

Tartu Ülikool

Loodus- ja tehnoloogiateaduskond

Keemia instituut

Riinu Ansper

**Allergilist nahapõletikku tekitavate polaarse kaksiksidemega ainete uurimine  
kvantitatiivsete struktuur-aktiivsus sõltuvustega**

Bakalaureusetöö

Juhendaja Uko Maran, Ph. D

Tartu 2013

## Sisukord

|   |    |
|---|----|
| Sisukord .....  | 2  |
| Sissejuhatus.....   | 4  |
| 1. Kirjanduse ülevaade.....   | 5  |
| 1.1.1. Kontaktdermatiidi mehhanism.....   | 5  |
| 1.1.2. Keemiline mehhanism .....  | 6  |
| 1.1.3. Lokaalsete lümfisõlmede katse .....  | 8  |
| 1.2. Allergilist nahapõletikku tekitavate ühendite ennustamine <i>in silico</i> ..... | 9  |
| 1.2.1. Klassifikatsioonimudelid .....   | 9  |
| 1.2.2. Lokaalsed mudelid ehk mehhanismipõhine lähenemine .....                        | 10 |
| 2. Andmed ja meetodika .....  | 12 |
| 2.1. Andmevalim .....   | 12 |
| 2.2. Kvantkeemilised arvutused .....  | 12 |
| 2.3. Molekulaarsed deskriptorid.....  | 13 |
| 2.4. Kvantitatiivsete struktuuri-aktiivsuse sõltuvuste koostamine .....               | 13 |
| 2.4.1. Kasutatud multi-lineaarsed regressioonimeetodid .....                          | 14 |
| 2.4.2. Mudelite hindamine ja valideerimine .....                                      | 15 |
| 3. Tulemused.....   | 17 |
| 3.1. Elektrofiilse kaksiksideme mehhanismi järgi toimivad ained.....                  | 17 |
| 3.2. Ained, milles kaksiksidet mõjutab keto-rühm.....                                 | 19 |
| Kokkuvõte.....  | 25 |
| Summary in English.....   | 26 |
| 4. Kirjandus.....   | 27 |
| 5. Lisad.....   | 30 |

## Lühendid

|                                   |   |
|-----------------------------------|---|
| AKD                               | allergiline kontaktdermatiit, allergiline nahapõletik ( <i>Allergic contact dermatitis</i> )  |
| BMLR                              | parim multi-lineaarne regressioon ( <i>Best multilinear regression</i> )  |
| DfW                               | Derek Windows'i operatsioonisüsteemile ( <i>Derek for Windows</i> )   |
| EC3                               | konsentratsioon (massiprotsentides), mis tekitab kolmekordse stimulatsiooniindeksi ( <i>Estimated concentration (%) of a substance expected to produce a stimulation index of 3</i> ) |
| FHDSA                             | vesiniksideme doonori osapindala ( <i>Fractional hydrogen donor surface area</i> )  |
| <sup>HA</sup> HDCA <sup>(1)</sup> | vesiniksidemete aktseptorist sõltuv laetud vesiniksideme doonori osapindala ( <i>Hydrogen acceptor dependent Hydrogen donors charged surface area-1</i> )                             |
| LLNA                              | lokaalne lümfisõlme proov ( <i>Local lymph node assay</i> )   |
| logP                              | oktanool-vesi jaotuskoefitsiendi logaritm   |
| pH                                | negatiivne logaritm happe ionide aktiivsusest lahuses   |
| QSAR                              | kvantitatiivne struktuuri-aktiivsuse sõltuvus ( <i>Quantitative structure-activity relationship</i> )   |
| R <sup>2</sup>                    | determinatsioonikoefitsient   |
| R <sup>2</sup> <sub>cv</sub>      | ristvalideeritud determinatsioonikoefitsient  |
| RAI                               | suhteline alkülatiooniindeks ( <i>Relative alkylation index</i> )   |
| SI                                | stimulatsiooniindeks ( <i>Stimulation index</i> )   |
| TIMES                             | koe metabolismi simulaator ( <i>Tissue metabolism simulator</i> )   |
| TOPKAT                            | mürgisuse prognoosimine arvutipõhise tehnoloogiaga ( <i>Toxicity prediction computer-assisted technology</i> )  |

## Sissejuhatus

Dermatiit on naha pindmise osa põletik ja see on üks tavalisemaid ja sagedamini esinevaid nahahaigusi. Kontaktdermatiidi korral tekib põletik pärast otsest kontakti keemilise ainega. Kontaktdermatiidi üks erijuhte on allergiline kontaktdermatiit. Selle puhul tekib põletikuline reaktsioon mõne aja möödudes pärast kontakti ärritava ainega ja tihti alles pärast korduvat kokkupuudet. Põletiku teket põhjustab ärriti esilekutsutud allergiline reaktsioon. Kui kemikaali vastu on tekkinud allergia, siis piisab dermatiidi tekkimiseks ka väikestest aine kogustest. [1]

Kontaktdermatiit esineb tööstusriikides sagedase kutsehaigusena. Allergilise kontaktdermatiidi korral on tekkinud ülitundlikkus ärriti vastu eluaegne ja sellest lähtuvad nahaprobleemid raskelt välditavad. Seega on tähtis haigust ennetada ja hinnata uute kemikaalide ohtu kontaktallegiat tekitada. Uut kemikaali saab usaldusväärselt hinnata loomkatsete abil. Samas *in vivo* katsed on kallid ja aeganõudvad ning seetõttu otsitakse teisi valikuvõimalusi *in vitro* ja *in silico* meetodites.

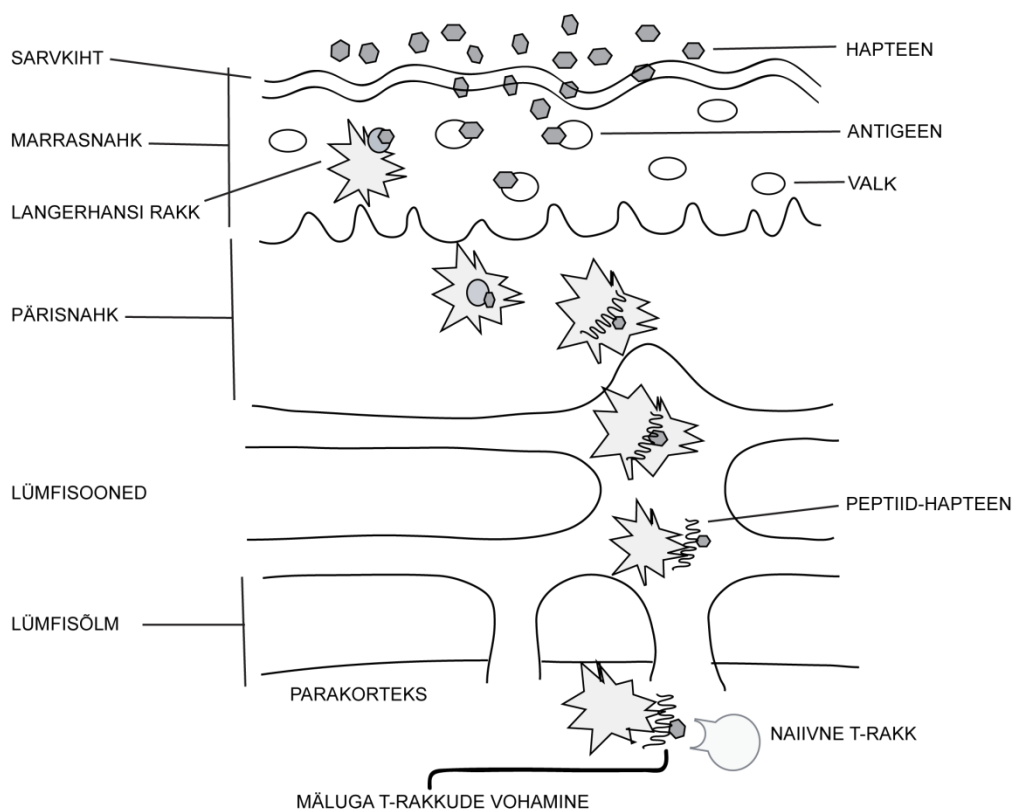
*In vitro* meetodeid arendatakse, kuid praeguse seisuga reguleerimiseks eesmärkideks sobiva täpsusega *in vitro* meetodeid ei leidu. *In silico* meetodeid peetakse paljulubavateks, sest arvuti abil prognoosimine säästaks aega ja raha ning aitaks loomkatsete arvu vähendada. *In silico* meetoditest on kvantitatiivsed struktuuri-aktiivsuse sõltuvused kõige efektiivsemad, sest need võivad olla rakendatavad paljudele kemikaalidele ja seejuures omada ka piisavat täpsust.

Töö kirjanduse ülevaate osas leidub kokkuvõtte kontaktdermatiidi tekkemehhanismist, senini kasutatud loomkatsetest ja tehtud *in silico* mudelitest. Praktilises osas uuriti elektrofiilse kaksiksidemetega ühendeid ja nende võimet allergilist kontaktdermatiiti tekitada. Selleks kasutati keemilise struktuuri ja naha tundlikkuse vahelisi kvantitatiivseid struktuuri-aktiivsuse sõltuvusi (*in silico* mudelid). Eraldi eesmärgiks oli *in silico* mudelites reaktsioonivõime iseloomustamiseks rakendada kvantkeemilisi molekulaardeskriptoreid.

# 1. Kirjanduse ülevaade

## 1.1.1. Kontaktdermatiidi mehhanism

Allergiline kontaktdermatiit kulgeb kahes faasis, millest esimene on induktsioon ja teine esilekutsumine. Induktsiooniprotsessis läbib kemikaal naha sarvkihi (*stratum corneum*) ja jõuab marrasnahka (*epidermis*). Praeguste teadmiste kohaselt põhjustab immunoloogilise reaktsiooni kemikaaliga reageerinud valk, mitte kemikaal ise [2]. Kehavõõrast madalmolekulaarset ühendit, mis reageerib valguga ja moodustab antigeeni, nimetatakse hapteeniks. Pärast reaktsiooni liigub hapteeni-valgu ühend Langerhansi rakkudesse (LR). Need küpsevad ja muutuvad seeläbi dendriitrakkudeks. Küpsemise käigus liiguvad LR lümfisõlmedesse ja lagundavad LR sisenenud kompleksi peptiidideks, millega on kovalentselt ühendatud hapteen. Lagundamise tulemusel tekkinud peptiidiahelad koos hapteeniga liiguvad raku pinnale. Lümfisõlmes esitletakse peptiid-hapteen ühendit (antigeeni) naiivsetele T-rakkudele. See algatab mälu T-rakkude (T-rakud, millel on antigeenile spetsiifiline retseptor) loome ja vohamise. Mälu T-rakud tunnevad edaspidise kokkupuute korral antigeeni ära ja algatavad allergilise reaktsiooni (Joonis 1). [3]



Joonis 1 Kontaktdermatiidi tekkemehhanism.

Kui aine tekitab põletikku vahetult pärast kokkupuudet, aga pikemaajalist allergiat ei teki, siis on tegu lihtsa kontaktdermatiidiga. Kuigi ka paljud allergilist kontaktdermatiiti tekitavad ained ärritavad nahka koheselt, ei pruugi enne nahaallergia väljakujunemist ohumärke avalduda. Allergilise nahapõletiku tekkeks on vajalik hapteni pikaajaline kokkupuude nahaga, misjärel tekib spetsiifiline antigeen ja seda ära tundvad mäluga T-rakud. [4]

### **1.1.2. Keemiline mehhanism**

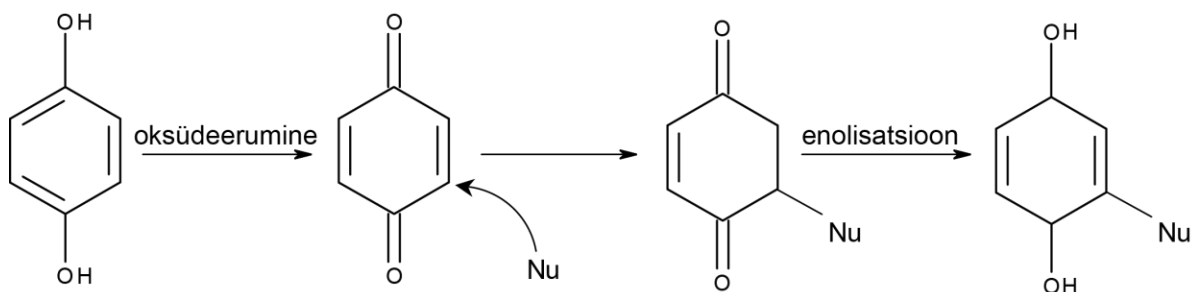
Allergilise nahapõletiku tekkimisel on olulised mitmesugused biokeemilised faktorid, nagu kemikaali reageerimisvõime valguga, kemikaali metabolism organismis, absorptsioon nahka ja seal olemise aeg. [4]

Esimeses etapis läbib hapteen naha sarvkihi. Paljud nahale kantud keskmise hüdrofoobsusega madalmolekulaarsed ühendid (molaarmass reeglina alla 500 g/mol) on võimelised sarvkihti läbima. Seega ei ole kemikaali võimet naha sarvkihti läbida seni peetud haiguse tekkimise juures määravaks faktoriks. [4]

Tunnustatud biokeemilise mudeli järgi reageerib hapteen marrasnahas sisalduvas valgus leiduva nukleofiiliga. Ei teata, kas hapteen seostub spetsiaalse valguga kindlas kohas nahas või saavad antigeeni tekitada erinevad valgud. Üldiselt vaadeldakse nukleofiilina lüsiini, arginiini aminorühmi või tsüsteiini tiolrühma. Kusjuures aminorühmi sisaldavatest aminohapetest arvatakse tähtsamaks lüsiini.

