Saasteaine BA sõltub tema polaarsest väliskeskkonnast vähe(m)polaarsesse sisekeskkonda ülevõtmise tõenäosusest p_I ja mehhanismile vastava retseptoriga

interakteerumise tõenäosusest p_2 . Eelnevast järeldub, et erinevate saasteainete isoefektiivsed kontsentratsioonid C on erinevad (Võrrand 4):

$$A = C \cdot p_1 \cdot p_2 \cdot c = \text{const} \quad (4)$$

Võrrandi 4 logaritmine annab Hanschi-tüüpi võrrandi (Võrrand 5):

$$\log\left(\frac{1}{C}\right) = \log(p_1) + \log(p_2) + c \quad (5)$$

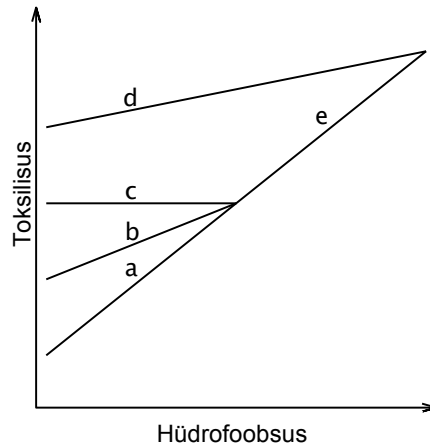
Hanschi meetod [60, 61] on üks populaarsemaid QSAR mudelite koostamise meetodeid. Tema järgi on uuritava ühendi BA seotud baasühendi BAga läbi asendajakoefitsientide hüdrofoobsete efektide π , streeriliste efektide E_s ja elektroonsete efektide σ kohta (Võrrand 6).

$$\log\left(\frac{1}{C}\right) = c_1 \cdot \pi + c_2 \cdot E_s + c_3 \cdot \sigma + c_0 \quad (6)$$

Mingit andmekomplekti (keemiline klass) iseloomustavate kordajate c_1 , c_2 , c_3 ja vabaliikme c_0 väärtused leitakse multilineaarsel regressioonil.

3.3.1 Toimeprofili kontseptsioon

Tabelis 1 nimetatud mehhanismid jagunevad kaheks vastavalt sellele kas nende toime on sõltumatu või sõltuv passiivsest transpordist. Esimesed, milleks on modifitseerimata pöörduvad ja vahetud pöördumatud mehhanismid, näitavad üle kogu hüdrofoobuse skaala enam-vähem ühtset sirget (Joonis 3: a+e ja d), kuna teised, milleks on modifitseeritud pöörduvad ja bioaktivatsiooniga algavad pöördumatud mehhanismid, näitavad murtud sirget (Joonis 3: b+e ja c+e).



Toksikodünaamiline kontroll: a - modifitseerimata pöörduv mehhanism, b - modifitseeritud pöörduv mehhanism (madal p_2), c - bioaktivatsiooniga algav pöördumatu mehhanism (kõrge p_2), e - pöördumatu mehhanism.

Toksikokineetiline kontroll: d - modifitseerimata pöörduv mehhanism, millesse on koondunud ka modifitseeritud pöörduv mehhanism ja bioaktivatsiooniga algav pöördumatu mehhanism.

Joonis 3. Toimeprofiilid.

Nähtuse seletamiseks formuleeriti nn. toimeprofiili (*response profile*) kontseptsioon. Murdepunkti asukoht Joonisel 3 sõltub mõnevõrra organismist ja eksperimendi läbiviimise tingimustest kuid asub üldjoontes $\text{Log}K_{ow}$ vahemikus 2 kuni 3. Murdepunktist allapoole jäävat osa nimetatakse toksikodünaamilise kontrolli piirkonnaks. Madala hüdrofoobsusega klassiliikmed liiguvad raku sisekeskkonnas suhteliselt väiksema takistusega, neile ei ole märkimisväärseks probleemiks läbida järgemööda ka mitut biomembraani. Koguaktiivsus on summa modifitseeritud pöörduva või bioaktivatsiooniga algava pöördumatu mehhanismi aktiivsusest A_{dep} ja alloleva modifitseerimata pöörduva mehhanismi aktiivsusest A_{indep} (Võrrand 7).

$$A = A_{dep} + A_{indep} \quad (7)$$

Murdepunktist ülespoole jäävat osa nimetatakse toksikokineetilise kontrolli piirkonnaks. Kõrge hüdrofoobsusega klassiliikmed peatatakse efektiivselt juba esimeses biomembraanis milleks on rakumembraan. Juhul kui nad ei suuda seda atakeerida mis on peaaegu alati ka tõsi, jääb koguaktiivsus määratuks vaid alloleva modifitseerimata pöörduva mehhanismi aktiivsusega A_{indep} (Võrrand 8).

$$A \approx A_{indep} \quad (8)$$

3.3.2 Baassirge toksilisuse kontseptsioon ja liigtoksilisus

Toimeprofiilide kontseptsioonile ei leidu terviklikku analoogi. Teadusharu alustalaks on Könemanni [14] esitatud nn. baassirge toksilisuse kontseptsioon (*baseline*

toxicity concept), mille järgi on iga saasteaine vähemalt sedavõrd aktiivne kui seda ennustab üks kokkuleppeline narkoosimudel. Lipnick jkt [62] lisasid sellele mõned aastat hiljem nn. liigtoksilisuse (*excess toxicity* - ET) mõiste, et ühtlustada tõlgendus proelektrofiilsete ja elektrofiilsete saasteainete kõrgenenud aktiivsuse kohta. Liigtoksilisus arvutatakse saasteaine tegeliku aktiivsuse A_{obs} suhtena ennustatud aktiivsusse A_{calc} (Võrrand 9), mis mõnikord logaritmitakse ülevaatlikkuse parandamise eesmärgil.

$$ET = \frac{A_{calc}}{A_{obs}} \geq 1 \quad (9)$$

Toimeprofiilide kontseptsiooni eellasteks saab pidada kirjeldusi tasemel "liigtoksilisus väheneb homoloogilises reas kuni muutub tühiseks" [63, 64].

Käesolevas töös kasutatakse liigtoksilisuse arvutamiseks mitut baassirget. Tegemist on põhimõttelise otsusega. Iga suurema keemilise klassi jaoks koostatakse baassirge võttes aluseks vastava kõige tüüpilisema mitte-modifitseeritud pöörduva mehhanismi. Sellisest lähenemisest on aromaatsete ühendite juures rohkem kasu kui alifaatsete ühendite juures. Määravaks saab jällegi funktsionaalrühmade parem omavaheline sidusus. Kui tuumaga on seotud hüdroksüülrühm siis käsitletakse ühendit fenoolina, kui aminorühm siis aniliinina, muidu tavalise benseenina. Sekundaarsed funktsionaalrühmad võimendavad primaarse funktsionaalrühma omadusi. Näiteks, kui aktiveerivate EDG lisamine fenoolile vähendab happelise dissotsiatsiooni tõenäosust ja edendab vesiniksideme aktseptorsuutlikkust doonorsuutlikkuse arvel, siis deaktiveerivate EDG ja EWG lisamine annab vastavalt vähem või rohkem vastupidise efekti. Fenooli-mehhanismide omavaheline võrdlemine üle fenoolse baassirge, aniliinimehhanismide üle aniliinse baassirge ja muude aromaatsete mehhanismide üle benseense baassirge lubab keskenduda senisest sisulisematele küsimustele.

3.3.3 Regressioonianalüüs

QSAR mudelite koostamiseks kasutatakse CODESSA tarkvaras [45] realiseeritud BMLR tehnikat [65]. BMLR tehnika algab tähtsusetute ($R^2 < 0,1$) ja puuduva väärtusega deskriptorite elimineerimisega. Seejärel leitakse parimad kahe-, kolme- jne. parameetrilised regressioonid lähtudes deskriptorite statistilisest olulisusest ja mittekollineaarsuse kriteeriumist ($R^2 < 0,6$). Arvutustes kasutatakse automaatselt normaliseeritud ja tsentreeritud deskriptorite skaalat. Lõplik mudel on omaduse parim kirjeldus etteantud parameetrite arvu juures antud deskriptorite kogumi juures.

Mingi deskriptori informatsioonitiheduse ning olulisuse ja põhjendatuse kohta mudelis saab hea ülevaate kui püüda teda asendada teiste deskriptoritega. Heaks näiteks on nendesamade kolme aromaatsete ühendite baassirge tuletamine deskriptoritevalimitega mis sisaldavad: (a) ainult \mathbf{LogK}_{ow} , (b) \mathbf{LogK}_{ow} ja kõiki teoreetilisi molekulaarseid deskriptoreid, (c) ainult kõiki teoreetilisi molekulaarseid deskriptoreid. Teise mudelisse lisandunud deskriptorid, kui üldse, vihjavad milliste hüdrofoobsete interaktsioonide kirjeldamise osas on \mathbf{LogK}_{ow} üksinda puudulik. Kolmas mudel näitab absoluutset keerukust. Vastavalt LSER võiks eeldada, et tema koostisesse hakkab kuuluma vähemalt üks geomeetiline või topoloogiline deskriptor (õõnsuse komponent) ning paar-kolm elektroonset deskriptorit (polaarsuse/polariseeritavuse ning vesiniksidemete komponent).

4 TULEMUSED JA ARUTELU

Alifaatsete ja aromaatsete ühendite andmekomplekt ning aromaatsete ühendite laiendatud andmekomplekt on esitatud käesoleva töö Lisadena 1 ja 2.

Iseorganiseeruva närvivõrgu osas on Joonistel 4 ja 5 reprodutseeritud lõplikud mehhanismide topoloogiad. Jälgimaks kaasnevat arutlust klassipiiride ja klassidele iseloomulike deskriptoriväärtuste üle soovitatakse lugejal konsulteerida vastavate Lisades toodud sarnasuskaartide ja kaalukaartidega. Regressioonianalüüsi osas saadud mudelid nagu ka viidatavad graafikud keemiliste klasside toksilisuse ja hüdrofoobsuse kohta asuvad samuti Lisades.

