



Univerzitet u Novom Sadu
TEHNOLOŠKI FAKULTET NOVI SAD
Katedra za primenjene i inženjerske hemije

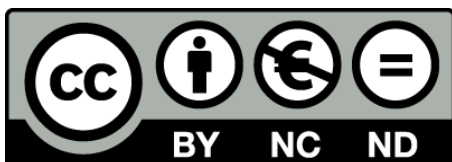
Strahinja Z. Kovačević, mast. inž. tehnol.

DOKTORSKA DISERTACIJA

**HEMOMETRIJSKO MODELOVANJE HROMATOGRAFSKOG
PONAŠANJA I BIOLOŠKE AKTIVNOSTI SERIJE
ANDROSTANSKIH DERIVATA**

Mentor: Dr Sanja Podunavac-Kuzmanović, redovni profesor

Novi Sad, 2015.



© COPYRIGHT BY STRAHINJA KOVAČEVIĆ, 2015

Ovaj rad je licenciran Creative Commons **Autorstvo-Nekomercijalno-Bez prerade** 4.0 međunarodnom licencom.

Autorstvo - Mora se navesti odgovarajuća zasluga, postaviti link do licence i ukazati na napravljene izmene u materijalu, što se može izvesti na bilo koji razuman način, ali ne na način koji sugeriše da davalac licence odobrava izmene ili upotrebu izmenjenog materijala.

Nekomercijalno - Materijal se ne može koristiti u komercijalne svrhe.

Bez prerade - U slučaju izmene i/ili nadogradnje, zabranjena je distribucija modifikovanog materijala.



University of Novi Sad
FACULTY OF TECHNOLOGY NOVI SAD
Department of Applied and Engineering Chemistry

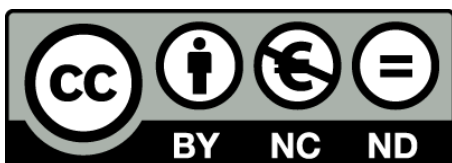
MSc Strahinja Z. Kovačević

DOCTORAL DISSERTATION

**CHEMOMETRIC MODELING OF CHROMATOGRAPHIC
BEHAVIOR AND BIOLOGICAL ACTIVITY OF A SERIES OF
ANDROSTANE DERIVATIVES**

Mentor: Dr. Sanja Podunavac-Kuzmanović, full professor

Novi Sad, 2015



© COPYRIGHT BY STRAHINJA KOVAČEVIĆ, 2015

This work is licensed under a Creative Commons **Attribution-NonCommercial-NoDerivatives** 4.0 International License.

Attribution - You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial - You may not use the material for commercial purposes.

NoDerivatives - If you remix, transform, or build upon the material, you may not distribute the modified material.



SADRŽAJ

REZIME (ABSTRACT)	4
SPISAK SKRAĆENICA	6
1. UVOD	9
2. TEORIJSKI DEO SA PREGLEDOM LITERATURE	11
2.1. Hemijska struktura steroidnih jedinjenja.....	11
2.2. Hromatografija steroidnih jedinjenja.....	13
2.3. Biološka aktivnost steroidnih jedinjenja	15
2.4. 17-Pikolil i 17-pikoliniliden androstanski derivati.....	19
2.5. Računarsko modelovanje molekula	20
2.6. Osnovne hemometrijske metode	27
2.6.1. Hemometrija kao multidisciplinarni naučni pristup.....	27
2.6.2. Osnove QSRR analize i pojam hromatografske lipofilnosti.....	29
2.6.3. Osnove QSAR analize.....	33
2.7. Hemometrijska statistika	38
2.7.1. Klasifikacione hemometrijske tehnike.....	39
2.7.2. Regresione hemometrijske tehnike	42
2.7.3. Formiranje i validacija QSRR i QSAR modela.....	47
3. EKSPERIMENTALNI DEO	51
3.1. Serija ispitivanih androstanskih derivata	51
3.2. Hromatografska analiza androstanskih derivata.....	56
3.3. Biološka aktivnost androstanskih derivata.....	57
3.4. Računarsko modelovanje molekula i hemometrijska obrada podataka.....	59