Naha eri osades on erinev pH, välimisest kihist vereringeni muutub pH 5,5st kuni 7,4ni. Lüsiini jäägid on terves selles pH-vahemikus protoneerunud. Arvatakse, et reaktsioon lüsiiniga saab toimuda tänu valgusisesele mikrokeskkonnale, kus mõni lüsiini jääk võib olla protoneerumata ja reaktsioonivõimeline. Tsüsteiiniga saab elektrofiil reageerida ka mujal (kui see ei ole seotud disulfidsilda). [4]

Erinevate hapteneide võime valguga reageerida sõltub nende elektrofiilsusest e võimest reageerida organismis olevate nukleofiilsete reaktsioonitsentritega. AKD tekitav aine, mis ei ole esialgsel kujul elektrofiil, võib saada tugevaks elektrofiiliks metaboolsete või abiootiliste muundumiste käigus; sellisel juhul on tegu prohapteneiga. Metaboolse muundumise tulemusena aine tihti oksüdeerub. Näiteks sobib dihidrokinoon, mis oksüdeerub organismis tugevaks elektrofiiliks bensokinooniks (Skeem 1). Metabolismi kaudu võib hapteen muutuda ka vähem reaktsioonivõimeliseks. [5]

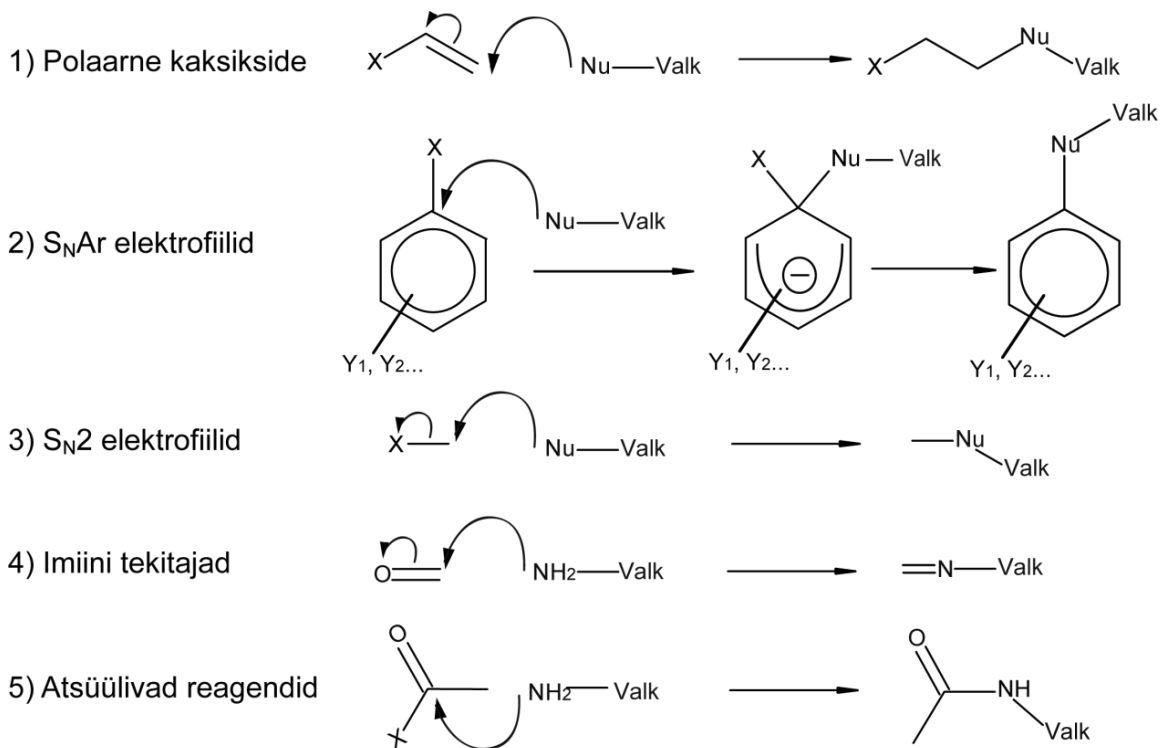


Skeem 1. Hüdrokinooni oksüdeerumine bensokinooniks.

Arvatakse, et reaktsioon hapteeni ja valgu vahel toimub hüdrofoobses keskkonnas (nagu rakumembraan). Selle põhjal on potentsiaalselt ohtlikumad hüdrofoobsed kemikaalid, sest need püsivad reaktsioonipiirkonnas kauem. [6]

Tõenäolise reaktsioonimehhanismi järgi jaotatakse AKDi tekitavad kemikaalid gruppidesse, (Skeem 2): [5]

- Polaarne kaksiksidi: Siia gruppi kuuluvad ühendid, millel on või millel saab metabolismi vmt kaudu tekkida polaarne elektrofiilne kaksiksidi või kolmikside, millele saab toimuda nukleofiilne liitumine.
- $S_NAr$  elektrofiilid: Sellised ühendid sisaldavad benseeni tuuma, millega on seotud mitu elektronegatiivset rühma ja hea lahkuv rühm (halogeen, pseudohalogeen).
- $S_N2$  elektrofiilid: Sisaldavad halogeeni või mõnda muud head lahkuvat rühma (sulfaadid, sulfoestrid), mis on seotud primaarse alküülse, aromaatses või  $sp^2$ -süsinikuga.
- Imiini tekitajad: Molekulid, mis sisaldavad reaktsioonivõimelist aldehyüdrühma. Samuti kuuluvad siia  $\alpha$ -,  $\beta$ - ja  $\gamma$ -diketoonid ja  $\alpha$ -ketoestrid. Mõned teised hetero-küllastumata süsteemid võivad käituda sarnaselt, nagu tiokarbonüülühendid, tiotsüanaadid ja isotiotsüanaadid.
- Atsüülivad reagentid: Karboksüülhappe derivaadid, kus asendaja on piisavalt happeline, et olla hea lahkuv rühm. Siia kuuluvad näiteks happe anhüdriidid.



Skeem 2. Võimalikud reaktsioonimehhanismid.

### 1.1.3. Lokaalsete lümfisõlmede katse

Kemikaali potentsiaali nahaga kokkupuutudes allergiat tekitada saab hinnata erinevate loomkatsete abil. Nendest kõige uuem ja efektiivsem on lokaalne lümfisõlme katse. Selle puhul puutub katseloom (hiir) kemikaaliga kokku ühe perioodi vältel ja tekitatakse ainult induktsioonifaas. Võimalike haptteenide olemasolu tuvastatakse kokkupuutele järgneva rakkude vohamise järgi lümfisõlmedes. Rakkude vohamist mõõdetakse tritiummürgistatud tümidiini meetodil [7]. Kui kontrollrühma puhul on tritiummürgistatud tümidiini lümfirakkudes kolm korda vähem kui kemikaaliga kokku puutunud hiirte lümfirakkudes, siis saadakse stimulatsiooniindeksiks kolm (SI = 3). Kemikaali potentsiaali allergilise reaktsiooni tekitada uuritakse mitme kontsentratsiooniga ja seeläbi saadakse kontsentratsiooni ja SI vaheline sõltuvus. Tulemusi väljendatakse ärriti kontsentratsioonina (massiprotsentides), mis tekitab SI = 3, seda väärtust nimetatakse EC3 (*estimated concentration (%) of a substance expected to produce a stimulation index of 3*). Niisiis võimaldab LLNA kemikaale EC3 järgi kategooriatesse jagada. Kemikaalide kategooriad EC3 järgi [8]:

|           |                                   |
|-----------|-----------------------------------|
| <1%       | tugev AKD tekitaja                |
| 1 – 10%   | keskmine                          |
| 10 – 35%  | nõrk                              |
| 35 – 100% | väga nõrk                         |
| > 100%    | ei tekita allergilist reaktsiooni |

EC3(%) määramatuseks loetakse  $\pm 2$  ühikut, ehk molaarse ja logaritmilise versiooni pEC3 (vt 2.1. Andmevalim) korral on see umbes  $\pm 0,3$  ühikut [9].

## 1.2. Allergilist nahapõletikku tekitavate ühendite ennustamine *in silico*

Allergilise nahapõletiku tekitamise võime ennustamiseks *in silico* on tehtud palju erinevaid mudeleid. Struktuuri-aktiivsuse sõltuvusmudelitest on enim tuletatud lokaalseid mudeleid. Lokaalseid mudeleid iseloomustab see, et neid saab kasutada vaid reaktsioonimehhanismilt või keemiliselt kuuluvuselt sarnaste ühendite puhul. Veel on tehtud klassifikatsioonimudeleid, mis on enamasti tuletatud statistiliste meetoditega kogu andmeseeria analüüsimisel, arvestamata haiguse mehhanismi mudeli tuletamisel. Sellistel mudelitel ideaaljuhul kehtivuspiire ei ole. Leidub ka allergilise nahapõletiku tekitamispotentsiaali hindavaid arvutiprogramme ehk n-ö ekspertsüsteeme. Need programmid otsivad üldjuhul uuritavast struktuurist riskantseid alamstruktuure, mis on eelnevalt ekspertide poolt kindlaks määratud. Sellisteks programmideks on TOPKAT, MultiCASE, DfW. Leidub ka programme, kus lisaks eelmainitule võetakse arvesse ka metaboolseid, abiootilisi muutusi (TIMES). [10]

### 1.2.1. Klassifikatsioonimudelid

Klassifikatsioonimudelite loomiseks on kasutatud erinevaid statistilisi meetodeid ja arendatud erinevaid mudeli koostamise lähenemisi. Mudelite saamiseks on kasutatud erinevaid deskriptorikomplekte [11, 12], statistilisi optimeerimismeetodeid ja statistiliste meetodite kombinatsioone [13, 14].

Molekuli AKD tekitamise potentsiaali hindamiseks tehtud klassifikatsioonimudelid sisaldavad tihti sarnaseid omadusi kirjeldavaid deskriptoreid. Peaaegu alati on esindatud deskriptorid, mis iseloomustavad molekuli suurust ning kuju, hüdrofoobsust, elektronjaotust või reaktsioonivõimelisust [13-15].

Enamik tehtud klassifikatsioonimudeleid jagab molekulid ohtlikkuse põhjal rühmadesse (tüüpiliselt kahte rühma, mõnikord kolme). Ennustamisvõimet iseloomustavate statistiliste

parameetrite poolest ei ole klassifikatsioonimudelid üldiselt lokaalsetest halvemad (kirjanduses olevad mudelid klassifitseerivad oma kontrollvalimi molekulidest 70% kuni 90% korrektselt). Regulaatorsetel eesmärkidel kasutatakse klassifikatsiooni mudeleid praegu siiski vähem. Põhjuseks on ilmselt see, et paljud neist ei ole ühteviisi rakendatavad erinevate molekulide korral ning nende viga on raske hinnata. Üldiselt võib öelda, et allergilist kontaktdermatiiti prognoosivad klassifikatsioonimudelid on praegu arengujärgus, kus ei ole jõutud universaalsete lahendusteni. [10]

### ***1.2.2. Lokaalsed mudelid ehk mehhanismipõhine lähenemine***

Mehhanismipõhiseid mudeleid on koostatud struktuuri või mehhanismi poolest sarnaste ühendite rühmadele. Nende kasutuspiirkond on defineeritud ja sellega seonduvalt on viga võimalik hinnata. Lokaalsetel mudelite puhul on probleemiks sobilike eksperimentaalselt mõõdetud ühendite vähesus, mille tõttu ei ole sellised mudelid tihti statistiliselt robustsed.

Mehhanistlike mudelite hulka kuuluvast RAI-mudelisse (*Relative Alkylation Index*) eeldatakse, et kemikaali võime allergilist nahapõletikku tekitada sõltub sellest, kui hästi see seostub valguga. Valguga seostumisel loetakse määravateks parameetriteks aine reaktsioonivõimelisust ja hüdrofoobsust. Reaktsioonivõimelisust iseloomustavateks suurusteks on erinevates mudelites kasutatud standardreaktsiooni kiiruskonstante, molekulaarorbitaalide parameetreid ja Tafti või Hammeti asendajakonstante. Hüdrofoobsuse kirjeldamiseks kasutatakse enamasti logP-d. Jaotuskoefitsient võeti originaalmudelisse kasutusele selleks, et iseloomustada, kui kaua püsib hapteen reaktsiooniosibilikus keskkonnas (hüdrofoobne, vt 1.1.1. Keemiline mehhanism). Oktanooli-vee jaotuskoefitsienti on mõningal määral seostatud ka marrasnaha läbimisega, kuid üldiselt eeldatakse taoliste mudelite koostamisel, et kemikaal (kui on suhteliselt väikese molaarmassiga) jõuab alati marrasnahka.

$$pEC3 = a * \log k + b * \log P + c \quad (1)$$

Võrrand (1) esitab RAI-mudeli üldkuju, kus  $p$  tähistab negatiivset logaritmi,  $EC3$  on keemilise ärritaja kontsentratsioon, mille puhul lümfisõlmedes vahavad rakud kolmekordselt;  $k$  on reaktsioonivõimet iseloomustav parameeter (siinkohal standardreaktsiooni kiiruskonstant);  $\log P$  – oktanooli-vee jaotuskoefitsiendi logaritmi;  $a, b, c$  – kordajad.

RAI-mudeleid on tehtud erinevatele keemiliste ühendite rühmadele: sultoonid [16], urusooli analoogid [17], aldehüüdid ja diketoonid [18, 19, 20], akrülaadid [21]. Samuti on mõningase

eduga rakendatud RAI-mudeleid ka mehhanismi järgi jaotatud nahaärritajate rühmade jaoks (vt 1.1.1. Keemiline mehhanism). [6] Rühmad, mida on RAI-mudeliga modelleeritud on  $S_NAr$  elektrofiilid [ 22 ], imiini tekitajad: [9],  $S_N2$  elektrofiilid [ 23 ]. Samas on elektrofiilse kaksiksidemega ühendite kirjeldamine struktuuri omaduste (sh RAI-mudeli) põhjal on osutunud keeruliseks.

Nende keemilise mehhanismi alusel koostatud rühmade jaoks, millel puudub toimiv lokaalne mudel, üritatakse rakendada analoogmeetodit. Analoogmeetodi kohaselt on sama keemilise mehhanismi järgi toimiva ühendi ohtlikkus (väljendatuna nt EC3 kaudu) sarnane siis, kui need ühendid on sarnased ohtlikkust määravate parameetrite (nt struktuur, *in chemico* reaktsioonivõime) poolest. Selle rakendamiseks on vaja ulatuslikult võrdlusandmeid. Nii luuakse näiteks polaarse kaksiksidemega ühendite rühma jaoks keemilist reaktiivsust iseloomustavat andmebaasi, mille jaoks mõõdetakse reaktsioonide eksperimentaalseid kiiruskonstante (mudelnukleofiiliks glutatioon). [24]

Käesoleva töö eesmärk on polaarseid kaksiksidet omavate ühendite rühmale erinevaid statistilisi meetodeid kasutades tuletada QSAR ja kasutada selles reaktsioonivõime kirjeldamiseks kvantkeemilisi molekulaar-deskriptoreid (täpsemalt reaktsioonivõime indekseid).