4.1 Alifaatsete ja aromaatsete ühendite andmekomplekt

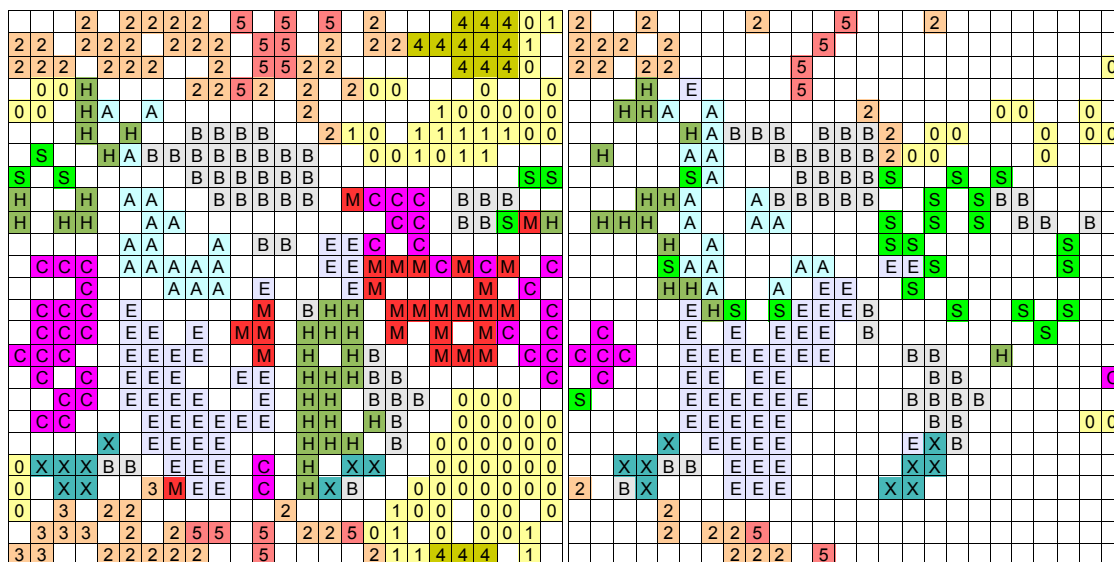
Treeningandmekomplekt ja testandmekomplekt, mis koosnesid vastavalt 275 ja 179 alifaatses ühendist ning 219 ja 40 aromaatses ühendist, pärinesid erinevatest allikatest ja olid erineva iseloomuga. Treenimiseks kasutatud alifaatsed ühendid olid monofunktsionaalsed: alkoholid, amiinid, karbonüülühendid (ketoonid, aldehüüdid), karboksüülhapete derivaadid (estrid, laktoonid, nitrilid), halogeenderivaadid. Testimiseks kasutatud alifaatsed ühendid olid polüfunktsionaalsed. Tegemist oli treeningandmekomplektist tuttavate funktsionaalrühmade kombineerimisega: dioolid, aminoalkoholid, halogeenitud alkoholid, halogeenitud karbonüülühendid, halogeenitud karboksüülhapete derivaadid. Kõigi aromaatsete ühendite ühistunnuseks oli fenoolne hüdroksüülrühm. Erinevalt treeningandmekomplektist võis testandmekomplektist leida mitmetuumalisi ja heteroatomide sisaldavaid struktuure.

4.1.1 Klassifitseerimine iseorganiseeruva närvivõrguga

Alifaatsed ja aromaatsed ühendid klassifitseeriti koos. Pöörduvate ja pöörduvatute mehhanismide eristamiseks valiti maksimaalne Fukui elektrofiilsusindeks süsinikule E_C ja maksimaalne Fukui nukleofiilsusindeks raskele aatomile N_A . Fukui indeksid sõltuvad piirorbitaalide energiatest ja lokaliseeritusest. Kaalukaartidelt võis leida, et viimane tingimus viis teravalt lahku alifaatsed ühendid ja aromaatsed ühendid. Reaktsioonimehhanismi ennustamiseks vajatakse rohkem informatsiooni E_C ja N_A tsentrite paiknemise kohta struktuuris. Kuna praeguse põlvkonna tarkvara seda põhimõtteliselt ei võimalda, tuli suur hulk otsuseid langetada käsitsi. Kui saasteainet iseloomustab kõrge E_C ja madal kuni keskmine N_A kontrollitakse elektrofiilsete liitumisreaktsioonide võimalikkust. Kui saasteainet iseloomustab keskmine kuni kõrge E_C ja kõrge N_A

kontrollitakse S_{N1} - või S_{N2} -tüüpi nukleofiilsete asendusreaktsioonide võimalikkust. Ilma E_C toetuseta kõrgeid N_A väärtuseid ignoreeritakse reaktiivsuse seiskohast. Küll on aga tegemist potentsiaalse vesiniksidemete aktseptoorsusega.

Modifitseeritud pöörduvate mehhanismide eristamiseks valiti maksimaalne positiivne osalaeng vesinikuaatomil Q^+_H ja maksimaalne negatiivne osalaeng raskel aatomil Q^-_A . Aromaatsete ühendite juures on deskriptorid märgatavas korrelatsioonis. Mida tugevam on asendajate induktsiooni- ja/või resonantsmõju, seda liikuvamaks muutub fookusesse tõstetud vesinik. Keskmiste Q^+_H väärtuste juures on tegemist vesiniksidemete donoorsusega, kõrgemal aga happelise dissotsiatsiooniga. Nende nelja deskriptori abil identifitseeritud mehhanismide topoloogia on esitatud Joonisel 4.



(a) Treeningandmekomplekt

(b) Testandmekomplekt

Alifaatsed ühendid: B - baassirge narkoos, A - amiinnarkoos, E - esternarkoos, X - ADH aktiivatsioon, C - karbonüülühendid (Nu liitumine karbonüülrühmale), M - Michaeli-tüüpi aktseptorid (Nu liitumine aktiveeritud C-C kordsele sidemele), H - halogeenderivaatide narkoos (S_{N2} -tüüpi nukleofiilne asendusreaktsioon), S - halogeenderivaatide reaktiivsus (S_{N1} -tüüpi nukleofiilne asendusreaktsioon).

Aromaatsete ühendid: 0 - mittepolaarne fenoolnarkoos, 1 - mittepolaarsed pro-kinoonid (fenoolsed pro-kinoonid), 2 - polaarne fenoolnarkoos, 3 - polaarset pro-kinoonid (aniliinsed pro-kinoonid), 4 - mittepolaarne WARU, 5 - polaarne WARU.

Joonis 4. Mehhanismide topoloogia alifaatsete ja aromaatsete ühendite andmekomplektile.

Alifaatsed ühendid

Baassirge narkoos (Joonis 4: 'B') jagas kõrvuti amiinnarkoosiga (Joonis 4: 'A') madalaimaid E_C väärtusi. Nende N_A väärtused on aga skaala erinevates otsades. Kohe tuli

ilmseks aga E_C petlikkus. Nimelt, baassirge narkoosi tuumikpiirkonnast leiti küllastatud alkoholid ja nitroalkaanid, kuid mitte ühtegi küllastamata alkoholi ega nitrili. Eelnevast kogemusest on aga hästi teada, et kõik nimetatud keemilised klassid on füsioloogilistes tingimustes ühevõrra inertsed. Närvivõrku uurides avastati kaks otsitavat keemilist klassi märksa kõrgemate E_C väärtuste juurest. Kui LUMO on tõsnud baassirge narkoosile iseloomulikku kõrgusesse siis on tegelikult päris ükskõik kas ta on vähem (sp^3 süsinik) või rohkem (sp^1 ja sp^2 süsinikud) lokaliseeritud. Testandmekomplektis sisaldunud aminoalkoholid ja aminoetrid paigutusid suuremas osas amiinnarkoosi klassi piiritlenud tühjadesse neuronitesse. Selle põhjuseks oli mõnevõrra suurenenud E_C ja vähenenud N_A .

ADH aktiveeruvad alkoholid kuuluksid deskriptorite väärtuste alusel baassirge narkoosi alla just nagu kõik muudki küllastamata alkoholid. Hoolikamal vaatlusel oli võimalik avastada, et kordse sideme ja hüdroksüülrühma sidestatus peegeldus väikestes kuid reeglipärastes nihetes, millest piisas täiesti eraldiseisva klassi piiritlemiseks (Joonis 4: 'X'). Q^+_H rohkem ja Q^-_A vähem tundsid ära, et kaheksal treeningandmekomplektis sisaldunud normaalahelalisel alkoholil asus hüdroksüülrühm otsmises asendis ja kahel seesmises asendis. Testandmekomplektis täiendavalt sisaldunud hargahelalised alkoholid ei tekitanud närvivõrgule probleeme.

Kõige kõrgemad E_C väärtuseid näitasid karbonüülühendid (Joonis 4: 'C'). Sinna lähedale jõudsid ka estrid ja laktoonid, kuid sellesse suhtuti teatava reservatsiooniga. Estrid hõlmasid väga laia piirkonna (Joonis 4: 'E'), sest neid iseloomustab uuritud keemilistest klassidest suurim konformatsiooniline paindlikkus (mõne kõrgema diestri konformeere leidus kuni paarikümnel neuronil). E_C järgi ei õnnestunud üheselt tuvastada Michaeli-tüüpi aktseptoreid (Joonis 4: 'M'). Nende LUMO on ühtlaselt laiali määratud üle kaksiksideme(te) ja aktiveerivast polaarsest rühmast moodustunud konjugatsiooniahela. Reaktsioonitsentriks oleva süsinikaatomi panus kahaneb proportsionaalselt konjugatsiooniahela pikkusega. Kuna puudus selge juhtiv deskriptor oli klass laiali pillutatud alaklassidena aktiveeriva rühma polaarsuse Q^-_A järgi (vastavalt nitrilid, aldehüüdid, ketoonid, estrid). Fukui indekse proportsionaalsus mitme samaväärse maksimumi korral tekitas suurt segadust. Kuna näiteks diketoonide E_C väärtuseks raporteeriti süstemaatiliselt vaid ligikaudu pool vastavate monoketoonide E_C väärtusest, siis tuli neid kaalukaardil otsida valge tooniga piirkonna asemel halli tooniga piirkonnast.

Nukleofiilsete asendusreaktsioonide jaoks peab struktuuris sisalduma arvestatav

elektrofiilsustsenter ja hea kuni väga hea lahkuv rühm. Mustri muudab veelgi keerukamaks vajadus arvestada S_N1 - ja S_N2 -tüüpi reaktsioonimehhanismidega. Kuna E_C ja N_A ei andnud üksinda ega koos Q^+_H ja Q^-_A kombineeritult piisavat indikatsiooni probleemi automaatseks lahendamiseks, teostati see käsitsi. Nagu tulemused näitasid sisaldus S_N1 -tüüpi reaktsioonimehhanism (Joonis 4: 'S') nende nelja deskriptori alusel täielikult S_N2 -tüüpi reaktsioonimehhanismi (Joonis 4: 'H') sees.