4. REZULTATI I DISKUSIJA	61
4.1. Hromatografska analiza androstanskih derivata	61
4.2. Računarsko modelovanje strukture androstanskih derivata	65
4.3. Analiza androstanskih derivata klasifikacionim hemometrijskim tehnikama	76
4.4. QSRR analiza retencionog ponašanja androstanskih derivata	83
4.4.1. Linearno QSRR modelovanje	83
4.4.2. Nelinearno QSRR modelovanje.....	103
4.4.3. Rangiranje QSRR modela.....	109
4.5. QSAR analiza antikancerogene aktivnosti androstanskih derivata.....	115
4.5.1. Odabir molekulskih deskriptora kao ulaznih varijabli QSAR modela.....	115
4.5.2. Nelinearno QSAR modelovanje.....	117
4.5.3. Rangiranje QSAR modela	125
5. ZAKLJUČAK (CONCLUSION)	131
6. LITERATURA	135
7. PRILOG	145
BIOGRAFIJA	155
KLJUČNA DOKUMENTACIJSKA INFORMACIJA	157
KEYWORDS DOCUMENTATION	160



ZAHVALNICA

Želim da se zahvalim svima koji su na bilo koji način doprineli realizaciji i kvalitetu ove doktorske disertacije.

*Neizmernu zahvalnost na ukazanom poverenju, dragocenim savetima, razumevanju i bezuslovnoj podršci tokom svih godina naše saradnje upućujem svom mentoru, **dr Sanji Podunavac-Kuzmanović**, redovnom profesoru Tehnološkog fakulteta u Novom Sadu.*

*Srdačno se zahvaljujem **dr Lidiji Jevrić**, vanrednom profesoru Tehnološkog fakulteta u Novom Sadu, na svesrdnoj pomoći, korisnim savetima i zalaganju tokom izrade ove doktorske disertacije i tokom studija.*

*Zahvalnost upućujem i **dr Slobodanu Gadžuriću**, vanrednom profesoru Prirodno-matematičkog fakulteta u Novom Sadu, na ukazanom poverenju i podršci, kao i zalaganju za moj angažman na projektu OI172012.*

*Zahvaljujem se **dr Radimiru Malbaši**, redovnom profesoru Tehnološkog fakulteta u Novom Sadu, na saradnji i interesovanju za ovu doktorsku disertaciju.*

*Posebnu zahvalnost upućujem **dr Evgeniji Đurendić**, redovnom profesoru Prirodno-matematičkog fakulteta u Novom Sadu, i **dr Jovani Ajduković**, docentu Prirodno-matematičkog fakulteta u Novom Sadu, na sintezi androstanskih derivata koji su predmet ispitivanja u ovoj doktorskoj disertaciji.*

*Zahvalnost upućujem i **dr Pavlu Jovanovu**, rukovodiocu laboratorije FINSLab Naučnog instituta za prehrambene tehnologije u Novom Sadu na pomoći pri HPLC analizi uzoraka.*

Najveće hvala upućujem svojoj porodici na ljubavi, razumevanju i nesebičnoj podršci.

April, 2015.

Strahinja



REZIME

Steroidna jedinjenja spadaju u grupu supstanci sa širokim spektrom biološkog delovanja i predstavljaju dobru polaznu osnovu za sintezu mnogih derivata sa željenim biološkim potencijalom. Organskim sintezama se došlo do velikog broja steroidnih derivata, od kojih su neki pokazali značajnu biološku aktivnost, kao što je citotoksičnost prema različitim ćelijskim linijama kancera. Karakterizacija novosintetisanih jedinjenja može se izvesti eksperimentalnim i računarskim (*in silico*) metodama. U ovoj doktorskoj disertaciji predstavljeno je eksperimentalno određivanje lipofilnosti 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata primenom visokopritisne tačne hromatografije na obrnutim fazama, a potom hemometrijska analiza hromatografskog ponašanja (hromatografske lipofilnosti) QSRR pristupom. Hemijska struktura analiziranih derivata opisana je numerički, pomoću izračunatih molekulskih deskriptora. U drugom delu doktorske disertacija predstavljena je QSAR analiza citotoksične aktivnosti 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata prema ćelijama androgen-receptor negativnog kancera prostate (AR-neg. PC-3). Odabir najkvalitetnijih QSRR i QSAR modela obavljen je na osnovu izračunatih statističkih parametara, a njihovo rangiranje izvedeno je primenom metode sume razlika rangova (SRD). Pored regresionih QSRR i QSAR hemometrijskih metoda, primenjene su klaster analiza i analiza glavnih komponenta sa ciljem utvrđivanja sličnosti ili razlika između analiziranih derivata na osnovu izračunatih molekulskih deskriptora.