Polaarse kaksiksidemega ühendite reaktsioonivõime iseloomustamiseks on varem üritatud kasutada kiiruskonstante (2009 aasta Roberts jt) [ 25 ]. Selles töös tuletati mudel kümnele ühendile ja  $R^2$  saadi 0,84. Hiljuti, 2013. aastal tegid Enoch jt katse kasutada polaarse kaksiksideme reaktiivsuse modelleerimiseks kvantkeemilisi arvutusi. Selleks arvutati tihedusfunktsionaali teooriaga energiaerinevused (nn aktivatsioonenergia) stabiilse vaheoleku ja lähteühendite summaarse energia vahel, kui mudelnukleofiiliks võeti metiool. Täiendava deskriptorina kasutati selles võrrandis reaktsioonitsentrite arvutatud solvendile kättesaadavat pindala. Suhteliselt piiratud andmevalimi põhjal (25 ühendit) saadi mudel, mille  $R^2 = 0,79$ . [26]

## 2. Andmed ja meetodika

### 2.1. Andmevalim

Andmevalimi koostamisel kasutati Gerberick jt 2005. [27] ja Kern jt 2010. [28] aastal avaldatud eksperimetaalseid EC3 väärtuseid omavate ühendite seeriaid. 2005. aasta andmeseeria jaotati reaktsioonimehhanismide järgi gruppidesse Roberts jt poolt 2007. aastal [23]. 2010. aasta andmeseerias olid kemikaalid juba arvatava reaktsioonimehhanismi järgi jaotatud. Kaks ühendit leiti lisaks veel Enoch 2013 aasta tööst [26].

Eelpool toodud artiklitest leiti 93 polaarse (elektrofiilse) kaksiksideme mehhanismi järgi toimivat ainet. Nendest 84 struktuuril oli EC3 väärtus numbriliselt olemas (info Tabel L1, struktuurid Tabel L2, toodud Lisades). Ülejäänud üheksale ühendile ei saadud EC3 väärtust eksperimendist tulenevalt arvutada ja seega ei olnud EC3 numbrilist väärtust artiklis toodud.

Eksperimentaalsed EC3 (%) andmed teisendati edasiste arvutuste jaoks molaarsele logaritmilisele  $pEC3$  skaalale (2), kus  $M$  on ühendi molaarmass, EC3 (%) viitab kontsentratsioonile (massiprotsentides), mille korral tekitab ühend  $SI = 3$ . Seega mida väiksem on EC3 (%) (ja mida suurem on  $pEC3$ ), seda suurem on selle ühendi võime tekitada AKD.

$$pEC3 = \log \frac{M}{EC3 (\%)} \quad (2)$$

Esialgne, 84 ühendiga andmevalim sisaldas mitme reaktsioonitsentriga ühendeid ja prohapteene. Mehhanismi järgi tõlgendatavate tulemuste saamiseks eraldati kogu andmevalimist piiratud andmevalim, millest jäeti välja prohapteened ja mitme reaktsioonitsentriga ühendid. Samuti jäeti piiratud andmevalimist välja need ühendid, mille korral mõjutas kaksiksidet mõni muu rühm kui keto-rühm. Piiratud andmevalimisse jäi 31 ühendit, millel on üks reaktsioonitsenter ja mille puhul on kaksiksidet mõjutavaks elektrofiilseks rühmaks keto-rühm.

### 2.2. Kvantkeemilised arvutused

Esialgse andmekomplekti 84 ühendi struktuurid genereeriti CAS-numbri järgi *ChemAxoni* integreeritud programmikomplektiga *JChem* [29] ja need struktuurid kontrolliti CAS-numbri (selle puudumisel ühendi nime) järgi üle *Scifinder*'i otsingumootoris. [30]. Kui algne ühend oli ioonsel kujul, siis see vastavalt protoneeriti või deprotoneeriti ja tehti neutraalseks. Edasised arvutused teostati laenguta ühenditega. Geomeetria optimeeriti poolempiirilisel MOPAC 6.0 programmpaketiga [31] kasutades AM1 paramerisatsiooni. Optimeerimiseks kasutati käsuriida:

AM1 PRECISE T=80000 NOINTER PRECISE GNORM=0.01 EF

Saadud optimeeritud geometriale arvutati molekulide edasiseks iseloomustamiseks veel elektrostaatilisi, steerilisi ja energeetilisi parameetreid. Selleks kasutati käsuriida:

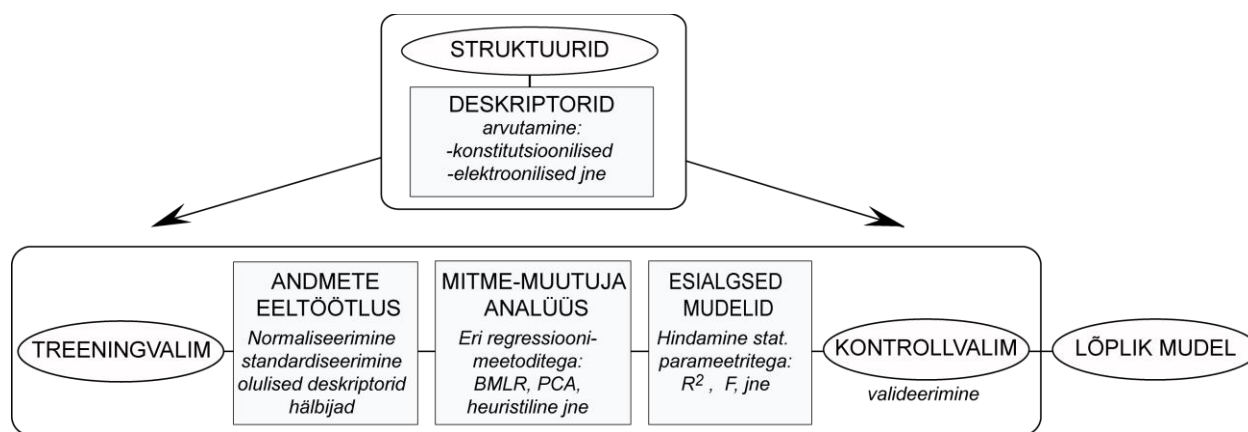
AM1 VECTORS BONDS PI POLAR PRECISE ENPART 1SCF

### 2.3. Molekulaarsed deskriptorid

Pärast poolempiirlisi arvutusi viidi ühendite optimeeritud geometriad ja kvantkeemiliste parameetrite väärtused üle programmi CODESSA, millega arvutati või ekstraheeriti topoloogilised, konstitutsioonilised, geomeetrilised, elektrostaatilised ja kvantkeemilised deskriptorid. Multilineaarsete regressioonide tuletamise meetodid (BMLR, heuristiline) kõrvaldavad statistiliselt ebaolulised deskriptorid automaatselt, samas võivad mehhanismi seisukohast ebaolulised deskriptorid mudelistesse sisse tulla. Seepärast kõiki CODESSA poolt arvutatud või ekstraheeritud deskriptoreid mudelite koostamisel ei kasutatud. Näiteks ei võetud arvesse algseid asukohaspetsiifilisi deskriptoreid. Sobivate deskriptorite algkomplekt kujunes mudelite tegemise käigus.

### 2.4. Kvantitatiivsete struktuuri-aktiivsuse sõltuvuste koostamine

QSAR on *in silico* mudel, mis seob omavahel keemilise struktuuri omaduse (see võib olla keemiline, füüsikaline, bioloogiline jne) ja seda struktuuri numbriliselt kirjeldatavad parameetrid. Üldine töövoog QSARi tuletamiseks sisaldab palju erinevaid etappe (Joonis 2).



Joonis 2. Etapid QSARi saamiseks.

Töövoo algul jaotatakse ühendid sageli kontrolli- ja treeninguvalimiks nii, et mõlemas alagrupis oleks esindatud andmevalimi kõik põhilised struktuuritüübid (seesugust jaotamist ei tehta kui mudelit väliselt ei valideerita). Seejärel arvutatakse deskriptorid. Andmete eeltöötleses deskriptorid normaliseeritakse ja tsentreeritakse; veel valitakse välja olulised deskriptorid ja jäetakse välja hälbivad ühendid. Valitud deskriptoritekomplektiga teostatakse mitme-muutuja analüüs parima mudeli saamiseks. [32]

Mitme-muutuja-analüüsi tulemusena saab teada kvantitatiivsed seosed sõltumatute muutujate (deskriptorid) ja sõltuva muutuja (omadus) vahel. Kui analüüs teostatakse lineaarsete regressioonimeetoditega (multi-lineaarsed regressioonimeetodid), siis saadakse tulemus üldkujul:

$$y = a_0 + a_1 * x_1 + \dots + a_n * x_n \quad (3)$$

kus  $y$  on omadus (sõltuv muutuja);  $x_1, x_2, x_n$  on deskriptorid (sõltumatud muutujad);  $a_1, a_2, a_n$  on kordajad. Pärast mitme-muutuja-analüüsi saadakse esialgsed mudelid, millest valitakse statistiliste parameetrite järgi välja parimad. [32]

Selles töös uuriti polaarse kaksiksidemega ühendite andmeseeriat. Täpsemalt seda, kas nende ühendite reaktsioonivõimet saab kirjeldada kvantkeemiliste deskriptoritega ja millised struktuuri omadused määravad nende ühendite võime käituda allergilise nahapõletiku tekitajana. Selleks koostati mitu struktuuri-aktiivsuse sõltuvust. Sobivate deskriptorite ja mudelite leidmiseks kasutati multi-lineaarseid regressioonimeetodeid: parimat multilineaarset regressiooni (BMLR) ja heuristilist regressiooni.

Kui mudelit on kavas kasutada ennustamiseks, siis on tähtis ka valideerimine. Mudelit saab valideerida nii mudeli siseselt kui ka väliselt.

#### **2.4.1. Kasutatud multi-lineaarsed regressioonimeetodid**

BMLR meetodi korral esialgu normaliseeritakse ja tsentreeritakse deskriptorite skaalad. Seejärel kõrvaldatakse omadusega mittekorreleeruvad ( $R^2 < 0,1$ ) ja puuduvate väärtustega deskriptorid. Pärast seda leitakse parimad kahe, kolme jne parameetrilised regressioonid. Nende regressioonide tegemisel võetakse arvesse deskriptorite statistilist olulisust ja mittekollineaarsuse tingimust. Viimane tähendab, et deskriptorite omavaheline korrelatsioon ei tohiks olla suurem kui  $R^2 < 0,6$ . Pärast parimate regressioonide tuletamist esitatakse deskriptorite skaalad taas algsel kujul. [33]

Ka heuristilise analüüsimeetodi korral deskriptorid esialgu normaliseeritakse ja tsentraliseeritakse. Seejärel eemaldatakse need deskriptorid, millel ei ole kõigi ühendite jaoks väärtust ning ka need, mille väärtus ühendite seas ei varieeru. Pärast seda koostatakse üheparameetrilised sõltuvused. Nende sõltuvuste statistiliste parameetrite (F-test,  $R^2$ , t-parameetri väärtus) järgi elimineeritakse järjekordne osa deskriptoreid. Seejärel koostatakse kaheparameetrilised sõltuvused, mille puhul võetakse arvesse deskriptorite mittekollineaarsuse tingimust (omavaheline korrelatsioon  $R^2 < 0,6$ ). Seejärel valitakse parimad kahe deskriptoriga võrrandid (vaikimisi valitakse kolm parimat) järgmise etapi teostamiseks. Järgnevalt täiendatakse parimaid paarseid korrelatsioone ükshaaval deskriptoriga. Uus deskriptor võetakse mudelisse, kui selle lisamisel saadud mudeli F-testi väärtus on suurem kui eelnenud mudeli oma. Viimaks esitatakse deskriptorite skaalad taas algsel kujul ja mudelid väljastakse. Väljastatakse 10 suurima  $R^2$  ja 10 suurima F-testi väärtusega mudelit, mille hulgast saab kasutaja teha valiku. [33]

#### 2.4.2. Mudelite hindamine ja valideerimine

Tulemuseks saadud mudelite kvaliteeti hinnatakse mitmete statistiliste parameetritega. Enimkasutatav statistiline parameeter *in silico* mudelite hindamiseks on determinatsioonikoefitsient ehk  $R^2$  (korrelatsioonikoefitsiendi ruut). See näitab, kui suure osa uuritava tunnuse varieeruvusest mudel ära kirjeldab.  $R^2$  on matemaatiliselt mudelile vastava hajuvuskomponendi ja uuritava tunnuse koguhajuvust kirjeldava hälvete ruutude summa jagatis (4). Mida lähemal on  $R^2$  väärtus ühele, seda täpsemalt prognoosib mudel tunnuse väärtust.

$$R^2 = \frac{\sum_{i=1}^n (P_{i(prog)} - \bar{P}_{(prog)})^2}{\sum_{i=1}^n (P_{i(eksp)} - \bar{P}_{(eksp)})^2} \quad (4)$$

Veel kasutatakse mudelite iseloomustamiseks F-testi ja standardhälbe ruutu ( $s^2$ ). F-test kontrollib, kas uuritava tunnuse parameetrite vahel on sõltuvus (nullhüpoteesiks on, et mudel ei kirjelda tunnust ja seega see ei sobitu andmestikuga). Kui F-testi väärtus jääb alla kriitilise väärtuse (enamasti  $\sim 1$ ), siis võib järeldada, et mudel ei kirjelda andmevalimit. Standardhälbe ruut ehk dispersioon kirjeldab kuidas on andmeseeria väärtused keskmise ümber jaotuvad; mida suurem on  $s^2$ , seda hajusamalt on andmed jaotunud. [34]

Saadud mudelite ennustusvõime tõestamiseks ja hindamiseks on vaja neid valideerida. Valideerida saab nii mudeliväliselt kui ka siseselt. Välise valideerimise korral jagatakse algne andmevalim kaheks osaks, millest ühte kasutatakse mudeli koostamiseks (treeningvalim) ja teist

valideerimiseks (kontrollvalim). Parimat treeningvalimiga tuletatud mudelit valideeritakse kontrollvalimiga. [32]