Testandmekomplektis sisaldunud üksikud elektrofiilsed karbonüülühendid tuvastati korrektselt. Halli piirkonna ääred, kus treenimisel leiti problemaatilised dikeetoonid ja Michaeli-tüüpi aktseptorid (Joonis 4: 'C', 'M'), olid asustatud halogeenitud karbonüülühenditega. Halogeenitud estrid kogunesid amiinnarkoosi ja esternarkoosi vahel asunud tühjadesse neuronitesse (Joonis 4: 'A', 'E'). Tegemist oli parima kohaga kus estrite tüüpilistele omadustele (kõrge E_C ja Q^-_A) õnnestus vähima vaevaga lisada halogeeniaatomi nukelofiilsus (kõrge N_A). Lõppviimistlus, et millised halogeenitud estrid toimivad S_N1 - ja millised S_N2 -tüüpi reaktsioonimehhanismi järgi, tuli jällegi teostada käsitsi.

Aromaatsed ühendid

Fenoolide klassifitseerimiseks piisas põhimõtteliselt kahe piirjoone tõmbamisest. Kõigepealt lahutati üksteisest mittepolaarsete (Joonis 4: '0', '1', '4') ja polaarsete fenoolide (Joonis 4: '2', '3', '5') domeenid. Otsuse tegemine kujunes oodatust raskemaks, sest piirjoon pidi omama ka rahuldavat keemilist selgitust. Saadaolevate omaduste hulgast peegeldas struktuuri EDG ja EWG sisaldust kõige paremini laengujaotuse deskriptor Q^-_A . Domeenide piirjoon tõmmati sellise Q^-_A väärtuse järgi, millest ühele poole jäid ühe või enama EWG asendatud fenoolid ning teisele poole ainult EDG asendatud fenoolid. EWG silmapaistvusele mõjus positiivselt erinevat tüüpi asendajate kombineerimine. Q^-_A puuduseks võib pidada asjaolu, et laengujaotuse (eba)ühtlus ei räägi iseenesest midagi selle kohta kas tuuma elektronitihedus on tervikuna vähenenud või suurenenud. Vastavat informatsiooni saab mingil määral hankida Fukui indeksitest E_C ja N_A .

Seejärel lahutati kumbki domeen Q^+_H alusel konkreetseteks WARU ja narkoosi klassideks. Mittepolaarne WARU mehhanism (Joonis 4: '4') koosnes vähemalt tri- ja tetrahalogeenitud fenoolidest. Halogeeniaatomid on mahukad ja vähepolaarsed ning avaldavad toimet peaaesjalikult induktsiooniefektide kaudu. Kloro- ja bromoderivaadid on väga kõrge hüdrofoobsusega, nende puhul on akuutse toksilisuse kõrval märkimisväärselt probleemiks toiduahelas akumulatsioon. Polaarne WARU mehhanism

(Joonis 4: '5') koosnes di- ja trinitrofenoolidest ning halogeenitud orto- ja para-mononitrofenoolidest. Nitrorühmad (neid võivad teoreetiliselt asendada ka muud tugevad EWG) avaldavad oma toimet läbi resonantsiefektide, mis on palju tugevamad kui induktsiooniefektid. Vastavad fenolaatioonid on juba piisavalt stabiliseeritud, et olla iseseisvalt vesilahustes püsivad. Fenolaatiooni püsivus osutub mingil hetkel juba puuduseks, sest dissotsiatsioonitasakaalu pööramine muutub raskeks. Sellega võib põhjendada näiteks trinitrofenooli kõrgemat isoefektiivset kontsentratsiooni võrreldes dinitrofenoolidega. Käesolevas töös kasutatud kahe WARU mehhanismi vastandamist võib pidada teadusharus esmakordseks.

Eelnevast kogemusest oli teada, et mittepolaarse fenoolnarkoosi hulgas peavad lõppude-lõpuks peituma ka mittepolaarsed pro-kinoonid (Joonis 4: '1') ja polaarsete fenoolnarkoosi hulgas polaarsete pro-kinoonid (Joonis 4: '3'). Klasside uurimisel leiti nad laialipillutatult, millest järeldus, et ükski neljast deskriptorist ei kirjelda omadust, mis oleks bioaktivatsiooni toimumise või mittetoimumise seisukohast määrav.

4.1.2 Leitud klasside regressioonianalüüs

Alifaatsed ühendid

Baassirge-, amiin- ja esternarkoos näitasid modifitseerimata pöörduvale mehhanismile tüüpilisi toimeprofile (Joonis 3: a+e). Praktilisteks vajadusteks piisab täiesti üheparameetristest $\text{Log}K_{ow}$ põhistest mudelitest. Lisaparameetri kaasamine on kahesuguse mõjuga. Ta annab väikese võidu treeningandmekomplekti peal, kuid avab täiendava võimaluse statistiliste hälbijate ilmutamiseks test- ja tööandmekomplektide peal.

Graafikul mis kujutas toksilisust funktsioonina $\text{Log}K_{ow}$ 'st asetused kõik kolme keemilist klassi üksteisest lahku. Amiinid olid toksikodünaamilise kontrolli piirkonnas ühesuguse $\text{Log}K_{ow}$ väärtuse juures 2-3 korda toksilisemad kui küllastatud alkoholid, mis omakorda olid 3-5 korda toksilisemad kui estrid. Baassirge toksilisuse kontseptsiooni alustalasid raputab eriti viimane tähelepanek. Kirjanduses on võimaliku selgitustena välja pakutud hüdroolüüsi (ühedest molekulist estrist tekib kaks molekuli samuti narkoosi järgi toimivaid hüdroolüüsiprodukte) [17]. On küsitav kas täiesti õige, kuid märksa lihtsama selgituse pakub LSER. Nimelt, kui tähele panna, et amiinid eksisteerivad vesilahuses positiivselt laetud ammooniumkatioonidena (vee protoneeriv toime), küllastatud alkoholid neutraalsete osakestena ja estrid negatiivset osalaengut (karboksüülne hapnik) kandvate osakestena, siis taandub kõik nende erinevatele elektroonsetele omadustele.

Vajadusest arvestada rohkemate teguritega räägib kõnekalt fakt, et peamiselt ühealuseliste normaalahelaliste alkoholide järgi koostatud baassirge narkoosi mudel näitas testandmekomplektis sisalduvate mitmealuseliste alkoholide jaoks arvestatavat ülehindamisviga ning hargahelaliste alkoholide (eriti tertsiaarsed alkoholid) jaoks alahindamisviga.

Aminoalkoholide ja aminoetrite andmetes puudus variatsioon. Neutraliseerimata ja neutraliseeritud (ammooniumisoolad) ühendite toksilisuse $p(IGC_{50})$ väärtused paiknesid vastavalt vahemikus 2.10 kuni 2.40 ühikut ja 1.10 kuni 1.50 ühikut. Nende toimeprofiil vastas (bioaktivatsiooniga algavale?) pöördumatule mehhanismile (Joonis 3: c). Kuna tegemist on väga hüdrofiilsete ühenditega siis võis nende MOA kujunemises rolli mängida ka näiteks osmootne šokk.

ADH aktiveeruvad alkoholid on esindatud tervikliku homoloogilise reaga $C_3 \dots C_9$. Tänu piisavale hulgale andmepunktidele nii toksikodünaamilise kontrolli kui toksikokineetilise kontrolli piirkonnas joonistus välja väga esinduslik bioaktivatsiooniga algava pöördumatu mehhanismi toimeprofiil (Joonis 3: b(c)+e). Murdepunktiga joont õnnestus modelleerida Kier&Hall 3'ndat järku indeksiga $^3\chi$, mis on märgatavalt mittelineaarne topoloogiline deskriptor. Mudel ennustas testandmekomplektist hästi normaalahelaliste alkoholide toksilisust kuid mitte hargahelaliste alkoholide toksilisust, sest $^3\chi$ läheb oma algatuslikult kaasa raskete aatomite vaheliste ühenduste (*connection paths*) keerukuse kasvuga.

Karbonüülühendite klassi moodustasid aldehüüdid, monoketoonid, diketoonid ja laktoonid. Graafikul mis kujutas toksilisust funktsioonina $\text{Log}K_{ow}$ 'st näitavasid nendest pöördumatule mehhanismile omast toimeprofiili (Joonis 3: c+e) ainult aldehüüdid. Võimalusest lahutada aldehüüdid ülejäänud karbonüülühenditest ning tuletada kaks mudelit loobuti, sest nagu nähtud eespool ADH aktiveeruvate alkoholide juures ja nagu saab veel näha allpool Michaeli-tüüpi aktseptorite juures, on murdepunktiga graafikute modelleerimine üpris vaevanõudev ja vähetulemuslik töö. Seega, aldehüüdide homoloogilise rea kuus alamat liiget $C_3 \dots C_6$ tunnistati statistilisteks hälbijateks. Neile lisandusid liigtoksilisuse alusel diketoonide hulgast kuus α -diketooni ja laktoonide hulgast kaks β -(propio)laktooni. Esiletõstetud ühendite liigtoksilisust võis selgitada orgaanilise keemia põhitõdede abil. Arvutused näitavad, et erinevalt monoketoonidest ning β - ja γ -diketoonidest on α -diketoonides potentsiaalne reaktsioonitsenter steeriliselt suhteliselt vähe varjestatud. Kaks kõrvutiasetsevat karbonüülrühma moodustavad tasapinna, mis on

nukleofiilidele ligipääsetav. Edaspidi, kui võimalikuks saab närvivõrku kaasata (automaatselt positsioneeritavaid-) fragmendipõhiseid deskriptoreid, peaks nii elektrofiilsustsenter kui nukleofiilsustsenter olema varustatud steerilist varjestatust puudutava informatsiooniga. Kolme- ja neljalülilise tsükliga laktoonid on toksilisemad kui viie- või kuuelülilise tsükliga laktoonid nn. pingeenergia tõttu. Peale tsükli katkemist (näiteks hüdrolüüsil) käitub tekkiv hüdrosüühape baassirge narkoosi mehhanismi kohaselt. Allesjäänud karbonüülühendid näitasid modifitseerimata pöörduvale mehhanismile omast toimeprofiili (Joonis 3: a+e). Mitme keemilise klassi kokkutoomiseks vajalik paindlikkus saavutati kahe deskriptori koostöös, millest esimene mõõtis suhtelist positiivse laenguga molekulipindala ja teine C-O sideme elektroonseid omadusi.