Ključne reči: 17-Pikolil i 17-pikoliniliden androstanski derivati, Hemometrija, QSRR analiza, QSAR analiza, Hromatografska analiza, Antikancerogena aktivnost.



ABSTRACT

Steroidal compounds belong to the group of substances with wide spectrum of biological activity and represent the basic material for synthesis of many derivatives with preferred biological potential. A grate number of steroidal derivatives have been obtained through organic syntheses, many of which have demonstrated significant biological activity, such as cytotoxicity toward various cancer cell lines. Characterization of newly synthesized compounds can be achieved experimentally and by computational approach (*in silico*). This doctoral dissertation describes experimental determination of lipophilicity of 17 α -picolyl and 17(*E*)-picolinylidene androstane derivatives applying reversed-phase high pressure liquid chromatography followed by QSRR chemometric analysis of chromatographic behaviour (chromatographic lipophilicity). Chemical structure of the analyzed derivatives was described numerically by *in silico* molecular descriptors. The second part of this dissertation describes QSAR analysis of cytotoxic activity of 17 α -picolyl and 17(*E*)-picolinylidene androstane derivatives toward androgen-receptor negative prostate cancer cell line (AR-neg. PC-3). Selection of the best QSRR and QSAR models was carried out based on their statistical parameters, and their ranking was done by sum of ranking differences (SRD) method. Besides the regression QSRR and QSAR chemometric methods, cluster analysis and principal components analysis were conducted in order to reveal possible similarities and dissimilarities among the studied derivatives on the basis of calculated molecular descriptors.

Keywords: 17-Picolyl and 17-picolinylidene androstane derivatives; Chemometrics; QSRR analysis; QSAR analysis; Chromatographic analysis; Anticancer activity.



SPIŠAK SKRAĆENICA

ADMET	Apsorpcija, distribucija, metabolizam, ekskrecija, toksičnost
ANN	Veštačka neuronska mreža
CA	Klaster analiza
CRRN	Poređenje rangova sa nasumičnim brojevima
CV	Unakrsna validacija
FF	Polje sila
GSA	Analiza globalne osetljivosti
HOMO	Najviša zauzeta (okupirana) molekulska orbitala
LR	Linearna regresija
LSER	Linearna korelacija slobodne energije solvatacije
LUMO	Najniža slobodna (prazna) molekulska orbitala
LV	Latentna varijabla
MLP	Višeslojna neuronska mreža
MLR	Multipla linearna regresija
MM	Molekulska mehanika
PCA	Analiza glavnih komponentata
PCR	Regresija glavnih komponentata
PLS	Regresija metodom parcijalnih najmanjih kvadrata
PR	Polinomska regresija
QSAR	Kvantitativna veza između strukture i aktivnosti
QSPR	Kvantitativna veza između strukture i fizičko-hemijske osobine
QSRR	Kvantitativna veza između strukture i retencije
RP-HPLC	Tečna hromatografija visokih performansi na obrnutim fazama
SRD	Suma razlike rangova
SS	Stepenasta selekcija (regresija)
TLC	Hromatografija na tankom sloju
VIF	Faktor inflacije varijanse
VIP	Značaj varijable u projekciji