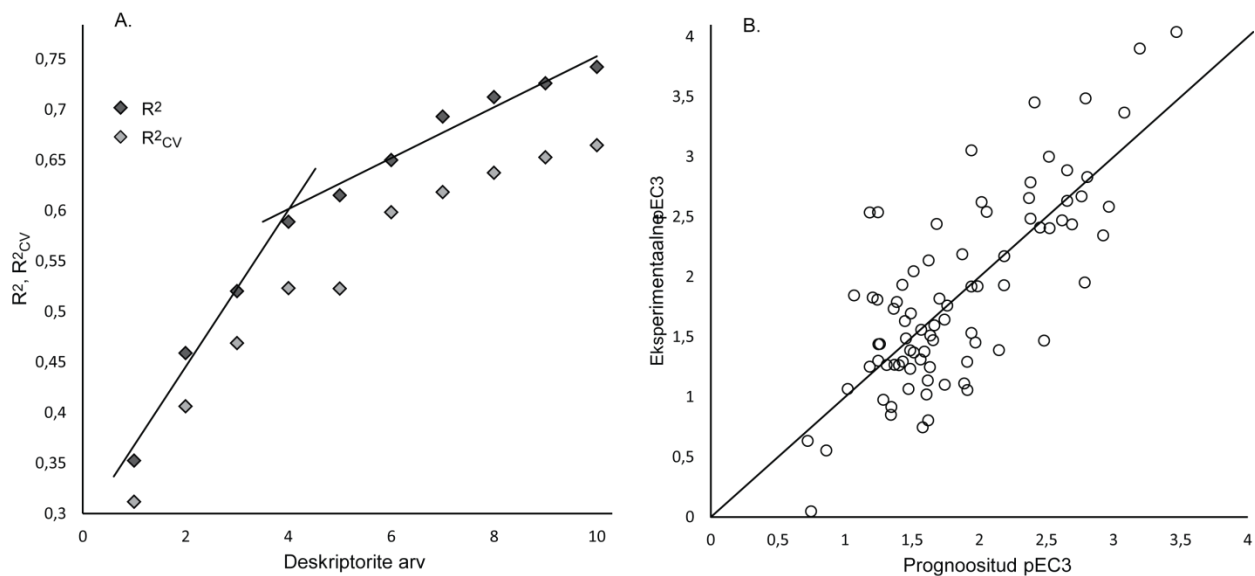
Sisemise valideerimise korral kasutatakse valideerimiseks samu ühendeid, millega mudel koostati. Käesolevas töös kasutati sisemist valideerimist. Täpsemalt rist-valideeritud determinatsioonikoefitsienti ( $R^2_{cv}$ ). Rist-valideeritud determinatsioonikoefitsiendi arvutamisel on keskse tähtsusega suurus nimega prognooside ruutude summa (*prediction of sum of squares*, PRESS). Mida väiksem on on PRESS väärtus, seda suurem on mudeli ennustusvõime. See suurus saadakse lahutades eksperimentaalsest tulemusest ennustatud tulemuse nii, et ennustamiseks kasutatud mudeli koostamisel ei kasutatud seda eksperimentaalset tulemust. Nii saadud vahede ruutude summa annab kokku PRESS väärtuse (5). Valemis on  $Y_i$  andmepunktile  $i$  vastav eksperimentaalne väärtus;  $\hat{Y}_{prog_{i(i)}}$  on andmepunkti  $i$  prognoositud väärtus kui andmepunkt  $i$  ise ei kuulunud treeningkomplekti. [35]

$$PRESS = \sum_{i=1}^n (Y_i - \hat{Y}_{prog_{i(i)}})^2 \quad (5)$$

### 3. Tulemused

#### 3.1. Elektrofiilse kaksiksideme mehhanismi järgi toimivad ained

Kõigepealt tuletati QSAR mudel kogu andmekomplektile, s.o polaarse kaksiksideme mehhanismi järgi toimivale 84 ainele. Selleks kasutati BMLR meetodit ja deskriptoritena kogumolekuli deskriptoreid (s.t deskriptorite hulgast eemaldati asukohaspetsiifilised deskriptorid). Esimestest katsetustest oli näha, et 1-(4-hüdroksüfenüül)-3-fenüülprop-2-een-1-oon (ID: 1; info: Tabel L1; struktuur Tabel L2 lisades) hälbis tunduvalt teistest ühenditest andmeseerias, seega jäeti see edasisest analüüsist välja. BMLR meetodi abil tuletati kuni 10 deskriptoriga mudelid ja neid analüüsiti optimaalse deskriptorite hulga leidmiseks. Selleks uuriti  $R^2$  (ka  $R^2_{cv}$ ) muutu deskriptorite arvu kasvades (Graafik 1.A.). Saadud seose põhjal järeldati, et optimaalne deskriptorite arv on 4 (Tabel 1, Graafik 1.B.), kuna edasised deskriptorid ei paranda olulisel määral korrelatsiooni ja ennustusvõime ( $R^2_{cv}$ ) väheneb võrreldes  $R^2$  väärtusega ja muutub samaaegselt väiksemaks.



Graafik 1.A. Deskriptorite arv ja sellele vastavate mudelite determinatsioonikoefitsiendid. B. Esialgsele 83 ühendiga andmevalimile tuletatud nelja deskriptoriga mudel.

Tabel 1. 83 ühendile tuletatud parim 4-parameetriline mudel, deskriptorid ja statistilised parameetrid

$$R^2 = 0,5888; R^2_{CV} = 0,5232; F = 27,92; s^2 = 0,2665, n = 83$$

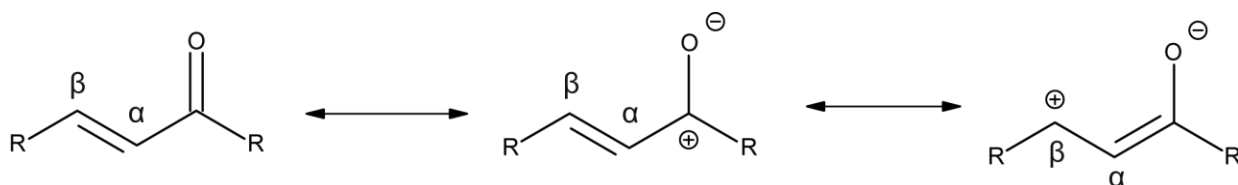
| nr | X       | $\pm\Delta X$ | t-test  | Deskriptor        |
|----|---------|---------------|---------|-------------------|
| 0  | 73,729  | 12,922        | 5,706   | Vabaliige         |
| 1  | -72,753 | 13,004        | -5,5946 | $\bar{V}_H$       |
| 2  | -67,921 | 14,985        | -4,533  | $N_A$             |
| 3  | -5,2492 | 1,2586        | -4,1706 | FHDSA             |
| 4  | 33,106  | 8,6989        | 3,8058  | $^{HA}HDCA^{(1)}$ |

Kõige olulisemate deskriptorite hulka kuuluvad t-testi absoluutväärtuse järgi (Tabel 1) kaks kvantkeemilist deskriptorit: vesiniku aatomi keskmine valents ( $\bar{V}_H$ ) ja süsiniku aatomi keskmine nukleofiilne reaktisoonivõime indeks ( $N_A$ ). Mõlemad deskriptorid võtavad arvesse andmeseerias olevate ühendite reaktisoonivõimet. Samuti on mõlemad negatiivse korrelatsiooniga, mis tähendab, et nende deskriptorite väärtuste vähenedes suureneb oht, et ühend tekitab AKD. Kuna reaktisoonivõime iseloomustamine kvantkeemiliste deskriptoritega (nagu reaktisoonivõime indeksid) oli üks käesoleva töö eesmärke, siis on nukleofiilse reaktiivuse indeksi sisaldumine mudelis oluline tulemus. Vesiniku aatomi keskmise valentsi ( $\bar{V}_H$ ) tähtsust on mõnevõrra keerulisem reaktisoonimehhanismi seisukohast interpreteerida ja seetõttu jäeti see edasises analüüsis kõrvale. Ülejäänud kaks deskriptorit on vesiniksidemete aktseptorist sõltuv laetud vesiniksideme doonori osapindala ( $^{HA}HDCA^{(1)}$ ) ja vesiniksideme doonori osapindala (FHDSA). Need deskriptorid iseloomustavad eri aspekte, mis on seotud vesiniksidemetega ja on tõenäoliselt seotud ühendi jõudmisega reaktisoonikeskkonda ning seal püsimisega.

Esimese mudeli ennustusvõime on suhteliselt madal, mis võib olla tingitud andmevalimi liiga suurest varieeruvusest (ligikaudu pooled ühenditest on prohaptendid). Tõenäoliselt on see varieeruvus (ja kaasnevad madala determinatsioonikoefitsienfiga mudelid) põhjuseks ka sellele, miks polaarse kaksiksideme rühma ühenditele ei ole varem avaldatud kõiki sellesse mehhanismirühma kuuluvaid ühendeid sisaldavaid mudeleid (vt 3.2.2. Lokaalsed mudelid ehk mehhanismipõhine lähenemine). Sellest lähtuvalt otsustati paremini tõlgendatavate ning suurema korrelatsioonikoefitsiendi ruuduga tulemuste saamiseks andmevalimit piirata nii, et ühendite struktuuride varieeruvus oleks võimalikult väike (vt 2.1. Andmevalim).

### 3.2. Ained, milles kaksiksidet mõjutab keto-rühm

Edaspidise töö käigus kasutati struktuurselt võimalikult vähe varieeruvat piiratud andmevalimit ja lähtuti konkreetsest reaktsioonimehhanismist. Algsest andmevalimist moodustavad piiratud andmevalimi ühendid kõige suurema struktuurselt sarnase (aga mitte prohaptena käituvat) alamhulga. Nimelt kõigi piiratud andmevalimi ühendite puhul polariseerib kaksiksidet keto-rühm. Kui elektrone äratõmbavaks rühmaks on keto-rühm, vastab alamstruktuurile mitu resonantspiirstruktuuri (Skeem 3). Resonantspiirstruktuuridest on näha, et elektrofiilne reaktsioonitsenter võib olla nii karbonüülse kui ka  $\beta$ -süsiniku juures. Seega määravad ühendi reaktsioonivõime fragmendi  $C=C-C=O$  elektroonsed omadused.



*Skeem 3. Piiratud andmevalimis sisalduvatele ühenditele iseloomuliku fragmendi resonantspiirstruktuurid*

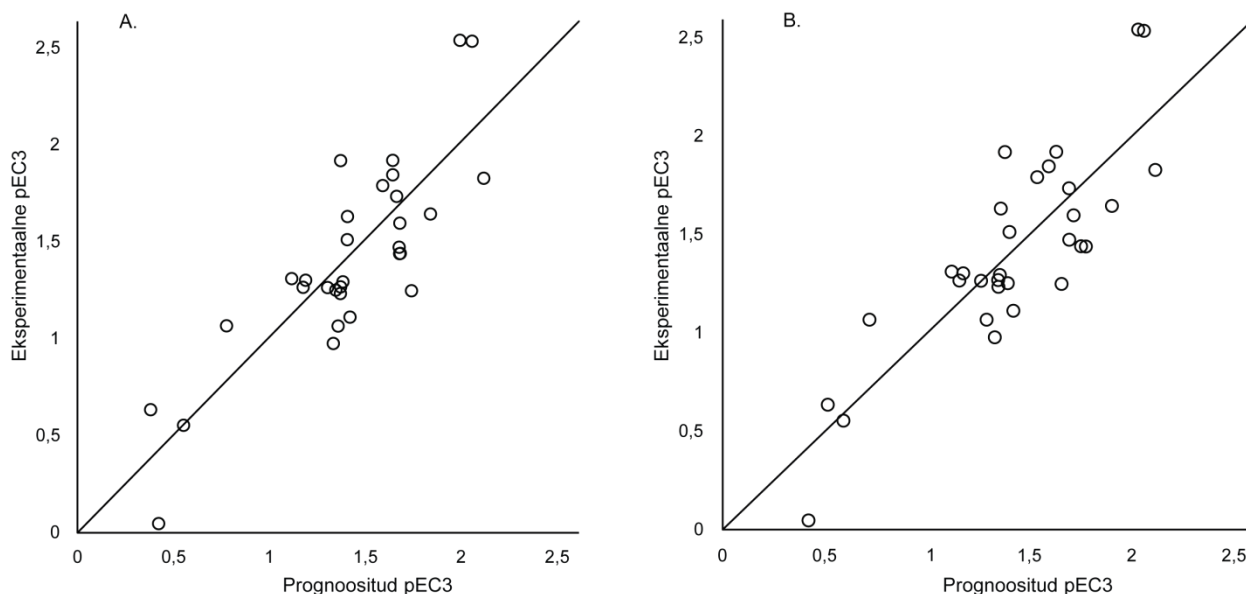
Selleks, et arvesse võtta  $C=C-C=O$  alamstruktuuri elektroonseid omadusi, arvutati alamstruktuurile fragmendideskriptorid ja need lisati kogumolekuli deskriptoritele. Seejärel, lähtudes suhtest minimaalselt üks molekulaardeskriptor kümne andmeseerias oleva ühendi kohta, teostati kolmeparameetriline heuristiline analüüs. Heuristilise analüüsi väljastatud kümmet parimat võrrandit analüüsiti ja leiti, et interpreteeritavuse poolest on parim võrrand, milles sisaldus fragmendi kirjeldav nukleofiilse reaktsioonivõime indeks ja lisaks sellele veel kaks kogumolekuli deskriptorit (Tabel 2.1., Graafik 2.A. Tabel L3 Lisades). See mudel on statistiliselt märkimisväärselt parem kui kogu andmevalimile tuletatud mudel (vt 3.1. Elektrofiilse kaksiksideme mehhanismi järgi toimivad ained).