Michaeli-tüüpi aktseptorites asub elektrofiilsustsenter β -süsinikul, mis on nukleofiilidele vaikimisi hästi ligipääsetav. Karbonüülühenditega võrreldes on ka reaktsiooni siirdeolek (*transition state*) energeetiliselt märksa soodsam, sest tugevasti polaarne aktiveeriv rühm aitab resonantsiefektidega laengutihedust ühtlustada. Kõik neli aktiveeriva rühma järgi eristatud alaklassi näitasid ilmekaid pöördumatule mehhanismile omaseid toimeprofiile (Joonis 3: c+e). Iga alaklass tundus omama iseloomulikku toksilisuse väärtust. Isoefektiivne kontsentratsioon vähenes üle kolme suurusjärgu reas nitrilid, estrid, ketoonid, aldehüüdid. Võrreldes muude uuritud juhtudega asus siin enamik andmepunkte toksikodünaamilise kontrolli regioonis mitte toksikokineetilise kontrolli regioonis. Modelleerimine on võimalik ainult siis kui alaklasse vaadeldakse koos. DeWeese jkt. [24] on nähtust väga tabavalt kirjeldanud kui analoogse andmekomplekti saamist homologsete andmekomplektide kombineerimisel ("*obtaining analogue dataset by combining homologue datasets*"). Kui üksikute alaklasside toksilisus oli enam-vähem konstantne ja ei andnud rahuldavat korrelatsiooni mitte ühegi ekstensiivuse ega intensiivsusparameetriga, siis neljast alaklassist moodustus "toksilisuse redel", mis andis hea korrelatsiooni paljude intensiivsusparameetritega. Karbonüülühendite paragrahvis ei viidudki aldehüüde eraldi alaklassi põhjusel, et puudus võimalus analoogse andmekomplekti moodustamiseks. Aldehüüdide toksilisus on toksikodünaamilise kontrolli regioonis konstantne nagu ka kõik nende karbonüülrühma kirjeldavate intensiivsusparameetrite väärtused. Koostatud mudel mõõtis maksimaalset elektronitihedust kaksiksidemel. Tähelepanu tasub pöörata sellele, et struktuurselt piiritlemata deskriptorid võivad anda ootamatuid tagasilööke. Treeningandmekomplektist tuli välja arvata üks struktuur (2,4-

dimetüül-2,6-heptadienaal), sest selles sisaldus konjugeerunud (so. aktiveeritud) kaksiksideme kõrval ka konjugeerumata kaksikside, mis juhtis deskriptori oma suurema elektronihedusega valedete jälgedele. Konjugeerumata kaksiksideme kõrval võib sama efekti anda ka konjugeerumata kolmikside.

Kvalitatiivses analüüsis oletati, et tavalised halogeenderivaadid osalevad S_N2 -tüüpi nukleofiilsetes asendusreaktsioonides. Graafikul mis kujutas toksilisust funktsioonina $\text{Log}K_{ow}$ 'st näitasid nad aga modifitseerimata pöörduvale mehhanismile omast toimeprofiili (Joonis 3: a+e). Teistest hälbis ainsana 1,3-dibromo-2-propanool. Tema liigtoksilisust selgitaks paremini hüpotees spetsiifilisest mehhanismist kui pöördumatust mehhanismist, sest vastav koloroderivaat 1,2-dikloro-2-propanool, nagu ka kõik monobromoderivaadid üldisemalt, oli mitteaktiivne. Koostatud üheparameetriline $\text{Log}K_{ow}$ põhine mudel näitas suurepäraseid tulemusi ka testandmekomplektis sisalduvate halogeenitud estrite jaoks. Baassirge narkoosi ja halogeenderivaatide narkoosi mudelid erinesid üksteisest nii sirge tõusu kui vabaliikme osas. See kinnitab veendumust, et suuremaid keemilisi klasse tasub nende mehhanismide eripära tundmaõppimiseks alati proovida modelleerida korra ka eraldiseisvana.

Treeningandmekomplekti kuulunud α -halonitriilide ning testandmekomplekti kuulunud α -haloestrite ja α -haloketoonide aktiivsuses ei tekkinud mingit kahtlust, sest nad näitasid pöördumatule mehhanismile iseloomulikku toimeprofiili (Joonis 3: c+e). Loomulikult pole üksnes selle graafiku alusel võimalik kinnitada ega ümber lükata, et reaktsioon toimub tõesti S_N1 -tüüpi nukleofiilse asendusreaktsioonina. Iga alaklassi jaoks koostati eraldi mudel. Kuna ühendi reaktsioonivõime sõltub rohkem dissotsieeruvate halogeeniaatomite arvust ja tüübist kui aktiveerivast rühmast, siis vastas iga alaklass ise piisavalt analoogse andmekomplekti nõudmistele. Elektroonsed deskriptorid ja laengujaotuse deskriptorid koosnevad kaheparameetrilised mudelid olid oletatava reaktsioonimehhanismi seisukohast interpreteeritavad ning hea statistilise kvaliteediga. Kahjuks ilmnisid deskriptorite täpsemal uurimisel juba tuttavad probleemid seoses umbmäärase positsioneeritusega.

Aromaatsed ühendid

Kõigi nelja mehhanismi koosmodelleerimine andis kesiseid tulemusi. Parameetrite valik oli dikteeritud mittepolaarse fenoolnarkoosi kui kõige suurema esindatusega mehhanismi poolt. Samas jäi üldise mudeli statistiline kvaliteet mittepolaarse fenoolnarkoosi

esialgsele kaheparameetrilisele mudelile selgesti kättesaamatuks. Statistiliste hälbijate identifitseerimine näitas, et tegemist oli 1,2- ja 1,4-benseendioolide mõjuga. Esialgne mudel alahindas benseendioolide toksilisust ja ülehindas nende tasakaalustamiseks vastavate alkoksüfenoolide toksilisust. Kuna nii hüdroksüülrühm kui alkoksürühm on lähedase tugevusega EDG, siis ei tule nende diskrimineerimisega toime ükski kvalitatiivses analüüsis kasutatud deskriptor. Mittepolaarse fenoolnarkoosi mudeli statistiline kvaliteet paranes hüppeliselt peale kõigi 1,2- ja 1,4-benseendioolide üleviimist eraldi seisvasse mittepolaarsesse pro-kinoonide mehhanismi. Parim mudel mittepolaarse fenoolnarkoosi jaoks koosnes $\text{Log}K_{ow}$ ja madalaima vaba molekuloorbitaali energiast E_{LUMO} . Sellised kaheparameetrilised mudelid on tänu oma lihtsusele võrdlemisi levinud ja soositud.

Käesolevas töös erines pro-kinoonide mehhanismide käsitus teadusharus tavapäraseks peetavast käsitlusest selle poolest, et nende identifitseerimine toimus alles teises järjekorras vastavate fenoolnarkooside mehhanismide seest, mitte esimeses järjekorras kogu andmekomplektist. Sellise otsusega välistati, et pro-kinoonse aktiivsusega saaks selgitada EWG asendatud või tetrahalogeenitud benseendioolide liigtoksilisust. Võrdluseks, Schultz jkt. [50] ja nende andmekomplekti taaskasutajad on pikalt polemiseerinud vajaduse üle tõsta tetrakloro- ja tetrafluoro-1,4-benseendiool (koos tetrametüül-1,4-benseendiooliga) eraldi alamehhanismi, sest neis puudub pro-kinoonide peamehhanismi seisukohast vajalik asendamata süsinikuaatom. Parim mudel mittepolaarsete pro-kinoonide jaoks koosnes ühest topoloogilisest deskriptorist ja kahest elektoonsest deskriptorist. Bioaktivatsiooniga algavate pöördumatute mehhanismide mudelite interpreteerimine on üldiselt üsna kaheldava väärtusega tegevus, sest arvestada tuleb konkureerivate ja/või astmeliste radadega kus tekkivad vaheühendid on lähteainest vägagi erinevate omadustega.

Polaarse fenoolnarkoosi esialgne mudel kannatas analoogiliselt mittepolaarse fenoolnarkoosi esialgsele mudelile harvade kuid vägagi silmapaistvate statistiliste hälbijate all. Probleem lahendati taaskord 1,2- ja 1,4-aminofenoolide üleviimisega eraldiseisvasse polaarsesse pro-kinoonide mehhanismi. Kordusmodelleerimine näitas, et see toksikofoor pole täielik (järelkult ka mitte sisuline), sest ta ei hõlma 1,3-aminofenoolide liigtoksilisust. Viimaste esindaja 5-amino-2-metoksüfenool viidi samuti üle polaarsesse pro-kinoonide mehhanismi. Parim mudel polaarse fenoolnarkoosi jaoks oli kolmeparameetriline. $\text{Log}K_{ow}$ suutmatust kohanduda dissotsiatsiooniefektidele korrigeeriti

rohkem suhtelise positiivse osalaenguga pindala abil ja vähem süsinikuaatomite sidemejärkude abil. Parim mudel polaarsete pro-kinnonide jaoks koosnes kolmest elektroonsest deskriptorist. Ekstensiivsuspameetri väljatõrjumist intensiivsuspameetritega selgitas asjaolu, et andmekomplekt asus täielikult toksikodünaamilise kontrolli piirkonnas. Mittepolaarsete ja polaarsete pro-kinoonide ühendamist ei üritatud.

Mittepolaarne ja polaarne fenoolnarkoos näitasid ühesugust modifitseerimata pöörduvale mehhanismile iseloomulikku toimeprofili (Joonis 3: a+e). See on üpriski hea argument kinnitamaks varasemat seisukohta, et mononitrofenoolid ei valda vesikeskkonnas märgatavat elektrofiilset reaktsioonivõimet.

Mittepolaarse WARU liigtoksilisuse arvutamiseks pole baassirgena võimalik kasutada mittepolaarse fenoolnarkoosi mudelit, sest ühendite E_{LUMO} väärtused asusid selle tööpiirkonna alumisest otsast rohkem kui 0.50 eV madalamal. Selline E_{LUMO} tüüpiliste vahemike võrdlemine pakkus aga veelkordset kinnitust, et tetrahalogeenitud benseendioolid käituvad pigem WARU mehhanismi järgi ($E_{LUMO} < -1.0$ eV) kui pro-kinoonse mehhanismi järgi ($E_{LUMO} > 0.0$ eV). Üheparameetrilise $LogK_{ow}$ põhise mudeli järgi oli hüdrofiilsete fluoroderivaatide liigtoksilisuse väärtused kuni 1.5 log ühikut ja hüdrofoobsete kloro- ja bromoderivaatide omad kuni 0.5 log ühikut. Arvestatava liigtoksilisuse säilimine ka kõige hüdrofoobsemate ühendite juures muutis tegeliku toimeprofili ennustatud toimeprofiilist veidi laugemaks (Joonis 3: b(c)+e). Parim mudel mittepolaarse WARU jaoks koosnes kahest elektroonsest deskriptorist mida oli võimalik interpreteerida induktsiooniefektide kaudu. Parim mudel polaarsete WARU jaoks koosnes $LogK_{ow}$ 'st ja kahest elektroonsest deskriptorist mida oli võimalik interpreteerida resonantsiefektide kaudu. Kui võrrelda omavahel polaarsete WARU ja polaarsete fenoolnarkoosi mudeleid siis hakkab silma, et $LogK_{ow}$ kirjeldab täielikult dissotsieerunud ühendeid paremini kui osaliselt dissotsieerunud ühendeid.