SKRAĆENICE I DEFINICIJE MOLEKULSKIH DESKRIPTORA KORIŠĆENIH PRI CA, PCA, QSRR I QSAR MODELOVANJU

AClogP	Deskriptor lipofilnosti
ALOGP	Deskriptor lipofilnosti
ALOGPs	Deskriptor lipofilnosti
AMR	<i>Ghose-Crippen</i> molarna refraktivnost
ASP	Asferičnost
Average logP	Prosečna vrednost deskriptora lipofilnosti
BI	Balaban indeks
BP	Temperatura ključanja
CAA	Konolijeva (<i>Connolly</i>) dostupna oblast
CC	Klasterski broj
CMA	Konolijeva (<i>Connolly</i>) molekulska oblast
CP	Kritični pritisak
CSEV	Konolijeva (<i>Connolly</i>) zapremina nedostupna rastvaraču
CT	Kritična temperature
CV	Kritična zapremina
DE	<i>Dreiding</i> -ova energija
Dipole	Dipolni moment
EEM_F1	Prva komponenta autokorelacionog vektora sigma <i>Fukui</i> indeksa
E_{GAP}	Energetska praznina između HOMO i LUMO molekulskih orbitala
E_{HOMO}	Energija najviše zauzete (okupirane) molekulske orbitale
E_{max}	Maksimalna vrednost elektrostatičkog potencijala
E_{sr}	Srednja vrednost elektrostatičkog potencijala
FPSA3	Deskriptor delimično naelektrisane pozitivne površine molekula tip 3
GFE	Gibsova slobodna energija
GMTIV	Gutmanov molekulski topološki indeks (<i>MTI by valence vertex degrees</i>)
H3D	3D- <i>Harary</i> indeks
HF	Toplota formiranja
HLC	Henrijeva konstanta
IGTC	Toplotni kapacitet idealnog gasa
J3D	3D-Balaban indeks
Jhete	Balaban indeks izveden iz ponderisane matrice udaljenosti elektronegativnosti
Jhetm	Balaban indeks izveden iz ponderisane matrice udaljenosti mase
Jhetp	Balaban indeks izveden iz ponderisane matrice udaljenosti polarizabilnosti
Jhetv	Balaban indeks izveden iz van der Valsove ponderisane matrice udaljenosti
KOWWIN	Deskriptor lipofilnosti
$\log P_{klop}$	Deskriptor lipofilnosti
$\log P_{phys}$	Deskriptor lipofilnosti
$\log P_{vg}$	Deskriptor lipofilnosti
$\log P_{wgt}$	Deskriptor lipofilnosti
MAXDN	Maksimalna elektrotopološka negativna varijacija
MAXDP	Maksimalna elektrotopološka pozitivna varijacija
MDCK	Deskriptor permeabilnosti <i>Madin-Darby</i> epitelnih ćelija psećeg bubrega
miLogP	Deskriptor lipofilnosti



MLOGP	Deskriptor lipofilnosti
MMFF94	<i>Molecular Mechanics Force Field</i> energija
MP	Temperatura topljenja
MR	Molarna refraktivnost
Mr	Molarna masa
MTI	Molekulski topološki indeks
NPA_AQc	Suma apsolutnih vrednosti procenjenih NPA (<i>Natural Population Analysis</i>) parcijalnih atomskih naelektrisanja na ugljenikovim atomima
NPA_Q5	Peta komponenta autokorelacionog vektora procenjenih parcijalnih NPA (<i>Natural Population Analysis</i>) atomskih naelektrisanja
Ov	Ovalnost
PI	Polarizabilnost
Pi_Q3	Treća komponenta autokorelacionog vektora <i>Hückel</i> -ovog pi-naelektrisanja
PPSA1	Delimično pozitivna površina tipa 1
PSA	Polarna površina
Rgyr	Radius žiracije
S+FaSSIF	Rastvorljivost u simuliranoj intestinalnoj tečnosti u stanju izgladnelosti (mg/ml)
S+HIVI_ST	pIC ₅₀ za inhibiciju HIV-1 integreze (<i>Simulation Plus Software</i>)
S+Pskin	Deskriptor kožne propustljivosti (<i>Simulation Plus Software</i>)
S+SF	Odnos ravnotežne rastvorljivosti jonizabilne soli ispitivanog jedinjenja i njene intrinzične rastvorljivosti
S1K	1- <i>Path Kier</i> alfa-modifikovan indeks oblika
S2K	2- <i>Path Kier</i> alfa-modifikovan indeks oblika
S3K	3- <i>Path Kier</i> alfa-modifikovan indeks oblika
SASA1.4	Površina molekula dostupna rastvaraču (prečnik molekula rastvarača iznosi 1.4 Å)
SCAA3	Ponderisana naelektrisana površina na akceptorskim atomima tipa 3
SMTIV	Šulcov (<i>Schultz</i>) molekulski topološki indeks (MTI by valence vertex degrees)
SP	Deskriptor kožne propustljivosti (<i>PreADMET Software</i>)
SPH	Sferičnost
TE	Ukupna energija
TIE	<i>E-state</i> topološki parametar
TOX_BCF	Biokoncentracioni faktor – podeoni koeficijent između ribljeg tkiva i okolne vodene sredine u ravnotežnom stanju (C_{fish}/C_{water})
vdWSA	Van der Valsova molekulska površina
vdWV	Van der Valsova molekulska zapremina
Vol	Zapremina molekula
Whete	Vinerov (<i>Wiener</i>) indeks izveden iz ponderisane matrice udaljenosti elektronegativnosti
Whetm	Vinerov (<i>Wiener</i>) indeks izveden iz ponderisane matrice udaljenosti mase
Whetp	Vinerov (<i>Wiener</i>) indeks izveden iz ponderisane matrice udaljenosti polarizabilnosti
Whetv	Vinerov (<i>Wiener</i>) indeks izveden iz van der Valsove ponderisane matrice udaljenosti
Whetz	Vinerov (<i>Wiener</i>) indeks izveden iz Z-ponderisane matrice (<i>Barysz</i> -ova matrica)
WI	Vinerov (<i>Wiener</i>) indeks
XLOGP2	Deskriptor lipofilnosti
XLOGP3	Deskriptor lipofilnosti
ΔE	Razlika između maksimalne i minimalne vrednosti elektrostatičkog potencijala