Tabel 2. 1. Piiratud andmevalimiga ja fragmendideskriptoritega tuletatud mudel. 2. Piiratud andmevalimiga ja aatomideskriptoritega tuletatud mudel.

| 1. $R^2 = 0,7043$ ; $R^2_{CV} = 0,5766$ ; $F = 21,43$ ; $s^2 = 0,0846$ , $n = 31$ |           |               |         |            |        |            |
|---|-----------|---------------|---------|------------|--------|------------|
| nr  | X         | $\pm\Delta X$ | T-test  | Deskriptor | $R^2$  | $R^2_{CV}$ |
| 0   | 1,5052    | 46,528        | 3,235   | Vabaliige  |        |            |
| 1   | 0,39815   | 0,064821      | 6,1422  | $\mu$      | 0,4670 | 0,3797     |
| 2   | -0,821138 | 0,3102        | -2,6479 | $^0AIC$    | 0,6661 | 0,5585     |
| 3   | -26,541   | 14,22         | -1,8665 | f- $N_A$   | 0,7043 | 0,5766     |

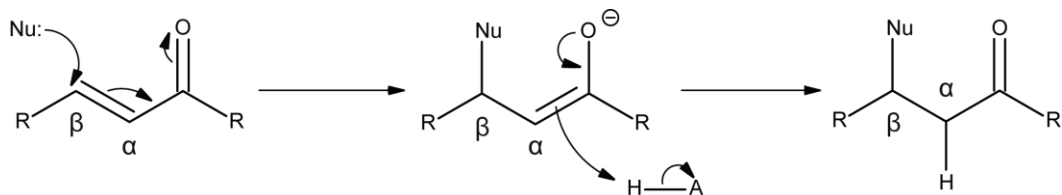
  

| 2. $R^2 = 0,6995$ ; $R^2_{CV} = 0,5672$ ; $F = 20,95$ ; $s^2 = 0,0859$ , $n = 31$ |          |               |         |                 |        |            |
|---|----------|---------------|---------|-----------------|--------|------------|
| nr  | X        | $\pm\Delta X$ | T-test  | Deskriptor      | $R^2$  | $R^2_{CV}$ |
| 0   | 1,5964   | 0,45911       | 3,4771  | Vabaliige       |        |            |
| 1   | 0,41514  | 0,063262      | 6,5623  | $\mu$           | 0,4670 | 0,3797     |
| 2   | -0,90353 | 0,29532       | -3,0594 | $^0AIC$         | 0,6661 | 0,5585     |
| 3   | -9,4804  | 5,4723        | -1,7324 | ( $\beta C$ )-N | 0,6995 | 0,5672     |



Graafik 2. A. Piiratud andmevalimiga ja fragmendideskriptoritega tuletatud mudel. B. Piiratud andmevalimiga ja aatomideskriptoritega tuletatud mudel

Järgnevalt prooviti reaktsioonivõimet veel spetsiifilisemalt kirjeldada. Selleks vaadeldi nukleofiilse liitumisreaktsiooni mehhanismi kui liitumine toimub polaarse kaksiksidemega ühenditele (Skeem 4). Mehhanismi järgi võiks eeltoodud fragmendi aatomitest tähtsaimat rolli omada  $\beta$ -süsinik, sest selle süsiniku juurde toimub kõige tõenäolisemalt nukleofiilne liitumine.



Skeem 4. Nukleofiilne liitumine polaarsele kaksiksidemele

Seega eeldati, et  $\beta$ -süsiniku reaktsioonivõime kirjeldab suurema osa kogu fragmendi reaktsioonivõimest. Sellest lähtuvalt arvutati  $\beta$ -süsinikule aatomideskriptorid ja need lisati kogumolekuli deskriptoritele. Sellele järgnes kolmeparameetiline heuristiline analüüs, mille väljunditeks saadud kümne parima sõltuvuse seas oli kõige paremini interpreteeritav mudel, mis sisaldas  $\beta$ -süsiniku nukleofiilse reaktsioonivõime indeksit. Ülejäänud kaks deskriptorit olid kogumolekuli deskriptorid (Tabel 2.2., Graafik 2.B., Tabel L3 Lisades). Teine (Tabel 2.1) ja kolmas (Tabel 2.2) mudel on statistiliste parameetrite ja deskriptorite poolest väga sarnased. Esimesed kaks deskriptorit on täpselt samad. Olulisim deskriptor t-testi väärtuse poolest on molekuli kogu dipoolimoment ( $\mu$ ). See on vektorsuurus, mis kirjeldab laengujaotust molekulis ja on seotud molekuli lahustumisega ja polariseeritavusega eri keskkondades (sh naha eri osades). Üheparameetiline sõltuvus  $pEC3$  ja dipoolimoment ( $\mu$ ) vahel annab märkimisväärse korrelatsiooni,  $R^2 = 0,4670$  mis näitab selle deskriptori olulisust EC3 modelleerimisel. Molekuli kogu dipoolmoment on seotud ka ühendi suurusega. Seda on näha Graafik 3.A., kus kõige madalamal asetsevad ühendid (ID: 80, ID: 84, ID: 77) on väikesed võrreldes keskmiste ja kõige kõrgemal asetsevate ühenditega. Kõige kõrgema kogu molekuli dipoolmomendi väärtusega on metüül-2-üünaat (ID: 17) ja metüülnon-2-üünaat (ID: 39), kõrge dipoolmomendi põhjuseks on ilmselt kolmikside.

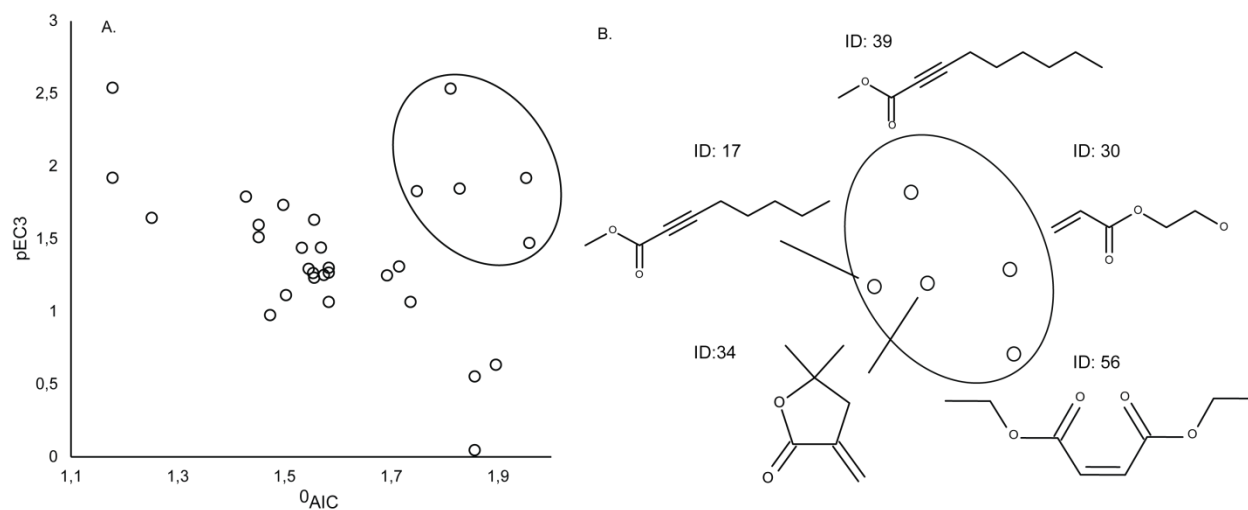
Teine deskriptor, keskmise informatsioonisisalduse nullindat järku indeks (*Average Information Content (order 0)*,  ${}^0AIC$ ), võtab arvesse molekulide atomaarset varieeruvust. See loetleb ja grupeerib ühendites sisalduvaid erinevaid aatomeid (6) [36], kus  $n$  on kogu aatomite arv molekulis,  $n_i$  on aatomite arv klassis  $i$  ja aatomite jagamine klassideks sõltub arvesse võetavast koordinatsioonisfäärast. Koordinatsioonisfääri järgi määratakse ka keskmine informatsioonisisalduse järk. Kui järk on 0, siis keemilisi sidemeid arvesse ei võeta ja klassidesse jagamine toimub keemiliste elementide järgi (Graafik 4). [36] Selle deskriptori lisamisega tõuseb

korrelatsioon koefitsiendi ruut umbes 0,2 võrra (Tabel 2.2.: kaheparameetrilise mudeli  $R^2 = 0,6661$ ).

$${}^k AIC = - \sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n} \quad (6)$$



Graafik 3. A. Ühendite jaotumine pEC3 ja kogu dipoolmomendi järgi B. A-osas ID-ga märgitud punktidele vastavad struktuurid.



Graafik 4. A. Ühendite jaotumine pEC3 ja keskmise informatsioonisisalduse ( ${}^0$  järku) järgi. B. Hälbijad keskmise nullindat järku informatsioonisisalduse järgi

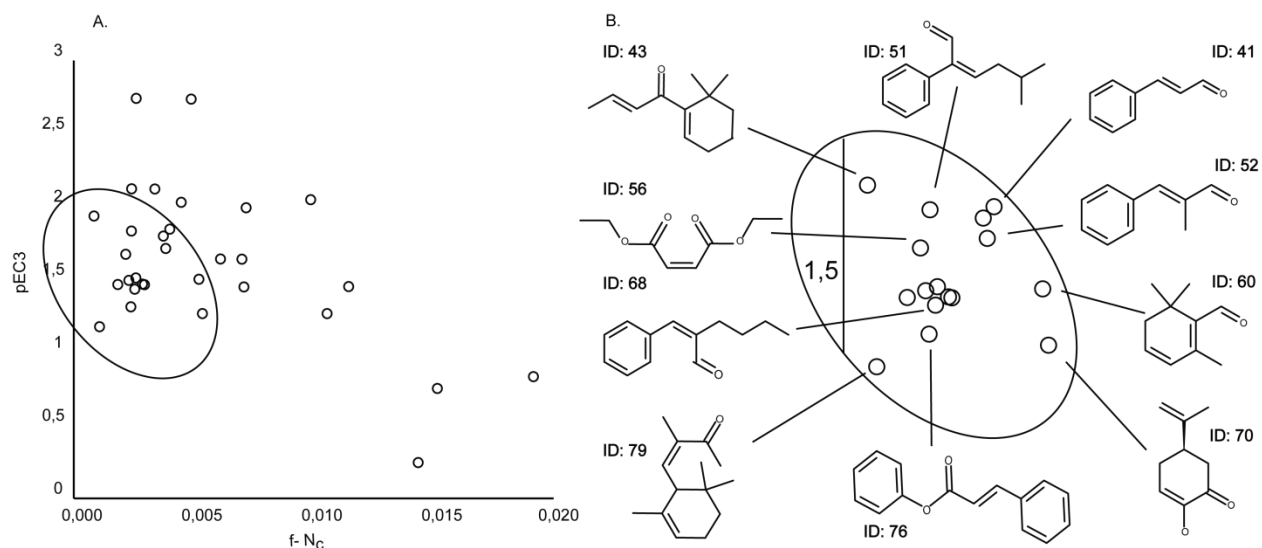
Neljandal graafikul on näha, et viis molekuli kalduvad üldisest trendist kõrvale. Kaks neist metüül-2-oktünoaat (Graafik 3.B. ID: 39) ja metüül-2-nonünoaat (Graafik 3.B. ID: 17) on

ainukesed ühendit valimis, mis sisaldavad kolmiksidet. Ülejäänud kolm sisaldavad suhteliselt palju hapniku aatomeid võrreldes ülejäänud ühenditega.

Teise ja kolmanda mudelite erinevus tuleb kolmandast deskriptorist. Teises mudelis (Tabel 2.2.) on kolmandaks deskriptoriks fragmendile C=C-C=O arvutatud nukleofiilse reaktiivsuse indeks. Nukleofiilse reaktiivsuse indeks arvutatakse kui keskmine kõigi süsiniku aatomite nukleofiilsest reaktiivsusest (CODESSA käsiraamatu järgi [36], (7)).

$$N_A = \sum_{i \in A} \frac{c_{iHOMO}^2}{\varepsilon_{LUMO} - \varepsilon_{HOMO}} \quad (7)$$

Summeerimine teostatakse üle kõigi aatomorbitaalide,  $c_{iHOMO}$  – on aatomorbitaali numbriga  $i$  koefitsiendid kõrgeima energiaga hõivatud molekulaarorbitaalil (HOMO),  $\varepsilon_{HOMO}$  – HOMO-orbitaali energia,  $\varepsilon_{LUMO}$  – mudelelektrofiili madalaima energiaga hõivamata orbitaali energia. Kuna see on konstantne suurus, siis CODESSAs on vaikumisi selle väärtuseks  $\varepsilon_{LUMO} = 1$  (suurel osal ühenditest on LUMO väärtus  $\sim 1$ ). [36]. Mida suurem on selle deskriptori väärtus, seda väiksem on selle aatomi võime käituda elektrofiilina ja põhjustada allergilist nahapõletikku. Hajuvusgraafikul, kus ordinaatteljel on pEC3 ja abstsiss-teljel fragmendi C=C-C=O nukleofiilse reaktsioonivõime indeks, on vastavalt näha negatiivne trend (Graafik 5.A).

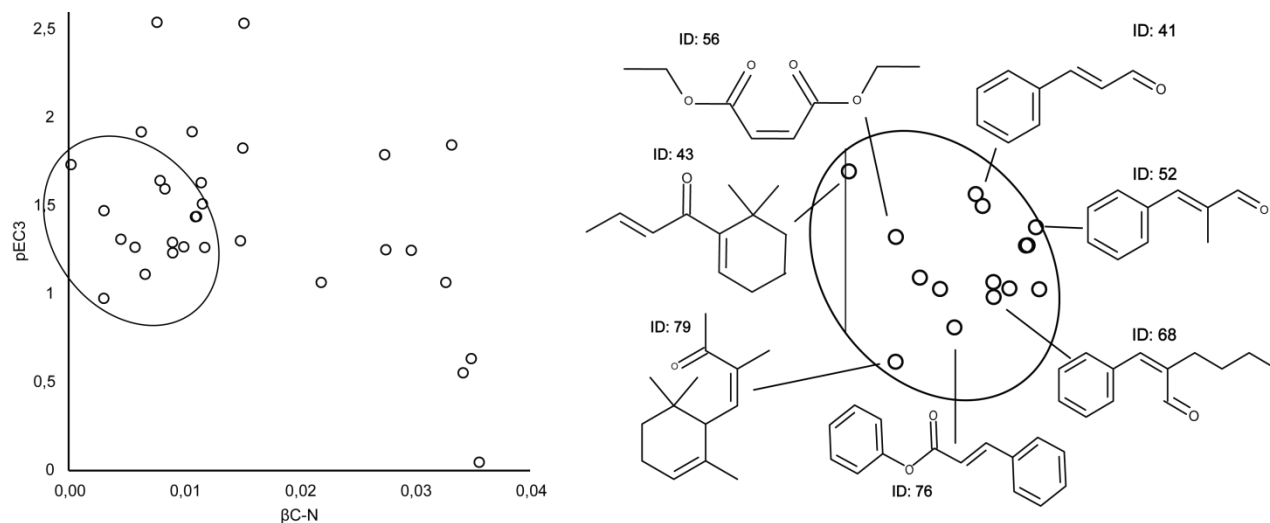


Graafik 5. A. Ühendite jaotumine fragmendi [C=C-C=O] keskmise nukleofiilsuse indeksi ja pEC3 järgi. B. Eraldi grupi moodustavad ühendid keskmise nukleofiilse reaktiivsuse indeksi järgi.