4.2 Aromaatsete ühendite laiendatud andmekomplekt

Esimese andmekomplektiga saadud tulemused näitasid, et töövoog omab suuremat suhtelist eelist aromaatsete ühendite puhul. Fenoolide täpsemaks uurimiseks ja nende suhestamiseks teiste keemiliste klassidega otsustati analüüsi korrata uue ja mitmekülgsema andmekomplektiga.

Erinevatest kirjandusallikatest koguti 465 ühendist koosnev andmekomplekt, mis koosnes ligikaudu 30% ulatuses benseenidest, 20% ulatuses aniliinidest ja 50% ulatuses

fenoolidest. Teadlikult välditi heteroaatomeid sisaldavaid ja mitmetuumalisi struktuure. Lisaks jälgiti, et kõik polaarsed funktsionaalrühmad saaksid osaleda vähem või rohkem vahetutes resonantsiefektides tuumaga. Seda mittetäitvad ühendid, näiteks β -, γ - jne. fenüülalkoholid, käituvad oma keemiliste omaduste poolest lähedasemalt alifaatsetele ühenditele kui aromaatsetele ühenditele. Lähteandmekomplekt jagati juhuvalikuga 3:1 suhtes 350 ühendist koosnevaks treeningandmekomplektiks ja 115 ühendist koosnevaks testandmekomplektiks. Paari ühendi korral tuli juhuvalikut käsitsi korrigeerida, et kindlustada, et treeningandmekomplekt jääks katma laiemat omadustevahemikku kui testandmekomplekt.

4.2.1 Klassifitseerimine iseorganiseeruva närvivõrguga

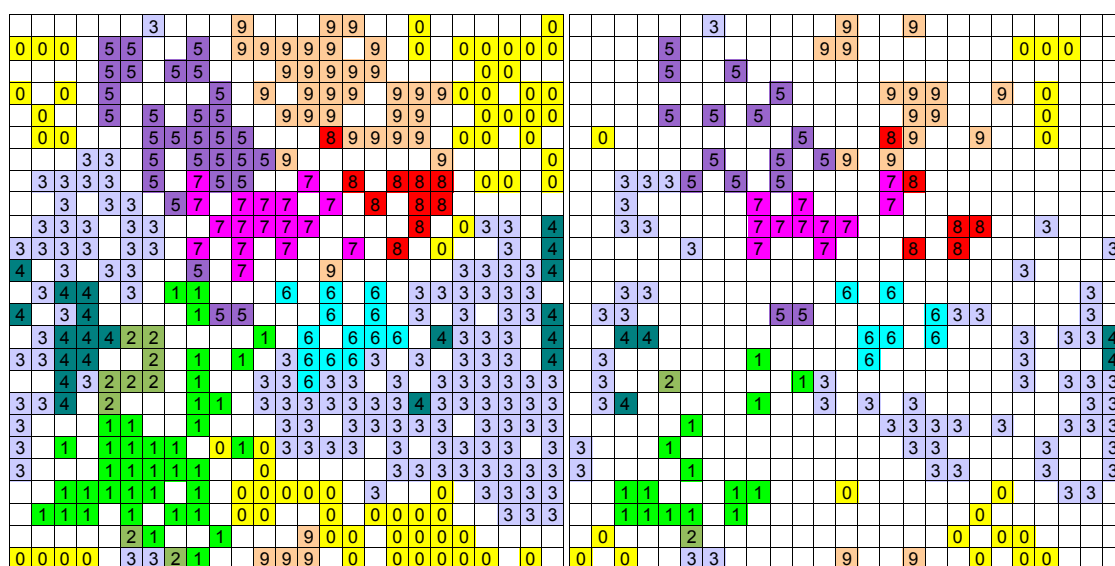
Aromaatsset ühendit on mõistlik kirjeldada tervikuna, kus üksikute funktsionaalrühmade omaduste asemel tõusevad esiplaanile molekuli omadused.

Esimene andmekomplekt kinnitas, et pöörduvad mehhanismid on võrreldes pöördumatute mehhanismidega selges ülekaalus. Omadus mille alusel märgitakse nende vaheline piirjoon peaks võimaldama senisest konkreetsema nn. lävendiväärtuse (*threshold value*) määramist. Lävendiväärtuste määramine toimub kvalitatiivse ja kvantitatiivse analüüsi tihedas koostöös. Kui esimene loetleb võimalikud variandid, siis teine valib neist toimeprofiilide iseloomulikkuse alusel välja kõige tõenäolisema.

Fukui indeksid N_A ja E_C vahetati välja vastavate piirorbitaalide energiatega E_{HOMO} ja E_{LUMO} vastu. Molekuloorbitaalide energiad on absoluutsed, mis tagab suurepärase võrreldavuse nii ühe keemilise klassi sees kui mitme keemilise klassi vahel. Deskriptorite madalad ja keskmised väärtused tõlgendatakse inter- ja intramolekulaarsete interaktsioonidena (van der Waalsi jõud, vesiniksidemed), kõrged väärtused reaktiivse toksilisusena. Omadustevektori koostamise käigus teostatava normaliseerimise tulemusena muutuvad E_{HOMO} ja E_{LUMO} sõltuvatest suurustest sõltumatuteks suurusteks. Probleem lahendati uue deskriptori E_{gap} sissetoomisega, mis näitab piirorbitaalide omavahelist eraldatust ($E_{gap} = E_{LUMO} - E_{HOMO}$). E_{gap} tuntakse paremini keemilise kõvadusena (*chemical hardness*). Orgaanilisest keemiast on hästi teada, et reagent eelistab interakteeruda temale lähedase kõvadusega substraadiga.

Piirorbitaalide energiad täiendavad laengujaotuse deskriptoreid Q^+_H ja Q^-_A molekuli polaarsuse määramisel. Aktiveerivad EDG tõstavad nivoosid, deaktiveerivad EDG madaldavad nivoosid vähe ning EWG palju. Asendajate efektid on aditiivsed. Mitu sama-

suunalist asendajat annavad tugevdatud trendi, mitu vastassuunalist asendajat (aktiveerivad EDG üheloop ning deaktiveerivad EDG ja EWG teiseloop) annavad kas nõrgendatud või päris ebaselge trendi läbi üksteise mõju tasakaalustamise. Nende viie deskriptori abil identifitseeritud mehhanismide topoloogia on esitatud Joonisel 5.



(a) Treeningandmekomplekt

(b) Testandmekomplekt

Aromaatsed ühendid: 0 - benseennarkoos, 1 - aniliinnarkoos, 2 - aniliinsed pro-kinoonid, 3 - mittepolaarne fenoolnarkoos, 4 - fenoolsed pro-kinoonid, 5 - polaarne fenoolnarkoos, 6 - mittepolaarne WARU, 7 - polaarne WARU, 8 - nitrobenseenide elektrofiilsus, 9 - nitrobenseenide narkoos.

Joonis 5. Mehhanismide topoloogia aromaatsete ühendite laiendatud andmekomplektile.

Kolm peamist keemilist klassi asetused üksteisest lahku Q^+_H alusel. Madalaimad Q^+_H väärtused on omased benseenidele, millest suure hulga moodustavad nitrobenseenid, keskmised Q^+_H väärtused aniliinidele ja keskmised kuni kõrged Q^+_H väärtused fenoolidele. Mida suurem on keemilise klassi protogeenne aktiivsus (see mõiste hõlmab eneses nii happelist dissotsiatsiooni kui vesiniksidemete teket), seda rohkem ja erinevaid toksilisuse mehhanisme tuleb tema puhul vesikeskkonnas kõne alla.

Benseenide piirkond (Joonis 5: '1') jagunes piirorbitaalide energiatega E_{HOMO} ja E_{LUMO} põhjal mitmeks osaks. Kõige madalamal asusid dinitrobenseenid ja polühalogeenitud mononitrobenseenid (Joonis 5: '8'), neist veidi kõrgemal ülejäänud mononitrobenseenid (Joonis 5: '9'). Tegemist on saasteainetega mille juurest tasuks esimesena otsida pöördumatule mehhanismile (so. elektrofiilsus) omast toimeprofiili (Joonis 3: d). Nitrorühm oli andmekomplektis kaugelt kõige tugevam EWG. Mononitrobenseenide põhiosa ($E_{LUMO} < -1.3 \dots -1.0$ eV) ja kõigi teiste benseenide ($E_{LUMO} > 0.5$ eV) vahele jäi lai vahe-

mik. Tõsi küll, nitrorühmaga võrreldava tuuma elektronitiheduse vähenemiseni jõudsid ka mõned täielikult halogeenitud ühendid nagu näiteks pentafluorobensaldehüüd ja pentafluorobensüülalkohol, kuid nende puhul on keemilise reaktsiooni kulgemine märksa vähetõenäolisem, sest puudub võimalus vaheühendite stabiliseerimiseks resonantsiefektide kaudu.

Nitrobenseenid võeti kokku esialgse nitrobenseenide toksilisuse mehhanismi nime all. Nagu eespool kirjeldatud saab selle konkreetseteks nitrobenseenide narkoosi ja nitrobenseenide elektrofiilsuse mehhanismideks jagada alles siis kui kvantitatiivse analüüsi käigus on leitud piirorbitaalide energiatega lävendiväärtus. Kõik ülejäänud benseenid moodustasid benseennarkoosi mehhanismi.

Polaarse WARU moodustanud nitroaniliinid ja nitrofenoolid asuvad närvivõrgul kohe nitrobenseenide kõrval (Joonis 5: '7'). Rekordkõrged Q^+_H väärtused osutavad, et nad ei ole vesikeskkonnas elektrofiilidena toimimiseks piisavalt stabiilsed. Tänu Fukui indeksite väljavahetamisele piirorbitaalide energiatega vastu oli polaarne WARU sarnasuskardil paremini piiritletud. Tema kirjeldus toksikofoorina jäi aga täpselt samaks, kui jätta arvestamata, et protogeensete rühmade nimistusse lisandus hüdroksüülrühma kõrvale samaväärsena ka aminorühm.