1. Uvod

Steroidna jedinjenja spadaju u grupu veoma važnih prirodnih proizvoda sa tetracikličnim jezgrom u strukturnoj osnovi. Steroidi (grč. *stereos* – čvrsto i *oeides* – slično), kao i veliki broj njihovih derivata, ispoljavaju širok spektar biološke aktivnosti, te su predmet ispitivanja mnogih istraživačkih grupa u svetu i kod nas. Steroidi su dosta izučavani još od izolacije holesterola u XIX veku, a od početka XX veka dodeljen je značajan broj Nobelovih nagrada za hemiju i medicinu za istraživanja koja su se bavila steroidnim jedinjenjima, uključujući žučne kiseline, sterole i steroidne hormone. Budući da je biološka uloga velikog broja prirodnih steroida biljnog i životinjskog porekla već ustanovljena (regulacija biohemijskih procesa u organizmu, uticaj na rast i razvoj organizma, polni razvoj i plodnost), započeta je laboratorijska sinteza različitih steroidnih derivata koji su u nekim slučajevima ispoljavali pojačanu biološku aktivnost u odnosu na polazna jedinjenja. Potrebno je istaći da čak i male strukturne modifikacije mogu imati izrazit uticaj na promenu biološkog potencijala steroidnog jedinjenja. Laboratorijskim *in vitro* i *in vivo* ispitivanjima došlo se do saznanja o antikancerogenoj (citotoksičnoj), antibakterijskoj, antifungalnoj i antiviralnoj aktivnosti nekih steroidnih derivata. Nakon dugogodišnjih medicinskih ispitivanja, započeta je njihova primena u terapiji različitih oboljenja, kao što su alergije, artritis, kancer i dr.

U grupu novosintetisanih steroidnih jedinjenja spadaju i 17-pikolil i 17-pikoliniliden androstanski derivati, koji su predmet ispitivanja ove doktorske disertacije. U ranije objavljenim istraživanjima utvrđena je njihova značajna antiproliferativna aktivnost prema nekoliko vrsta ćelijskih linija kancera, što ih izdvaja kao potencijalne terapeutike za tretman ovih oboljenja. S obzirom na to da su u pitanju novi androstanski derivati, svaki doprinos njihovoj karakterizaciji u pogledu fizičko-hemijskih osobina i predviđanju njihovog biološkog potencijala je dragocen.

Hemometrija, kao multidisciplinarna naučna oblast, postala je neizostavni deo analize podataka iz oblasti hemije, tehnologije, medicine, farmacije, mikrobiologije i mnogih drugih oblasti. Jedan od glavnih činilaca ovog pristupa su statističke metode primenjene sa ciljem klasifikacije i diskriminacije objekata (u ovom slučaju hemijskih jedinjenja) ili utvrđivanja korelacije između zadatih varijabli. Rezultati koje hemometrija pruža mogu da eliminišu faktor „probe i greške“ i tako omogućće uštedu vremena i smanjenje finansijskih troškova, pogotovo u slučaju dizajniranja novih lekova. Hemometrijskom analizom i računarskim modelovanjem molekula može se steći uvid u strukturne osobine jedinjenja koje mogu biti presudne za njegov biološki potencijal.