Negatiivse trendi (Graafik 5.A) juures on näha, et osa ühenditest moodustab kõrvale kalduva grupi. Selles grupis on ühendid, millest enamiku puhul on reaktsioonitsenter steeriliselt

ebasoodne. Veel on nendel struktuuridel ühist see, et polaarse kaksiksideme juures on elektrondonoorseid rühmi (bensüül, metüül jmt). Need ühendid ka pEC3 väärtuste poolest sarnased (Graafik 5.B.). Kolmanda mudeli (Tabel 2.2.) kolmas deskriptor on  $\beta$ -süsinikule (Skeem 4) arvutatud nukleofiilse reaktsioonivõime indeks. Ühendite jaotumine pEC3 ja nukleofiilse reaktsioonivõime indeksi järgi (Graafik 5.B) näitab, et eraldi grupi moodustavad samad ja sarnased ühendid nagu fragmendile arvutatud nukleofiilse reaktsioonivõime indeksi korral (Graafik 6).

Teise ja kolmanda mudeli statistilised parameetrid erinevad teineteisest väga vähe. Seega on suurem osa fragmendi nukleofiilsest reaktiivsusest kirjeldatav  $\beta$ -süsiniku nukleofiilse reaktiivsusega. Samas ühedeskriptoriliste sõltuvuste determinatsioonikoefitsientid pEC3ga fragmendile arvutatud nukleofiilse reaktsioonivõime indeksil ja  $\beta$ -süsinikule arvutatud nukleofiilse reaktsioonivõime indeksil vastavalt  $R^2 = 0,280$  ja  $R^2 = 0,186$ . Need on küllalt erinevad võrreldes teise ja kolmanda mudeli determinatsioonikoefitsientide väikese erinevusega. Järelikult tuleks reaktsioonivõime arvestamiseks eelistada fragmendile arvutatud deskriptorit, sest see (erinevalt  $\beta$ C-N) võtab arvesse elektritiheduse delokalatsiooni, mis suure tõenäosusega omab olulist mõju reaktsioonivõimele. Samuti on fragmendile arvutatud nukleofiilse reaktsioonivõime indeksiga võrrand  $R^2_{CV}$  alusel natuke parema ennustusvõimega (Tabel 2.1.  $R^2_{CV} = 0,5766$  võrreldes Tabel 2.2.  $R^2_{CV} = 0,5672$ ).



Graafik 6. Ühendite jaotumine  $\beta$ -süsiniku nukleofiilse reaktiivsuse indeksi ja pEC3 järgi ja eraldi grupi moodustavad ühendid

## Kokkuvõte

Allergiline kontaktdermatiit on levinud nahahaigus, millel on haigestunule eluaegne mõju. Kuna seda haigust ei saa lõplikult ravida, on oluline ennetamine ja selle jaoks on vaja hinnata kemikaalide potentsiaali allergilist kontaktdermatiiti tekitada. Ainete ohtikkuse hindamine loomkatsete abil on küll usaldusväärne, kuid samaaegselt kallis ja aeganõudev ning seega vajatakse alternatiivseid testimismeetodeid. Praeguse seisuga on perspektiivikad *in silico* meetodid, mis võimaldavad efektiivselt vähendada loomkatsete arvu ja suurendada hinnatud kemikaalide hulka.

Töös anti ülevaade allergilise kontaktdermatiidi toimemehhanismist ja *in silico* modelleerimisvõimalustest. Arvutusliku eksperimendi eesmärgiks oli polaarset kaksiksidet sisaldavate ühendite reaktsioonivõime kirjeldamiseks kasutada kvantkeemilisi reaktsioonivõime deskriptoreid. Selle jaoks tuletati esialgu mudel kõigile kirjandusest leitud ühenditele, mis toimivad polaarsete kaksiksideme mehhanismi järgi (nukleofiilne liitumine polaarsetele kaksiksidemele) ja millel on mõõdetud EC3 väärtus. Esialgne mudel tuletati 83 ühendiga see sisaldas nelja molekulaardeskriptorit. Selle mudeli statistilised parameetrid ei olnud rahuldavad ja mudeli sisu oli mehhanistlikult raskesti interpreteeritav. Järgnevalt andmevalimit piirati: andmevalimisse jäeti ühendid, mis sisaldasid polaarset kaksiksidet ja sellega seotud karbonüülset süsinikku, kokku 31 ühendit. Piiratud andmevalimi ühendite reaktsioonitsentritele arvutati fragmendideskriptorid ja nukleofiiliga potentsiaalselt reageerivale süsinikule aatomideskriptorid. Tuletati kaks kolmedeskriptorilist mudelit, millest ühes on sees reaktsioonivõimet iseloomustav fragmendideskriptor ja teises reaktsioonivõimet iseloomustav aatomideskriptor (ülejäanud deskriptorid on samad). Nende mudelite statistilised parameetrid on sarnased (fragmendideskriptoriga  $R^2 = 0,7043$ , aatomideskriptoriga  $R^2 = 0,6695$ ).

Viimased kaks tuletatud mudelit näitavad, et keemilisi interaktsioone, mis määravad polaarsete kaksiksidemega ainete võime tekitada allergilist nahapõletikku, saab kirjeldada reaktsioonivõimet iseloomustavate kvantkeemiliste molekulaarsete deskriptoritega, kuid nendele lisaks on vaja ka kogumolekuli deskriptoreid, mis kirjeldavad molekuli üldisi omadusi (laengujaotust, polaarsust/polariseeritavust ja atomaarset mitmekülgsust).

## Summary in English

### A study of skin sensitization potency for Michael acceptors in the LLNA using Quantitative structure-activity relationships

Skin sensitization is a widespread skin disease which has a lifelong impact. The possibility of individuals becoming sensitized is a major problem therefore there is a need to accurately identify chemicals that have the potential to cause skin sensitization. Reliable evaluation of chemicals for skin sensitization can be made using animal testing, but as *in vivo* tests are expensive and time-consuming there is a growing need for the alternative methods. The methods foreseen partially replace animal tests are *in silico* methods. *In silico* methods like QSARs can effectively reduce animal testing and increase the number of evaluated chemicals.

This thesis provides a brief overview of the mechanism of skin sensitization and the current approaches for modelling the potencies of chemicals. The aim of this study was to apply quantum chemical descriptors in order to describe the reactivity of Michael acceptors. To begin with, a model was derived for all Michael acceptors found in the literature, which have been measured by LLNA and thus have an EC3 value. The preliminary model was derived with 83 chemicals and contained four molecular descriptors. The statistical parameters of the equation were not satisfactory and it was on the whole difficult to interpret mechanistically. Subsequently the dataset was narrowed down to substances, which included a double bond next to a carbonyl carbon. Altogether, there were 31 substances in the restricted dataset. After that, more specific descriptors were calculated. Namely fragment descriptors for the C=C-C=O fragment that is responsible for the activity and atomic descriptors for the most probable atom ( $\beta$ -carbon) to undergo reaction. Two three-parameter models were derived, first of which contains fragment-based nucleophilic reactivity index and second of which includes atomic nucleophilic reactivity (other descriptors are identical). These models were statistically greatly similar as the determination coefficient was  $R^2 = 0,7043$  for the first and  $R^2 = 0,6695$  for the second.

These last two models derived with the narrowed dataset indicate that it is feasible to use quantum chemical descriptors for developing skin sensitization models for Michael acceptors but they should be accompanied by the whole-molecule descriptors that capture general properties of the molecules (charge distribution, polarity/polarizability, atomic diversity).

#### 4. Kirjandus

1. Vorvick, L. J.; Zieve, D. T. *Contact dermatitis*. <http://www.nlm.nih.gov/medlineplus/ency/article/000869.htm> viimati uuendatud 21.11.2011.
2. Landsteiner, K.; Chase, M. V. Studies on the sensitization of animals with simple chemical compounds IX. Skin sensitization induced by injection of conjugates. *J. Exp Med.* **1941**, 73(3), 431 – 438.
3. Hennino, A.; Vocanson, M.; Chavagnac, C.; Saint-Mezard, P.; Dubois, B.; Kaiser, D.; Nicolas, J.-F. Update on the pathophysiology with special emphasis on CD8 effector T cells and CD4 regulatory T cells. *An Bras Dermatol.* **2006**, 80(4), 335 – 347.
4. Camilla, K.; Smith Pease. From xenobiotic chemistry and metabolism to better prediction and risk assessment of skin allergy. *Toxicology.* **2003**, 192, 1 – 22.
5. Aptula, A. O.; Patlewicz, G.; Roberts, D. W. Skin sensitization: mechanistic applicability domains for structure-activity relationships. *Chem. Res. Toxicol.* **2005**, 18, 1420 – 1426.
6. Roberts, D. W.; Aptula, A.; Determinants of skin sensitisation potential. *Journal of applied toxicology.* **2008**, 28, 377 – 387.
7. Pollard, P. C.; Moriarty, D. J. Validity of the tritiated thymidine method for estimating bacterial growth rates: measurement of isotope dilution during DNA synthesis. *Appl. Environ. Microbiol.* **1984**, 48(6), 1076 – 1083.
8. Gerberick, G. F.; Ryan, C. A.; Dearman, R. J.; Kimber, I. Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. *Methods (Elsevier).* **2007**, 41 – 60.
9. Roberts, D. W.; Aptula, A. O.; Patlewicz, G. Mechanistic applicability domains for non-animal based prediction of toxicological endpoints. QSAR analysis of the Schiff base applicability domain for skin sensitization. *Chem. Res Toxicol.* **2006**, 19, 1228 – 1233.
10. Patlewicz, G.; Worth, A. Review of data sources, QSARs and Integrated testing strategies for skin sensitization. *JRC Scientific and Technical Reports.* **2008**.
11. Li, Y.; Pan, D.; Liu, J.; Kern, P. S.; Gerberick G. F.; Hopfinger, A. J.; Tseng, Y. J. Categorical QSAR models for skin sensitization based upon local lymph node assay classification measures part 2: 4D-fingerprint three-state and two-2-state logistic regression models. *Toxicological sciences.* **2007**, 99(2), 532 – 544.
12. Estrada, E.; Patlewicz, G.; Chamberlain, M.; Basketter, D.; Larbey, S. Computer-aided knowledge generation for understanding skin sensitization mechanisms: the TOPS-MODE approach. *Chem. Res. Toxicol.* **2003**, 16, 1226 – 1235.
13. Yuan, H.; Huang, J.; Cao, C. Prediction of skin sensitization with particle swarm optimized support vector machine. *Int. J. Mol. Sci.* **2009**, 10, 3237 – 3254.

14. Gunturi, S. B.; Theerthala, S. S.; Patel, N. K.; Bahl, J.; Narayanan, R. Prediction of skin sensitization potential using D-optimal design and GA-kNN classification methods. *SAR and QSAR in Environmental Research*, **2010**, 21, 305 – 335.
15. Golla, S.; Madihally, S.; Robinson, R. L.; Gasem, K. A. M. Quantitative structure-property relationship modeling of skin sensitization: A quantitative prediction. *Toxicology in Vitro*. **2009**, 23, 454 – 465.
16. Roberts, D. W.; Williams, D. L. The derivation of quantitative correlations between skin sensitisation and physico-chemical parameters for alkylating agents and their application to experimental data for sultones. *J Theor Biol*. **1982**, 99, 807 – 825.
17. Roberts, D. W.; Benezra, C. Quantitative structure-activity relationships for skin sensitization potential of urushiol analogues. *Contact Dermatitis*. **1993**, 29, 78 – 83.
18. Roberts, D. W.; York, M.; Basketter, D. A. Structure-activity relationships in the murine local lymph node assay for skin sensitization: Alpha, beta-diketones. *Contact Dermatitis*. **1999**, 41, 14 – 17.
19. Patlewicz, G.; Basketter, D. A.; Smith, C. K.; Hotchkiss, S. A.; Roberts, D. W. Skin sensitization structure-activity relationships for aldehydes. *Contact Dermatitis*. **2001**, 44, 331 – 336.
20. Roberts, D. W.; Patlewicz, G. Mechanism based structure-activity relationships for skin sensitisation-the carbonyl group domain. *SAR QSAR Environ Res*. **2002**, 13, 145 – 152.
21. Roberts, D. W. Structure-activity relationships for skin sensitisation potential of diacrylates and dimethacrylates. *Contact Dermatitis*. **1987**, 17, 281 – 289.
22. Roberts, D. W.; Aptula, A. O.; Patlewicz, G. Y. Chemistry-based risk assessment for skin sensitization mechanistic modeling for the S<sub>N</sub>Ar domain. *Chemical Research in Toxicology*. **2011**, 24, 1003 – 1011.
23. Roberts, D. W.; Patlewicz, G.; Kern, P. S.; Gerberick, F.; Kimber, I.; Dearman, R. J.; Ryan, C. A.; Basketter, D. A.; Aptula, A. O. Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chemical Research in Toxicology*. **2007**, 20, 1019 – 1030.
24. Devillers, J., Toim. *Ecotoxicology modeling*, Springer: New York; 2009; 1k 61 – 79.
25. Roberts, D. W.; Natsch, A. High throughput kinetic profiling approach for covalent binding to peptides: application to skin sensitization potency of Michael acceptor electrophiles. *Chem. Res. Toxicol*. **2009**, 22, 592 – 603.
26. Enoch, S. J.; Roberts, D. W. Predicting skin sensitization potency for michael acceptors in the LLNA using quantum mechanics calculations. *Chemical Research in Toxicology*. **2013**, 26, 767 – 774.
27. Gerberick, G. F.; Ryan, C. A.; Kern, P. S.; Schlatter, H.; Dearman, R. J.; Kimber, I.; Patlewicz, G. Y.; Basketter, D.A. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis*. **2005**, 16, 157 – 202.