Mittepolaarse WARU ja mittepolaarse fenoolnarkoosi mehhanismide vaheline üleminek jäi endistviisi pidevaks (Joonis 5: '6'). Esimese andmekomplekti abil paikapandud piirjoone tõlkimine Q^+_H pealt frontaalorbitaalide energiatega peale andis vähemalt halogeenitud fenoolide osas täiendavat selgust. E_{LUMO} lävendiväärtust -0.50 eV rahuldasi fenoolid millel asus hüdroksüülrühma suhtes orto- või para-asendis vähemalt kolm deaktiveerivat EDG või nõrka EWG. Mittepolaarse WARU esindajateks peeti teiste hulgas kolme dihalogeenitud 2-hüdroksübenseeni ja ühte dihalogeenitud 2-tsüanofenooli. Aniliinidest ületas lävendiväärtuse paraja varuga ainult pentafluoroaniliin. Sellest võis järeldada, et erinevalt resonantsiefektidest ei suuda induktsiooniefektid aminorühma lämmastiku elektronitihedust niivõrd vähendada, et tema prootonid muutuksid olulisel määral liikuvateks. Ühe aniliini lisamist fenoolide sekka välditi, sest see oleks halvendanud väljavaateid mehhanismipõhise (fenoolse hüdroksüülrühma happeline dissotsiatsioon) mudeli koostamiseks.

Järelejäänud ühendid jagati aniliinnarkoosi ning mittepolaarse ja polaarne fenoolnarkoosi mehhanismideks. Aminofenoolid ja benseendiolid, mis teevad läbi bioaktivat-

siooni kinoonideks, olid närvivõrgul sattunud aniliinnarkoosi ja mittepolaarse fenoolnarkoosi äärealadele kõrvuti (Joonis 5: '2', '4'). Seda piirkonda iseloomustavad väga kõrged piirorbitaalide energiatega väärtused, kusjuures nende vaheline kaugus E_{gap} on vähenenud, sest hõivatud orbitaal jõuab vabale orbitaalile suhteliselt kiiremini järele. Kirjeldatud molekuloorbitaalide pilt oli andmekomplektis unikaalne. Aminofenoolide alkoksüaniliinidest eristamisega sai närvivõrk iseseisvalt hakkama, sest Q^+_H diskrimineerib vaba hüdroksüülrühma järgi (Joonis 5: '1', '2'). Benseendioolide lahutamine alkoksüfenoolidest oli samade deskriptorite baasil võimatu (Joonis 5: '3', '4'). Vajaduse korral võib kaaluda närvivõrgu sisendruumi laiendamist eriotstarbelise koostisliku deskriptori, näiteks fenoolsete hüdroksüülrühmade arvu lisamise teel.

Käesolevas töös pakuti välja, et pro-kinoonsete struktuuride MOA tuleneb peaaesjalikult nn. oksüdatiivsest stressist (*oxidative stress*) kui nad kasutavad ära oma unikaalset piirorbitaalide pilti interakteerumaks ensüümide oksüdaasidega ja reduktaasidega. Selle abil õnnestus muuhulgas seletada 1,3-derivaatide aktiivsust ning konjugeeritud kõrvalahelatega 1,2- ja 1,4-derivaatide mitteaktiivsust.

4.2.2 Leitud klasside regressioonianalüüs

Benseennarkoos on oodatavalt küll kõige lihtsam mehhanism, kuid üheparameetiline $\text{Log}K_{ow}$ põhine mudel ei saavutanud head tulemust. Jääkvigade (*residual errors*) analüüs näitas heterogeensust, mida tuleb ikka ja jälle seletada asendajate polaarsuse ning vesiniksidemete moodustamise võimega. Ülehindamisviga oli iseloomulik asendamata süsivesinikele ja EDG asendatud derivaatidele, alahindamisviga EWG asendatud derivaatidele. Kõige paremini olid ennustatud bensüülalkoholid, mis analoogiliselt alifaatsetele alkoholidele valdavad tasakaalustatud vesiniksidemete doonor- ja akseptorsuutlikkust. Alaklasside koondamine on teostatav intensiivsuseparameetri(te) lisamise teel. Praegusest mahukamate andmekomplektide juures tuleks kaaluda erinevuste mahasurumise asemel nende rõhutamist. Paremal juhul võiks see välja viia benseennarkoosi lahutamiseni mittepolaarseks ja polaarseks benseennarkoosiks (analoogia fenoolnarkoosiga) ning bensüülalkoholide narkoosiks.

Aniliinnarkoosi jaoks koostatud esialgne mudel soovitas kõik aminofenoolid sõltumata nende rühmade omavahelisest paigutusest üle viia eraldiseisvasse aniliinsesse pro-kinoonide mehhanismi. Üheparameetiline $\text{Log}K_{ow}$ põhine mudel jättis jällegi soovida. Jääkvigade analüüs näitas proportsionaalsust elektronitihedusega. Mõningane

ülehindamisviga oli iseloomulik aktiveerivate EDG asendatud derivaatidele, märksa suurem alahindamisviga aga EWG (eriti nitrorühm!) asendatud derivaatidele. Üllataval kombel sobis mudelisse lisaparameetriks paremini vesinikuaatomite keskmine sidemejärk kui näiteks samasisulised piirorbitaalide energiad.

Aniliinnarkoosi mudeli abil arvatud aminofenoolide liigtoksilisuse väärtused jäid 1,2- ja 1,4-derivaatide jaoks tüüpiliselt vahemikku 0.7 kuni 2.0 log ühikut ja 1,3-derivaatide jaoks vahemikku 0.2 kuni 0.4 log ühikut. Tegemist on hüdrofiilsete ühenditega, mis paiknesid toksikodünaamilise kontrolli piirkonna alumises otsas (Joonis 3: c(b)). 1,2- ja 1,4-derivaadid on 1,3-derivaatidest väiksema keemilise kõvadusega tänu tugevatele resonantsiefektidele EDG rühmade vahel. Parim aniliinsete pro-kinoonide mudel haaras kinni sellisest kunstlikult tekkinud korrelatsioonist liigtoksilisuse ja E_{gap} vahel. Esimeses andmekomplektis leiti aminofenoolid polaarsete fenoolide (peamiselt mononitrofenoolid) hulgast. Liigtoksilisuse hindamisest loobumine oli nendes tingimustes ainuõige samm, sest nagu näha puudub neil klassidel arvestatav ühisosa nii struktuuri kui mehhanismi plaanis.

Mittepolaarse fenoolnarkoosi jaoks koostatud esialgne mudel esitles statistiliste hälbijatena 1,2- ja 1,4-benseendioole aga mitte 1,3-benseendioole. Võrreldes 1,3-aminofenoolidega raskendas 1,3-benseendioolide jaoks piirorbitaalide energiatega lävendiväärtuste saavutamist hapniku suurem elektronegatiivsus ning hüdroksüülrühma jagamata elektronipaari raskem delokaliseeritavus. 1,3-benseendioolide profiili täpsustamiseks läheks vaja 5-amino-2-metoksüfenoolile sarnanevaid rohkete aktiveerivate EDG asendatud derivaate. Mittepolaarse fenoolnarkoosi mudeli abil arvatud benseendioolide liigtoksilisuse väärtused jäid 1,2-derivaatide jaoks vahemikku 0.7 kuni 2.0 log ühikut ja 1,4-derivaatide jaoks vahemikku 1.2 kuni 3.0 log ühikut. Siiski, identse asenduspildi korral omasid esimesed teistest veidi madalamaid isoeffektiivseid kontsentratsioone. Andmekomplekti iseloomustas väike variatiivsus. Parim fenoolsete pro-kinoonide mudel koosnes ühest topoloogilisest deskriptorist ja ühest elektroonsesest deskriptorist.

Mittepolaarse ja polaarsete fenoolnarkoosi ning mittepolaarse ja polaarsete WARU jaoks saadud tulemused olid oma interpretatsiooni poolest väga lähedased esimese andmekomplektiga saadud tulemustele. $\text{Log}K_{ow}$ oli kõige edukam mittedissotsieeruvate osakeste jaotustasakaalu kirjeldamisel, ning suhteliselt edukam täielikult dissotsieerunud ühendite kirjeldamisel kui osaliselt dissotsieerunud ühendite kirjeldamisel. Fenoolne

hüdrosüülrühm pakub aminorühmaga võrreldes märksa keerukamaid interaktsioone, millega arvestamiseks vajati hapnikukeskseid elektroonseid deskriptoreid.

Esiolgne nitrobenseenide toksilisuse mehhanism tõi korruga nähtavale kaks erinevat toimeprofiili. Dinitrobenseenid ja vähemalt trifluoreeritud mononitrobenseenid näitasid pöördumatule mehhanismile omast toimeprofiili (Joonis 3: d ja c), kõik ülejäänud nitrobenseenid aga modifitseerimata pöörduvale mehhanismile omast toimeprofiili (Joonis 3: a+e). Elektrofiilsuse E_{LUMO} lävendiväärtuseks kujunes nendel andmetel -1.85 kuni -1.70 eV.

Nitrobenseenid osalevad nukleofiilsetes asendusreaktsioonides. Kineetiliste mõõtmiste järgi on limiteerivaks staadiumiks sideme teke siseneva rühmaga mitte sideme katkemine lahkuva rühmaga [26]. Nukleofiil atakeerib eelistatult kõige väiksema elektrotihedusega (so. suurima positiivse osalaenguga) tuumasüsinikku. Kui struktuuris sisalduvad deaktiveerivad EDG (kujutavad endast ühtlasi head lahkuvat rühma) siis kujuneb märklauaks üks neid siduv süsinikuaatom, kui mitte, siis süsinikuaatom, mis on konjugeeritud asendis rohkemate nitrorühmadega. Mudelühendiks võib pidada 1-fluoro-2,4-dinitrobenseeni, mida tuntakse biokeemias paremini Sangeri reagendina ja mis leiab kasutamist proteiinide analüüsil otsimiste aminorühmade märkimisel.

Nitrobenseenide elektrofiilsuse mehhanismi täiendaval analüüsimisel ilmnes, et neljateistkümne halogeenitud derivaadi toksilisuse $p(IGC_{50})$ väärtused paiknesid vahemikus 4.71 kuni 5.82 ühikut ja viie halogeenimata derivaadi väärtused vahemikus 3.76 kuni 4.30 ühikut. Halogeenitud nitrobenseenid oli kogu töö ulatuses ainuke keemiline klass, mis näitas vaieldamatut liigtoksilisust ka toksikokineetilise kontrolli piirkonnas (Joonis 3: d). Mõlemad keemilised klassid modelleeriti siiski koos. Parim mudel koosnes kahest (valents)elektronidele keskendunud topoloogilisest deskriptorist. Mudeli statistiline kvaliteet oli hea kuid ta polnud eelkirjeldatud reaktsioonimehhanismi seisukohast interpreteeritav.