S obzirom na to da je biološki potencijal, u smislu antiproliferativne aktivnosti, 17-pikolil i 17-pikoliniliden androstanskih derivata prema određenim ćelijskim linijama kancera već ustanovljen, značaj ovog istraživanja ogleda se u formiranju matematičkih modela koji bi imali sposobnost predviđanja antikancerogene aktivnosti novih, još uvek nesintetisanih, androstanskih derivata sa izraženijom antiproliferativnom aktivnošću prema ćelijama androgen-receptor negativnog kancera prostate (AR-neg. PC-3). Budući da je lipofilnost jedan od najbitnijih fizičko-hemijskih parametara biološki aktivnih jedinjenja, definisanje lipofilnosti 17-pikolil i 17-pikoliniliden androstanskih derivata hromatografskom tehnikom, kao i predviđanje lipofilnosti novih androstanskih derivata sa potencijalnom biološkom aktivnošću, svakako dobijaju na značaju.

Stoga, ciljevi ove doktorske disertacije su: (1) ispitivanje ponašanja 17-pikolil i 17-pikoliniliden androstanskih derivata u kolonskoj hromatografiji na obrnutim fazama sa ciljem procene njihove hromatografske lipofilnosti, (2) utvrđivanje kvantitativne veze između hromatografske lipofilnosti (retencije) i hemijske strukture hemometrijskim QSRR (engl. *Quantitative Structure-Retention Relationship*) pristupom, (3) utvrđivanje kvantitativne veze između njihove antikancerogene aktivnosti prema ćelijama kancera prostate (AR-neg. PC-3) i hemijske strukture hemometrijskim QSAR (engl. *Quantitative Structure-Activity Relationship*) pristupom, (4) utvrđivanje sličnosti između jedinjenja na osnovu hemijske strukture hemometrijskim klasifikacionim tehnikama i (5) odabir reprezentativnih matematičkih modela za predviđanje hromatografske lipofilnosti (hromatografskog ponašanja) i antikancerogene aktivnosti ispitivanih androstanskih derivata. Hemijska struktura jedinjenja opisana je numerički, molekulskim deskriptorima, dobijenim pomoću sofisticiranih računarskih programa za molekulski dizajn.

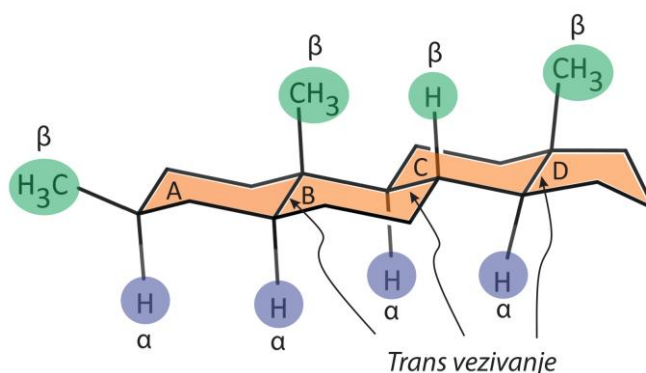
Ostvarivanje postavljenih ciljeva trebalo bi da dovede do značajnih smernica koje bi imale primenu kod budućih sinteza strukturno sličnih androstanskih derivata sa povećanom antikancerogenom aktivnošću prema ćelijama kancera prostate.



2. Teorijski deo sa pregledom literature

2.1. Hemijska struktura steroidnih jedinjenja

Glavna strukturna karakteristika steroida je perhidrociklopentanofenantrenski sistem prstenova, tzv. gonan. Ovaj sistem od četiri zasićena prstena (A, B, C i D) često sadrži i supstituente koji bitno utiču na hemijsko i biološko ponašanje molekula. S obzirom na tetraedarsku strukturu jednostrukih veza između atoma ugljenika, ovaj sistem prstenova nije u potpunosti planaran. Cikloheksanski prstenovi (A, B i C) u steroidnom jezgru uglavnom zauzimaju konformaciju stolice, dok se ciklopentanski prsten (D) nalazi u konformaciji koverta. Susedni prstenovi se mogu nalaziti približno u istoj ravni (*trans* pozicija) ili pod određenim uglom (*cis* pozicija) u odnosu jedan na drugi. *Trans* vezivanjem prstenova u konformaciji stolice postiže se najmanji napon svih prstenova, pri čemu su supstituenti ili vodonikovi atomi na svim čvorovima u aksijalnom položaju, kako je prikazano na slici 2.1. Atomi ili supstituenti koji se nalaze ispod ravni steroidnog jezgra obeležavaju se slovom α , dok oni iznad ravni slovom β .