28. Kern, P. S.; Gerberick, G. F.; Ryan, C. A.; Kimber, I.; Aptula, A.; Basketter, D. A. Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. *Dermatitis*. **2010**, 21, 8 – 32.
29. JChem for Office; internetis kättesaadav: <http://www.chemaxon.com/products/jchem-for-office/> (viimati vaadatud 19.04.2014)
30. SciFinder; internetis kättesaadav: <https://www.cas.org/products/scifinder/> (viimati vaadatud 19.04.2014)
31. Stewart, J. J. P. *MOPAC Program Package 6.0*, QCPE, No.445, 1989.
32. Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittkul, V. A practical overview of quantitative structure-activity relationship. *EXCLI Journal*. **2009**, 8, 74 – 88.
33. Peixun, L.; Long, W. Current mathematical methods used in QSAR/QSPR studies. *International Journal of Molecular Sciences*. **2009**, 10, 1978-1998.
34. Everitt, B. S. *The Cambridge dictionary of statistics (Second edition)*, Cambridge University Press: New York, 2002, lk 78.
35. Rawlings, J. O.; Pantula, S. G.; Dickey, D. A. *Applied Regression Analysis: A Research Tool (Second edition)*, Springer-Verlag: New York, 1998, lk 230 – 231.
36. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA: Reference Manual*, Gainesville: University of Florida, 1994.

## LISAD

**Tabel L1.** Kirjandusest leitud polaarse (elektrofiilse) kaksiksidemega ühendid, mille on mõõdetud EC3 (%) väärtus.

| ID | Nimetus  | CAS-number | EC3 (%) | Piiratud valimis | Viide |
|----|--|------------|---------|------------------|-------|
| 1  | 1-(4-hüdroksüfenüül)-3-fenüül-2-propeen-1-oon                                    | 2657-25-2  | 0.002   |                  | [28]  |
| 2  | tsükloheksa-2,5-dieen-1,4-dioon  | 106-51-4   | 0.0099  |                  | [27]  |
| 3  | N,N'-(2,5-diaminotsükloheksa-2,5-dieen-1,4-diülideen) bisbenseen-1,4-diamiin     | 20048-27-5 | 0.04    |                  | [28]  |
| 4  | 4-nitro-benseen-1,2-diamiin  | 99-56-9    | 0.05    |                  | [28]  |
| 5  | 4-[(2-hüdroksüetüül)amino]-3-nitrofenool   | 65235-31-6 | 0.07    |                  | [28]  |
| 6  | Benseen-1,4-diool  | 123-31-9   | 0.11    |                  | [27]  |
| 7  | 3-fenüül-1-(2,3,4-trihüdroksüfenüül)-prop-2-een-1-oon                            | 1482-74-2  | 0.11    |                  | [28]  |
| 8  | Benseen-1,4-diamiin  | 106-50-3   | 0.16    |                  | [27]  |
| 9  | 2-metüülbenseen-1,4-diamiin  | 95-70-5    | 0.2     |                  | [27]  |
| 10 | 4-amino-3-nitrofenool  | 610-81-1   | 0.2     |                  | [28]  |
| 11 | Dodeküül-3,4,5-trihüdroksübensoaat   | 1166-52-5  | 0.3     |                  | [27]  |
| 12 | 2-aminofenool  | 95-55-6    | 0.4     |                  | [27]  |
| 13 | 2-nitrobenseen-1,4-diamiin   | 5307-14-2  | 0.4     |                  | [27]  |
| 14 | 5-[(2-hüdroksüetüül)amino]-2-metüülfenool  | 55302-96-0 | 0.4     |                  | [27]  |
| 15 | 3-kloro-2,6-dihüdroksü-4-metüülbensaldehüüd                                      | 57074-21-2 | 0.4     |                  | [28]  |
| 16 | 2-metüülbenseen-1,4-diamiinsulfaat (1:1)   | 615-50-9   | 0.4     |                  | [28]  |
| 17 | Metüül-2-üünaat  | 111-12-6   | 0.45    | Jah              | [28]  |
| 18 | benseen-1,3-diamiin  | 108-45-2   | 0.49    |                  | [27]  |
| 19 | 1-(2,4-dihüdroksüfenüül)-3-fenüülprop-2-een-1-oon                                | 1776-30-3  | 0.56    |                  | [28]  |
| 20 | N-[2-[(4-amino-3-metüülfenüül)etüülamino]etüül]-metaansulfoonamiid sulfaat (2:3) | 25646-71-3 | 0.6     |                  | [27]  |

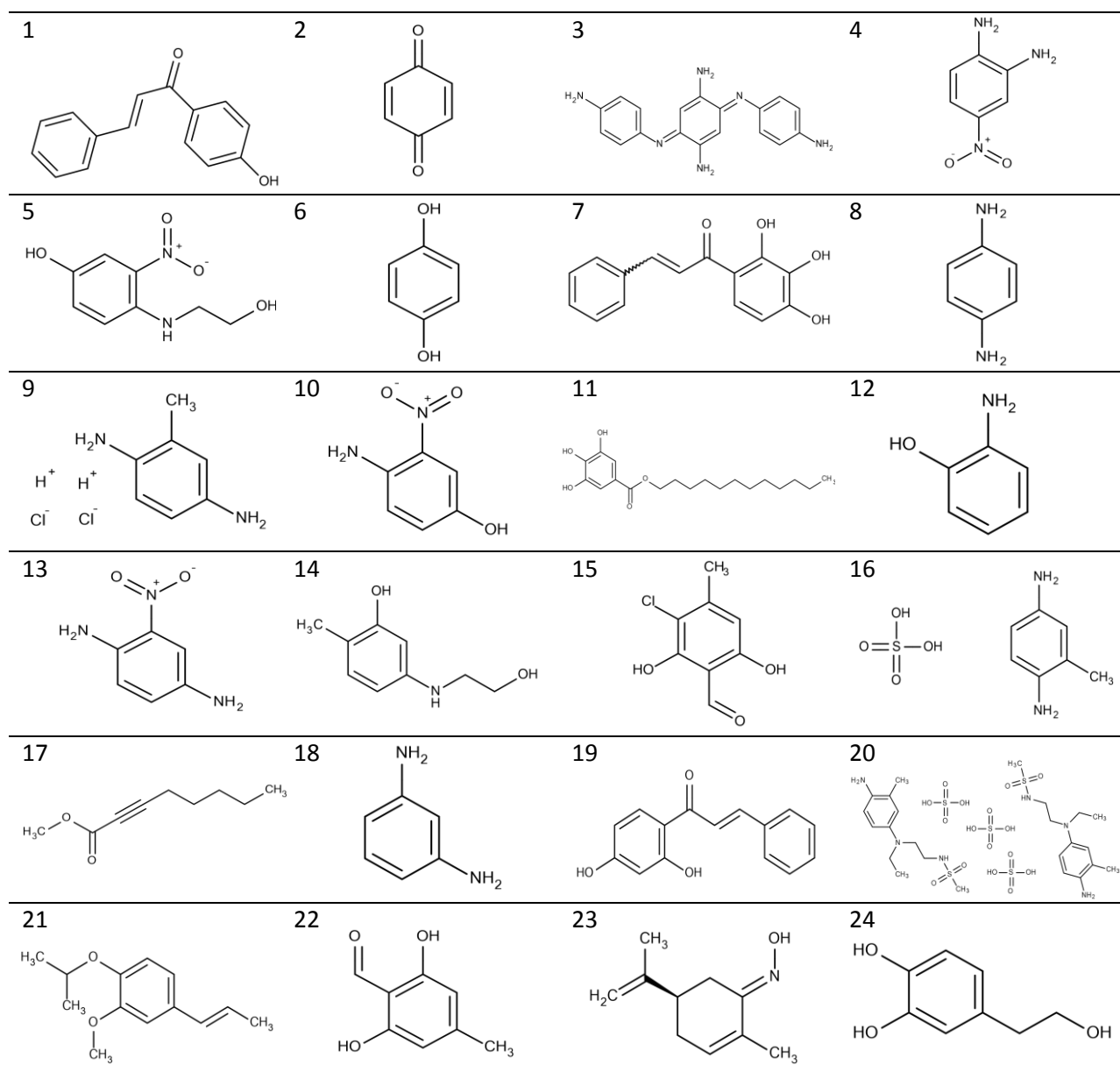
|    |  |            |      |     |      |
|----|--|------------|------|-----|------|
| 21 | 2-metoksü-1-(1-metüületoksü)-4-(prop-1-<br>eenüül)benseen      | 29653-00-7 | 0.6  |     | [27] |
| 22 | 2,6-dihüdroksü-4-metüülbensaldehüüd                            | 526-37-4   | 0.6  |     | [28] |
| 23 | 2-metüül-5-(1-metüületenüül)-tsükloheks-2-<br>een-1-oon oksiim | 55658-55-4 | 0.6  |     | [28] |
| 24 | 4-(2-hüdroksüetüül)-benseen-1,2-diool                          | 10597-60-1 | 0.6  |     | [28] |
| 25 | 3,3-difenüülprop-2-<br>enaal                                   | 1210-39-5  | 0.6  | Jah | [28] |
| 26 | 2[4-(metüülamino)fenool] sulfaat                               | 55-55-0    | 0.8  |     | [27] |
| 27 | 2-bromo-2-(bromometüül)-pentaandinitriil                       | 35691-65-7 | 0.9  |     | [27] |
| 28 | 2-metoksü-4-(prop-1-eenüül)-fenool                             | 97-54-1    | 1.2  |     | [27] |
| 29 | 3-metüül-4-fenüül-1,2,5-tiadiasool-1,1-<br>dioksiid            | 3775-21-1  | 1.4  |     | [27] |
| 30 | 2-hüdroksüetüül-akrülaat                                       | 818-61-1   | 1.4  | Jah | [27] |
| 31 | 4-amino-3-metüül-fenool  | 2835-99-6  | 1.45 |     | [28] |
| 32 | 2-metoksü-6-metüül-4-(prop-1-eenüül)-<br>fenool                | 13041-12-8 | 1.6  |     | [27] |
| 33 | etenüülpüridiin  | 1337-81-1  | 1.6  |     | [27] |
| 34 | Dihüdro-5,5-dimetüül-3-metülelen-2(3H)-<br>furanoon            | 29043-97-8 | 1.8  | Jah | [27] |
| 35 | 2-amino-6-kloro-4-nitrofenool                                  | 6358-09-4  | 2.2  |     | [27] |
| 36 | 2-[(4-amino-2-nitrofenüül)amino]-etanool                       | 2871-01-4  | 2.2  |     | [27] |
| 37 | 1-metoksü-4-(prop-1-eenüül)-benseen                            | 104-46-1   | 2.3  |     | [27] |
| 38 | Dek-2-enaal  | 3913-71-1  | 2.5  | Jah | [27] |
| 39 | Metüülnon-2-üünaat   | 111-80-8   | 2.5  | Jah | [27] |
| 40 | $\alpha$ -(fenüülmetülelen)-benseenatseetaldehüüd              | 1755-47-1  | 2.5  | Jah | [26] |
| 41 | 3-fenüül-prop-2-enaal  | 104-55-2   | 3    | Jah | [27] |