Nitrobenseenide narkoosi liikmed olid vaadeldavad nitreeritud benseennarkoosi liikmetena. Nitrorühm on tugev elektroni aktseptor ja massiivne ning ta tuli suurepäraselt toime kolme benseennarkoosi alaklassi koondamisega. Nitrobenseenide narkoosi üheparameetiline $\text{Log}K_{ow}$ põhine mudel ületas oma statistiliselt kvaliteedilt kõiki eeluletatud narkoosimudeleid. Mononitrobenseenide demonstreeritavalt modifitseerimata pöörduv mehhanism ei tohiks jätta mingit kahtlust, et sama kehtib ka mononitrofenoolide kohta.

5 KOKKUVÕTE

Vesikeskkonnas esinevad toksilisuse mehhanismid jagunevad nelja põhiklassi. Esimesel tasemel eristatakse pöörduvaid ja pöördumatuid mehhanisme. Teisel tasemel eristatakse konkreetsemalt modifitseerimata ja modifitseeritud pöörduvaid mehhanisme ning vahetuid ja bioaktivatsiooniga algavaid pöördumatuid mehhanisme. Saasteaine toksilisuse mehhanism on nelja põhiklassi lõikes etteennustatav tema füsikokeemilistest omadustest. Kahe andmekomplekti võrdlev analüüs näitas, et klassifikatsioon füsiko-keemilistele omadustele vastavusse seatud teoreetiliste molekulaarsete deskriptorite baasil omab võrreldes klassifikatsiooniga struktuurivalemite baasil suhteliselt suuremat eelist aromaatsete ühendite puhul kui alifaatsete ühendite puhul. Iseorganiseeruv närvivõrk võimaldab klassifitseerimist teostada mitmesuguse detailsusega. Käesolevas töös lahutati mitu teadusharuks ühtseks ja lõplikuks peetavat toksilisuse mehhanismi edasisteks alamehhanismideks. Saadud klassid on pidevad mitte diskreetsed. Iga klassi koosseisus saab välja tuua väga tüüpilisi ühendeid, keskmiselt tüüpilisi ühendeid ja vähetüüpilisi ühendeid. Viimased ilmuvad regressioonianalüüsil sageli statistiliste hälbijatena.

Toksikoloogilistele mehhanismidele regressioonianalüüsil tuletatud mudelite keerukus ja neis sisalduvate parameetrite iseloom sõltub otseselt põhiklassist. Modifitseerimata pöörduvad mehhanismid on suurepäraselt kirjeldatud juba ühest ekstsensivsusparameetrist (eksperimentaalne jaotuskoefitsient, geomeetrised või topoloogilised deskriptorid) koosneva mudelitega. Modifitseeritud pöörduvad mehhanismid vajavad modifitseerituse kirjeldamiseks ühte kuni kahte intensiivsuspameetrit (elektroonsed deskriptorid). Pöördumatute mehhanismide üldistamiseks formuleeriti toimeprofili kontseptsioon. Selle järgi avaldub homoloogilise rea madalamate liikmete toime elusorganismi vastu reaktiivse toksilisusena ja kõrgemate liikmete oma narkoosina. Graafikul mis kujutab saasteaine toksilisust funktsioonina tema hüdrofoobsusest esineb iseloomulik murdepunktiga joon. Mehhanismi seisukohast interpreteeritavate mudelite saamiseks tuleb andmekomplekt murdepunktile vastava hüdrofoobsuse väärtuse järgi kaheks jagada. Reaktiivse toksilisuse modelleerimine jääb ka peale selliseid ettevalmistusi probleemataks, sest intensiivsuspameetrid hakkavad õigesti tööle alles siis kui mitu homoloogilist andmekomplekti on kombineeritud üheks analoogiliseks andmekomplektiks.

Kvalitatiivse ja kvantitatiivse analüüsi kooskasutamine andis sünergilise efekti. Lahendada õnnestus mitmeid probleeme (näiteks füsikokeemiliste omaduste lävendi-

väärtuste määramine) millega üks ega teine üksinda toime ei tuleks. Töövoog omab tulemuste parandamiseks ja täpsustamiseks arvestatavat varu kui kasutada õnnestuks veelgi suuremaid andmekomplekte ning spetsiifilisemaid teoreetilisi molekulaarseid deskriptoreid. Lõpuks oleks huvitav uurida tema kohandatavust akuutse toksilisuse kõrval teistele bioloogilise aktiivsuse liikidele.

Töö tulemused on vormistatud kahe Lisas esitatud käsikirjana. Esimene käsikiri on esitatud avaldamiseks. Teine käsikiri on ettevalmistamisel.

6 SUMMARY

An integral workflow has been designed and implemented for the rapid assessment of *T. pyriformis* acute aquatic toxicity data. Known mechanisms of toxic action of aliphatic and aromatic compounds were organised into a classification tree with four terminal nodes. The first two nodes came to represent unmodified and modified reversible mechanisms in which the xenobiotic interacts with the biomatter by means of physical interactions such as van der Waals forces and hydrogen bonding. The other two nodes came to represent direct and indirect (i.e. involving metabolic activation) irreversible mechanisms in which the xenobiotic forms covalent bonds with biomatter.

The mapping of compounds to the classification tree was performed using the technique of self-organising maps. Alternative sets of theoretical molecular descriptors were identified that exhibited relevancy to the problem of discrimination between these terminal nodes. The level of detail of the map can be varied greatly by adjusting its configuration and adding/removing dedicated descriptors. The comparison of results for aliphatic and aromatic compounds demonstrated remarkable differences. The mechanism of action of aliphatic compounds depends on the properties of few dominate but discrete functional groups. Current descriptors are not particularly successful in descending to fragmental level, because they lack the required structural (meta)information. The mechanism of action of aromatic compounds depends on the electron density on the ring. With the careful planning of experiment it was possible to work out everything solely on the basis of the energies of frontier molecular orbitals. Many observations were new to the field.

A multilinear regression model was developed for an each group of compounds assigned to a particular mechanism of toxic action. The establishment of such direct correspondence is one of the key features of the workflow. When required to estimate the toxicity of an unknown compound, it only takes to calculate a few theoretical molecular descriptors to identify the mechanism of toxic action and to parameterise the model associated with it. After the eradication of need for human interaction, the accuracy of predictions can be further promoted by dissecting larger mechanisms into smaller and better understood mechanisms.

Unmodified reversible mechanisms demonstrated good to excellent results with one-parameter partitioning coefficient based models. Modified reversible mechanisms

needed one or two auxiliary intensive parameters to be added to it. Reversible mechanisms are the most difficult to model. The composition of models varies depending on the relative abundance of lower and higher homologues in the dataset. With the aid of the concept of response profiles, this phenomenon could be interpreted as the change over of the rate limiting step.

7 KASUTATUD KIRJANDUS

1. The Toxic Substances Control Act (TSCA). 1976. 15 U.S.C. s/s 2601 et seq. <http://www4.law.cornell.edu/uscode/15/ch53.html>
2. European Commission. 1996. Technical Guidance Document in support of commission directive 93/67/EEC on risk assessment of new notified substances and commission regulation (EC) No 1488/94 on risk assessment of existing substances - Part III, pp 505-570.
3. European Commission. 2001. White paper on the Strategy for a future Chemicals Policy. COM(2001) 88. http://europa.eu.int/comm/environment/chemicals/pdf/0188_en.pdf
4. McKim, J. M., Bradbury, S. P., Niemi, G. J. Fish Acute Toxicity Syndromes and Their Use in the QSAR Approach to Hazard Assessment. *Environ. Health Perspect.*, 1988, 71, 171-186.
5. Schultz, T. W. TETRATOX: *Tetrahymena pyriformis* population growth impairment endpoint: a surrogate for fish lethality. *Toxicol. Methods*, 1997, 7, 289-309.
6. Schultz, T. W., Sinks, G. D., Bearden, A. P. QSAR in Aquatic Toxicity: A Mechanism of Action Approach Comparing Toxic Potency to *Pimephales promelas*, *Tetrahymena pyriformis*, and *Vibrio fischeri*. In: *Comparative QSAR*, Devillers, J. (Ed.), Taylor & Francis, 1997, pp. 51-109.
7. Bearden, A. P., Schultz, T. W. Comparison of *Tetrahymena* and *Pimephales* toxicity based on mechanism of action. *SAR QSAR Environ. Res.*, 1998, 9, 127-153.
8. Seward, J. R., Hamblen, E. L., Schultz, T. W. Regression comparisons of *Tetrahymena pyriformis* and *Poecilia reticulata* toxicity. *Chemosphere*, 2002, 47, 93-101.
9. Dimitrov, S. D., Mekenyan, O. G., Sinks, G. D., Schultz, T. W. Global modeling of narcotic chemicals: ciliate and fish toxicity. *J. Mol. Struct.*, 2003, 622, 63-70.
10. Fukuto, T. R. Mechanism of action of organophosphorus and carbamate insecticides. *Environ. Health Perspect.*, 1990, 87, 245-255.
11. Terada, H. Uncouplers of Oxidative Phosphorylation. *Environ. Health Perspect.*, 1990, 87, 213-218.
12. Lipnick, R. L., Johnson, D. E., Gilford, J. H., Bickings, C. K., Newsome, L. D. Comparison of fish toxicity screening data for 55 alcohols with the quantitative structure-activity relationship predictions of minimum toxicity for nonreactive nonelectrolyte organic compounds. *Environ. Toxicol. Chem.*, 1985, 4, 281-296.
13. Lipnick, R. L. Outliers: their origin and use in the classification of molecular mechanisms of toxicity. *Sci. Total Environ.*, 1991, 109/110, 131-153.
14. Könemann, H. Quantitative structure-activity relationships in fish toxicity studies. Part 1: relationship for 50 industrial pollutants. *Toxicology*, 1981, 19, 209-221.
15. Schultz, T. W., Seward-Nagel, J., Foster, K. A., Tucker, V. A. Population Growth Impairment of Aliphatic Alcohols to *Tetrahymena*. *Environ. Toxicol.*, 2004, 19, 1-10.
16. Sinks, G. D., Carver, A. T., Schultz, T. W. Structure-toxicity relationships for aminoalkanols: a comparison with alkanols and alkanamines. *SAR QSAR Environ. Res.*, 1998, 9, 217-228.
17. Jaworska, J. S., Hunter, R. S., Schultz, T. W. Quantitative Structure-Toxicity Relationships and Volume Fraction Analyses for Selected Esters. *Arch. Environ.*