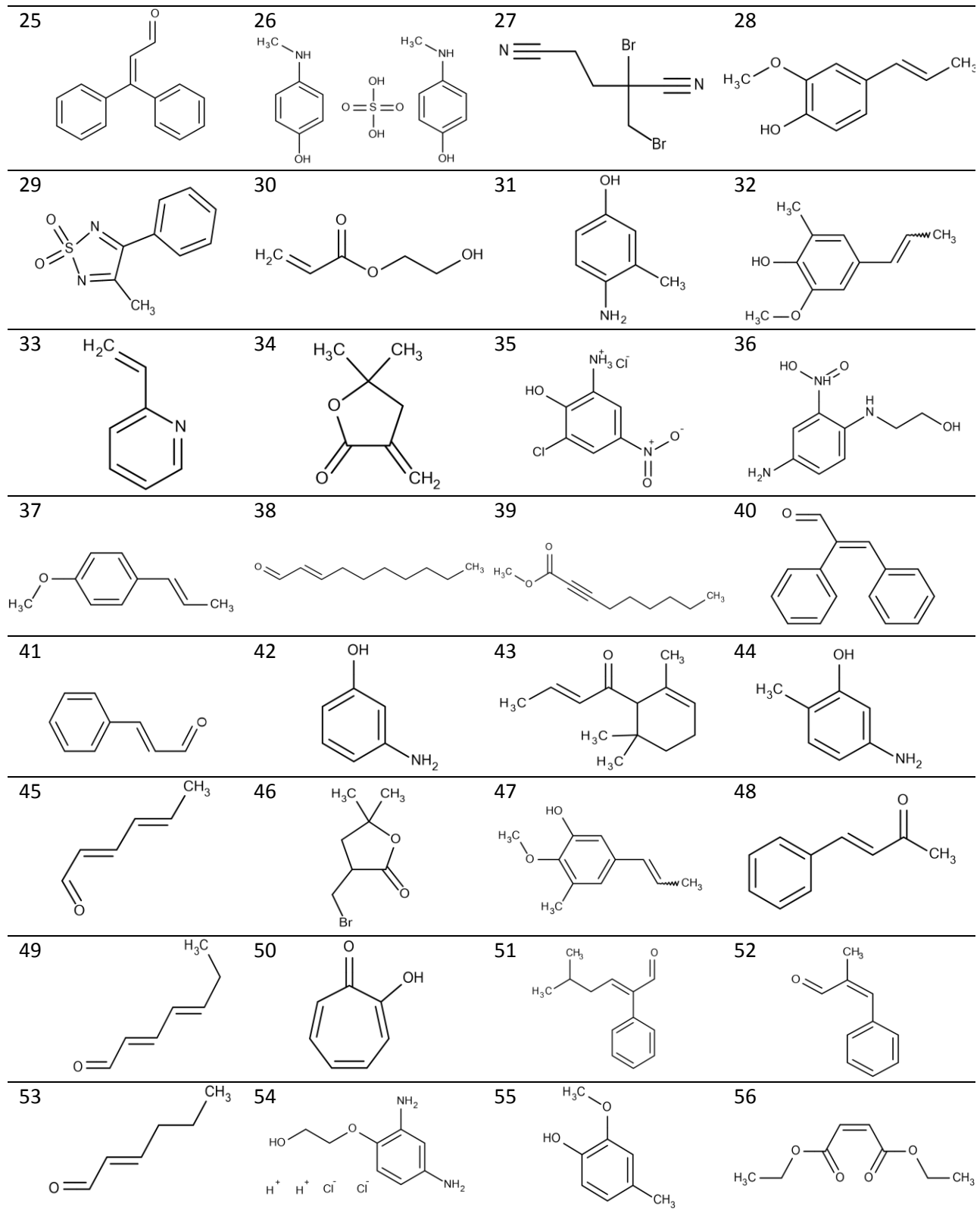
|    |  |             |     |     |      |
|----|--|-------------|-----|-----|------|
| 42 | 3-aminofenool  | 591-27-5    | 3.2 |     | [27] |
| 43 | 1-(2,6,6-trimetüül—tsükloheks-2-eeen-1-üül)-<br>but-2-eeen-1-oon | 24720-09-0  | 3.3 | Jah | [26] |
| 44 | 5-amino-2-metüülfenool   | 2835-95-2   | 3.4 |     | [28] |
| 45 | 2,4-heks-2,4-dienaal (2E, 4E)                                    | 142-83-6    | 3.5 | Jah | [28] |
| 46 | Dihüdro-5,5-dimetüül-3-metüleen-2(3H)-<br>furanon                | 29043-97-8  | 3.6 |     | [27] |
| 47 | 2-metoksü-3-metüül-4-(prop-1-eeen-1-üül)-<br>fenool              | 186743-29-3 | 3.6 |     | [27] |
| 48 | 4-fenüül-3-buteen-2-oon  | 122-57-6    | 3.7 | Jah | [27] |
| 49 | Hept-2,4-dienaal   | 5910-85-0   | 4   | Jah | [27] |
| 50 | 2-hüdroksü-tsüklohepta-2,4,6-trieen-1-oon                        | 533-75-5    | 4.3 |     | [28] |
| 51 | a-(3-metüülbutülideen)-<br>benseenatsetaldehüüd                  | 21834-92-4  | 4.4 | Jah | [28] |
| 52 | 2-metüül-3-fenüül-prop-2-enaal                                   | 101-39-3    | 4.5 | Jah | [27] |
| 53 | Heks-2E-enaal  | 6728-26-3   | 5.5 | Jah | [27] |
| 54 | 2-(2,4-diaminofenoksü)-etanool hüdrokloriid<br>(1:2)             | 66422-95-5  | 5.5 |     | [28] |
| 55 | 2-metoksü-4-metüül-fenool  | 93-51-6     | 5.8 |     | [27] |
| 56 | 1,4-dietüül but-2Z-eendiaat <i>diethyl maleate</i>               | 141-05-9    | 5.8 | Jah | [27] |
| 57 | 4-metoksü-benseenmetanool  | 105-13-5    | 5.9 |     | [28] |
| 58 | 2-metoksü-4-propüül-fenool                                       | 2785-87-7   | 6.8 |     | [27] |
| 59 | 3,5-diamino-2,6-dimetoksüpüridiin-<br>dihüdrokloriid             | 56216-28-5  | 6.9 |     | [28] |
| 60 | 1,1,3-trimetüül-2-formüültsükloheksa-2,4-<br>dieen               | 116-26-7    | 7.5 | Jah | [27] |
| 61 | 4-(1-metüületenüül)-tsükloheks-1-eeen-1-<br>karboksüaldehyüd     | 2111-75-3   | 8.1 | Jah | [27] |

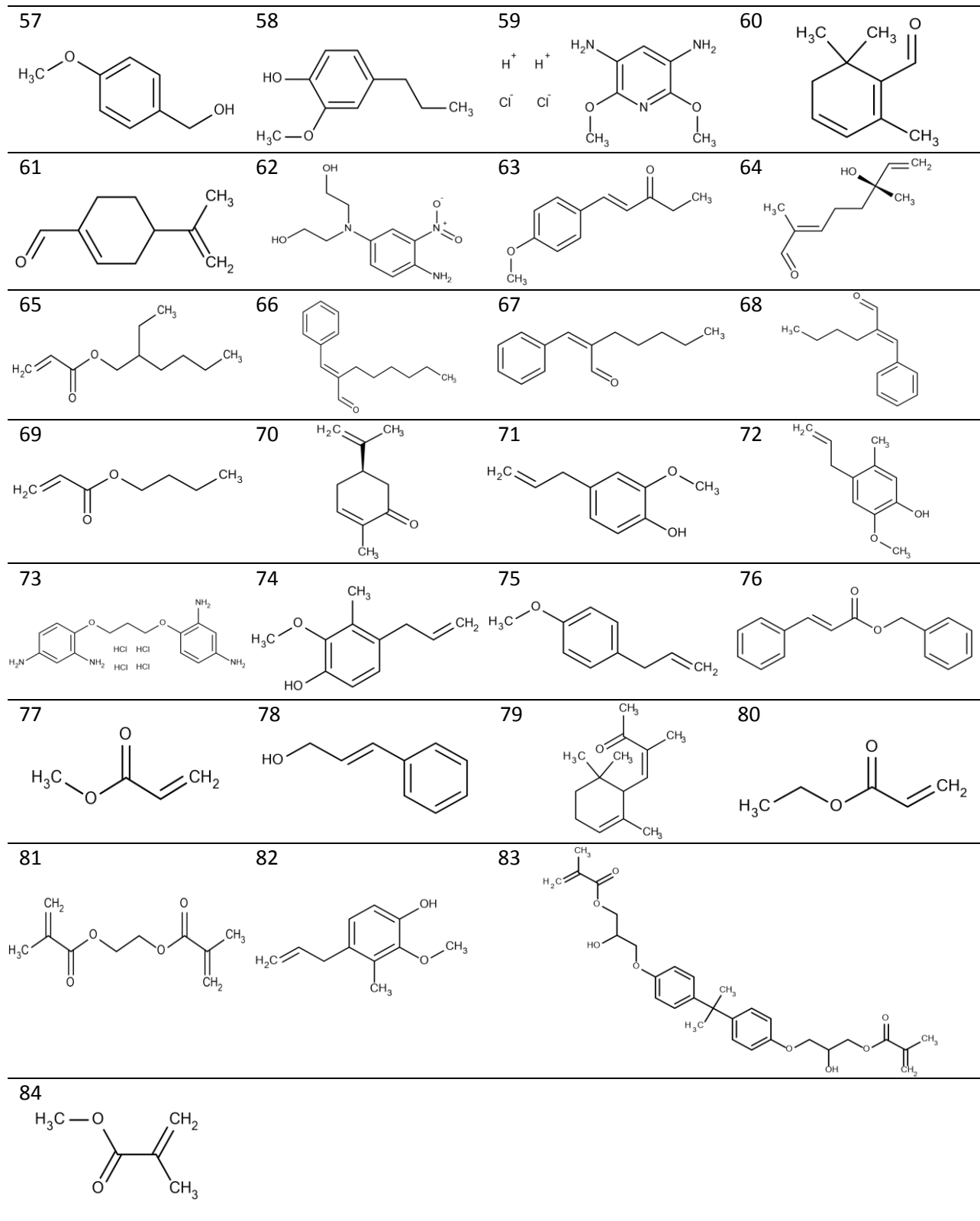
|    |   |             |      |     |      |
|----|---|-------------|------|-----|------|
| 62 | 1-amino-2-nitro-4-bis-(2-hüdoksüetüül)-<br>amino-bensool                  | 29705-39-3  | 8.2  |     | [28] |
| 63 | 1-(4-metoksüfenüül)-1-penteen-3-oon                                       | 104-27-8    | 9.3  | Jah | [27] |
| 64 | 6-hüdoksü-2,6-dimetüül-okt-2,7-dienaal                                    | 54664-89-0  | 9.5  | Jah | [28] |
| 65 | 2-etüülheksüülakrülaad  | 103-11-7    | 10   | Jah | [28] |
| 66 | 2-(fenüülmetüleen)-oktanaal   | 101-86-0    | 11   | Jah | [27] |
| 67 | 2-(fenüülmetüleen)-heptanaal  | 122-40-7    | 11   | Jah | [27] |
| 68 | 2-(fenüülmetüleen)-heksanaal  | 7492-44-6   | 11   | Jah | [27] |
| 69 | Butüülakrülaad  | 141-32-2    | 11   | Jah | [28] |
| 70 | 2-metüül-5R-(1-metüületenüül)-tsükloheks-<br>2-eeen-1-oon                 | 6485-40-1   | 12.9 | Jah | [28] |
| 71 | 2-metoksü-4-(prop-2-eeen-1-üül)-fenool                                    | 97-53-0     | 13   |     | [27] |
| 72 | 2-metoksü-5-metüül-4-(prop-2-eeen-1-üül)-<br>fenool                       | 186743-25-9 | 13   |     | [27] |
| 73 | 1,3-bis-(2,4-diaminofenoksü)-propaan x<br>4HCl                            | 74918-21-1  | 14.7 |     | [28] |
| 74 | 2-metoksü-6-metüül-4-(prop-2-eeen-1-üül)-<br>fenool                       | 186743-24-8 | 17   |     | [27] |
| 75 | 1-metoksü-4-(prop-2-eeen-1-üül)-benseen                                   | 140-67-0    | 18   |     | [27] |
| 76 | Fenüülmetüül 3-fenüül-prop-2-enaad  | 103-41-3    | 18.4 | Jah | [28] |
| 77 | metüülakrülaad  | 96-33-3     | 20   | Jah | [28] |
| 78 | 3-fenüül-prop-2-eeen-1-ool  | 104-54-1    | 21   |     | [27] |
| 79 | 3-metüül-4-(2,6,6-trimetüül-tsükloheks-2-<br>eeen-1-üül)-but-3-eeen-2-oon | 127-51-5    | 21.8 | Jah | [28] |
| 80 | etüülakrülaad   | 140-88-5    | 28   | Jah | [27] |
| 81 | Etüleenglükooldimetakrülaad   | 97-90-5     | 28   |     | [27] |
| 82 | 2-metoksü-3-metüül-4-(prop-2-eeen-1-üül)-<br>fenool                       | 186743-26-0 | 32   |     | [27] |

|    |  |           |    |          |
|----|--|-----------|----|----------|
| 83 | 1,1'-[(1-metüületülideen)bis[4,1-fenüleenoksü(2-hüdroksü-3,1-propaandiüül)]2-metüül-prop-2-enaat | 1565-94-2 | 45 | [28]     |
| 84 | metüül metakrülaat   | 80-62-6   | 90 | Jah [28] |

**Tabel L2.** Kirjandusest leitud polaarse (elektrofiilse) kaksiksidemega ühendite struktuurid.







**Tabel L3.** Mudel 2 (fragmendideskriptoriga mudel) ja mudel 3 ( $\beta$ -süsiniku aatomideskriptoriga mudel): prognoositud pEC3 väärtused ja nende erinevus eksperimentaalse pEC3ga.

| ID | pEC3  | 2. mudel   |          | 3. mudel   |          |
|----|-------|------------|----------|------------|----------|
|    |       | prog. pEC3 | erinevus | prog. pEC3 | erinevus |
| 17 | 2.535 | 2.058      | -0.477   | 2.058      | -0.477   |
| 25 | 2.541 | 1.996      | -0.545   | 2.031      | -0.510   |
| 30 | 1.919 | 1.373      | -0.546   | 1.380      | -0.539   |
| 34 | 1.846 | 1.645      | -0.201   | 1.594      | -0.252   |
| 38 | 1.790 | 1.592      | -0.198   | 1.538      | -0.253   |
| 39 | 1.828 | 2.119      | 0.291    | 2.114      | 0.286    |
| 40 | 1.920 | 1.645      | -0.275   | 1.631      | -0.289   |
| 41 | 1.644 | 1.841      | 0.197    | 1.903      | 0.259    |
| 43 | 1.734 | 1.665      | -0.069   | 1.693      | -0.041   |
| 45 | 1.439 | 1.681      | 0.242    | 1.776      | 0.337    |
| 48 | 1.597 | 1.682      | 0.085    | 1.715      | 0.119    |
| 49 | 1.440 | 1.684      | 0.244    | 1.752      | 0.312    |
| 51 | 1.631 | 1.409      | -0.222   | 1.360      | -0.271   |
| 52 | 1.512 | 1.407      | -0.104   | 1.402      | -0.110   |
| 53 | 1.252 | 1.348      | 0.097    | 1.393      | 0.142    |
| 56 | 1.473 | 1.678      | 0.205    | 1.694      | 0.222    |
| 60 | 1.302 | 1.191      | -0.111   | 1.176      | -0.126   |
| 61 | 1.268 | 1.372      | 0.104    | 1.347      | 0.079    |
| 63 | 1.311 | 1.118      | -0.193   | 1.120      | -0.191   |
| 64 | 1.248 | 1.742      | 0.494    | 1.657      | 0.409    |
| 65 | 1.266 | 1.178      | -0.088   | 1.157      | -0.108   |
| 66 | 1.294 | 1.384      | 0.091    | 1.355      | 0.061    |
| 67 | 1.265 | 1.305      | 0.041    | 1.264      | -0.001   |
| 68 | 1.233 | 1.372      | 0.138    | 1.348      | 0.115    |
| 69 | 1.066 | 0.778      | -0.288   | 0.718      | -0.348   |
| 70 | 1.066 | 1.360      | 0.294    | 1.291      | 0.225    |
| 76 | 1.112 | 1.421      | 0.309    | 1.421      | 0.308    |
| 77 | 0.634 | 0.383      | -0.251   | 0.515      | -0.119   |
| 79 | 0.976 | 1.335      | 0.359    | 1.330      | 0.354    |
| 80 | 0.553 | 0.554      | 0.001    | 0.591      | 0.038    |
| 84 | 0.046 | 0.424      | 0.378    | 0.420      | 0.373    |