- Contam. Toxicol.*, 1995, 29, 86-93.
18. Veith, G. D., Lipnick, R. L., Russom, C. L. The toxicity of acetylenic alcohols to the fathead minnow, *Pimephales promelas*: narcosis and proelectrophile activation. *Xenobiotica*, 1989, 19, 5, 555-565.
 19. Schultz, T. W., Bryant, S. E., Lin, D. T. Structure-Toxicity Relationships for *Tetrahymena*: Aliphatic Aldehydes. *Bull. Environ. Contamin. Toxicol.*, 1994, 52, 279-285.
 20. Schultz, T. W., Cronin, M. T. D. Response-Surface Analyses for Toxicity to *Tetrahymena pyriformis*: Reactive Carbonyl-Containing Aliphatic Chemicals. *J. Chem. Inf. Comput. Sci.*, 1999, 39, 304-309.
 21. Schultz, T. W., DeWeese, A. D. Structure-Toxicity Relationships for Selected Lactones to *Tetrahymena pyriformis*. *Bull. Environ. Contam. Toxicol.*, 1999, 62, 463-468.
 22. Schultz, T. W., Netzeva, T. I., Roberts, D. W., Cronin, M. T. D. Structure-Toxicity Relationships for the Effects to *Tetrahymena pyriformis* of Aliphatic, Carbonyl-Containing, α,β -Unsaturated Chemicals. *Chem. Res. Toxicol.*, 2005, 18, 330-341.
 23. Akers, K. S., Sinks, G. D., Schultz, T. W. Structure-toxicity relationships for selected halogenated aliphatic compounds. *Environ. Toxicol. Pharmacol.*, 1999, 7, 33-39.
 24. DeWeese, A. D., Schultz, T. W. Structure-Activity Relationships for Aquatic Toxicity to *Tetrahymena*: Halogen-Substituted Aliphatic Esters. *Environ. Toxicol.*, 2001, 16, 54-60.
 25. Hermens, J. L. M. Electrophiles and Acute Toxicity to Fish. *Environ. Health Perspect.*, 1990, 87, 219-225.
 26. Sykes, P. A guidebook to mechanism in organic chemistry. 6th ed., Prentice Hall, 1986.
 27. Veith, G. D., Broderius, S. J. Rules for Distinguishing Toxicants That Cause Type I and Type II Narcosis Syndromes. *Environ. Health Perspect.*, 1990, 87, 207-211.
 28. Schultz, T. W., Lin, D. T., Arnold, L. M. QSARs for monosubstituted anilines eliciting the polar narcosis mechanism of action. *Sci. Total Environ.*, 1991, 109/110, 569-580.
 29. Schultz, T. W., Holcombe, G. W., Phipps, G. L. Relationships of quantitative structure-activity to comparative toxicity of selected phenols in the *Pimephales promelas* and *Tetrahymena pyriformis* test systems. *Ecotoxicol. Environ. Saf.*, 1986, 12, 146-153.
 30. Schultz, T. W., Lin, D. T., Wesley, S. K. QSARs for monosubstituted phenols and the polar narcosis mechanism of toxicity. *Qual. Assur. Good Pract. Regul. Law*, 1992, 1, 132-143.
 31. Veith, G. D., Broderius, S. J. Structure-toxicity relationships for industrial chemicals causing type (II) narcosis syndrome. In: *QSAR in Environmental Toxicology - II*, Kaiser, L. K. E (Ed.), D. Reidel Publishing Co., Dordrecht, Holland, 1987, pp. 385-391.
 32. Vaes, W. H. J., Urrestarazu Ramos, E., Verhaar, H. J. M., Hermens, J. L. M. Acute toxicity of nonpolar versus polar narcosis: is there a difference? *Environ. Toxicol. Chem.*, 1998, 17, 7, 1380-1384.
 33. Cronin, M. T. D., Schultz, T. W. Structure-toxicity relationships for phenols to *Tetrahymena pyriformis*. *Chemosphere*, 1996, 32, 8, 1453-1468.
 34. Mekenyan, O. G., Veith, G. D. A relationship between descriptors for

- hydrophobicity and soft electrophilicity in predictive toxicology. *SAR QSAR Environ. Res.*, 1993, 1, 335-344.
35. Cronin, M. T. D., Gregory, B. W., Schultz, T. W. Quantitative Structure-Activity Analyses of Nitrobenzene Toxicity to *Tetrahymena pyriformis*. *Chem. Res. Toxicol.*, 1998, 11, 902-908.
 36. Cronin, M. T. D., Manga, N., Seward, J. R., Sinks, G. D., Schultz, T. W., Parametrization of Electrophilicity for the Prediction of the Toxicity of Aromatic Compounds. *Chem. Res. Toxicol.*, 2001, 14, 1498-1505.
 37. Katritzky, A. R., Oliferenko, P., Oliferenko, A., Lomaka, A., Karelson, M. Nitrobenzene toxicity: QSAR correlations and mechanistic interpretations. *J. Phys. Org. Chem.*, 2003, 16, 811-817.
 38. Brunmark, A., Cadenas, E. Redox and addition chemistry of quinoid compounds and its biological implications. *Free Radical Biol. Med.*, 1989, 7, 435-477.
 39. Schultz, T. W., Sinks, G. D., Cronin, M. T. D. Quinone-induced toxicity to *Tetrahymena*: structure-activity relationships. *Aquat. Toxicol.*, 1997, 39, 267-278.
 40. a) Mohamadi, F., Richards, N., G. J., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., Still, W. C. MacroModel - an Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.*, 1990, 11, 440.
b) MacroModel version 7.0, Schrödinger, Inc., Portland (2000).
 41. a) Chang, G., Guida, W. C., Still, W. C. An Internal-Coordinate Monte Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.*, 1989, 111, 4379-4386.
b) Saunders, M., Houk, K. N., Wu, Y. D., Still, W. C., Lipton, M., Chang, G., Guida, W. C. Conformations of Cycloheptadecane. A Comparison of Methods for Conformational Searching. *J. Am. Chem. Soc.*, 1990, 112, 1419-1427.
 42. a) Halgren, T. A. Merck Molecular Force Field. I. – V. *J. Comput. Chem.*, 1996, 17, 490-519, 520-552, 553-586, 587-615, 616-641.
b) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries. *J. Comput. Chem.*, 1999, 20, 730-748.
 43. Stewart, J. J. P. *MOPAC 6.0 Program Package*. QCPE #455, (1989).
 44. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.*, 1985, 107, 3902-3909.
 45. CODESSA version 2.20, Florida (1994).
 46. Karelson, M. *Molecular Descriptors in QSAR/QSPR*. John Wiley & Sons, New York, 2001.
 47. CLogP, Daylight Chemical Information Systems, Inc., <http://www.daylight.com/daycgi/clogp>
 48. Verhaar, H. J. M., van Leeuwen, C. J., Hermens, J. L. M. Classifying environmental pollutants. Part 1: Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere*, 1992, 25, 4, 471-491.
 49. Jäckel, H., Nendza, M. Reactive substructures in the prediction of aquatic toxicity data. *Aquat. Toxicol.*, 1994, 29, 305-314.
 50. Schultz, T. W., Sinks, G. D., Cronin, M. T. D. Identification of Mechanisms of Toxic Action of Phenols to *Tetrahymena pyriformis* from Molecular Descriptors.

- In: *Quantitative Structure-Activity Relationships in Environmental Sciences - VII*. Chen, F. C., Schüürmann, G. (Eds.), SETAC Press, Pensacola, Florida, 1997, pp. 329-342.
51. Aptula, A. O., Netzeva, T. I., Valkova, I. V., Cronin, M. T. D., Schultz, T. W., Kühne, R., Schüürmann, G. Multivariate Discrimination between Modes of Toxic Action of Phenols. *Quant. Struct.-Act. Relat.*, 2002, 21, 12-22.
 52. Ren, S. Determining the Mechanism of Toxic Action of Phenols to *Tetrahymena pyriformis*. *Environ. Toxicol.*, 2002, 17, 119-127.
 53. Schüürmann, G., Aptula, A. O., Kühne, R., Ebert, R.-U. Stepwise Discrimination between Four Modes of Toxic Action of Phenols in the *Tetrahymena pyriformis* Assay. *Chem. Res. Toxicol.*, 2003, 16, 974-987.
 54. Ren, S. Two-step Multivariate Classification of the Mechanisms of Toxic Action of Phenols. *QSAR Comb. Sci.*, 2003, 22, 596-603.
 55. Cronin, M. T. D., Aptula, A. O., Duffy, J. C., Netzeva, T. I., Rowe, P. H., Valkova, I. V., Schultz, T. W. Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to *Tetrahymena pyriformis*. *Chemosphere*, 2002, 49, 1201-1221.
 56. Ren, S. Phenol mechanism of toxic action classification and prediction: a decision tree approach. *Toxicol. Lett.*, 2003, 144, 313-323.
 57. Spycher, S., Pellegrini, E., Gasteiger, J. Use of Structure Descriptors to Discriminate between Modes of Toxic Action of Phenols. *J. Chem. Inf. Model.*, 2005, 45, 200-208.
 58. Kohonen, T. *Self-Organising Maps*. 3rd ed., Springer, Berlin, 2001.
 59. Zupan, J., Gasteiger, J. *Neural Networks in Chemistry and Drug Design*. 2nd ed., Wiley-VCH, Weinheim, Germany, 1999.
 60. Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., Streich, M. The Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients. *J. Am. Chem. Soc.*, 1963, 85, 2817-2824.
 61. Hansch, C., Fujita, T. ρ , σ , π Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.*, 1964, 86, 1616-1626.
 62. Lipnick, R. L., Watson, K. R., Strausz, A. K. A QSAR study of the acute toxicity of some industrial organic chemicals to goldfish. Narcosis, electrophile and proelectrophile mechanisms. *Xenobiotica*, 1987, 17, 1011-1025.
 63. Deneer, J. W., Seinen, W., Hermens, J. L. M. The acute toxicity of aldehydes to the guppy. *Aquat. Toxicol.*, 1988, 12, 185-192.
 64. Deneer, J. W., Sinnige, T. L., Seinen, W., Hermens, J. L. M. A quantitative structure-activity relationship for the acute toxicity of some epoxy compounds to the guppy. *Aquat. Toxicol.*, 1988, 13, 195-204.
 65. Draper, N. R., Smith, H. *Applied Regression Analysis*. Wiley, New York, 1981.

8 LISAD

1. Ruusmann, V., Schultz, T. W., Maran, U. Qualitative and Quantitative Aspects of Acute Aquatic Toxicity to *Tetrahymena pyriformis*. *Chem. Res. Toxicol.*, 2006, submitted 12.04.2006.
2. Ruusmann, V., Maran, U. Mechanisms of Acute Aquatic Toxicity of Aromatic Compounds to *Tetrahymena pyriformis*. In preparation.