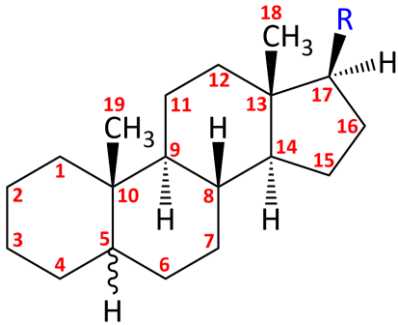
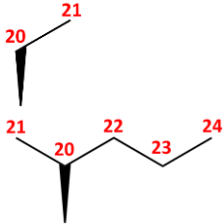
Slika 2.1. Struktura steroidnog jezgra sa metil supstuentima u položajima 3, 10 i 13

Položaj i priroda supstuenata često imaju presudan uticaj na ukupnu lipofilnost molekula, a samim tim i na njegovo ponašanje u biološkim i hromatografskim sistemima. Steroidi se smatraju relativno nepolarnim (lipofilnim ili hidrofobnim) jedinjenjima, međutim prisustvo hidrosidnih i okso grupa daje im amfifilni karakter, što je posebno karakteristično za žučne kiseline [1]. Steroidi se svrstavaju u grupu nehidrolizujućih lipida koji ne sadrže glicerol i masne kiseline i, prema tome, ne podležu saponifikaciji pri zagrevanju sa alkalnim hidrosidima [2]. Često su pratioci masnih supstanci sa kojima se izdvajaju prilikom ekstrakcije.



Sistemska nomenklatura definiše nazive steroidnih jedinjenja na osnovu prirode supstituenta primarno u poziciji 17 steroidnog jezgra. Osnovna struktura androstana i nekih njegovih analoga prikazana je u tabeli 2.1.

Tabela 2.1. Osnovna struktura androstana i nekih njegovih analoga

Supstituent (R)	Naziv
	Androstan
H (i H atom u položaju 10)	Estran
	Pregnan
	Holan

Najvažnije grupe steroidnih jedinjenja su *steroli*, *žučne kiseline* i *steroidni hormoni* [1]. Za *sterole* je karakteristično prisustvo hidrosidne grupe u C-3 položaju i jedna ili više dvostrukih veza u B prstenu steroidnog skeleta. Najpoznatiji predstavnik ove grupe steroida je holesterol, koji je bitan prekursor sinteze mnogih drugih biološki važnih steroida. *Žučne kiseline* su steroidna jedinjenja kod kojih su A i B prstenovi povezani u *cis* konformaciji, dok se sve hidrosidne grupe nalaze u α -položaju. Bočni niz na C-17 položaju, koji je najčešće izgrađen od pet ugljenikovih atoma, sadrži karboksilnu grupu. *Žučne kiseline* su krajnji proizvodi metabolizma holesterola, a najvažnije su glicinski i taurinski derivati holne kiseline, dezoksiholne i henodezoksiholne kiseline [3]. *Steroidni hormoni* su grupa lipofilnih signalnih molekula koji imaju važnu ulogu u rastu i razvoju organizma, polnom razvoju i regulaciji metabolizma. U organizmu se formiraju iz holesterola. Za testosteron, kortizol, aldosteron i progesteron karakteristična je okso grupa u C-3 položaju, konjugovana sa dvostrukom vezom između C-4 i C-5 atoma prstena A. Značajne razlike između ovih hormona uočavaju se na prstenu D. Struktura estradiola je karakteristična po prisustvu hidrosidne grupe u C-3 položaju, kao i aromatičnom A prstenu.



2.2. Hromatografija steroidnih jedinjenja

Steroidna jedinjenja su u većini bioloških uzoraka uglavnom zastupljena u veoma niskim koncentracijama, često zajedno sa strukturno sličnim supstancama, te su za njihovu detekciju i kvantifikaciju neophodne analitičke tehnike visoke specifičnosti i osetljivosti. Hromatografske tehnike ispunjavaju oba ova uslova i nalaze široku primenu ne samo u kvalitativnoj i kvantitativnoj analizi steroida, već i u određivanju njihove lipofilnosti. Hromatografsko određivanje lipofilnosti biološki aktivnih jedinjenja je u odnosu na druge eksperimentalne metode (metoda mućkanja, metoda sporog mešanja, pH-metrijska metoda) reproduktivnije, jednostavnije, brže i koristi male količine uzorka [4].

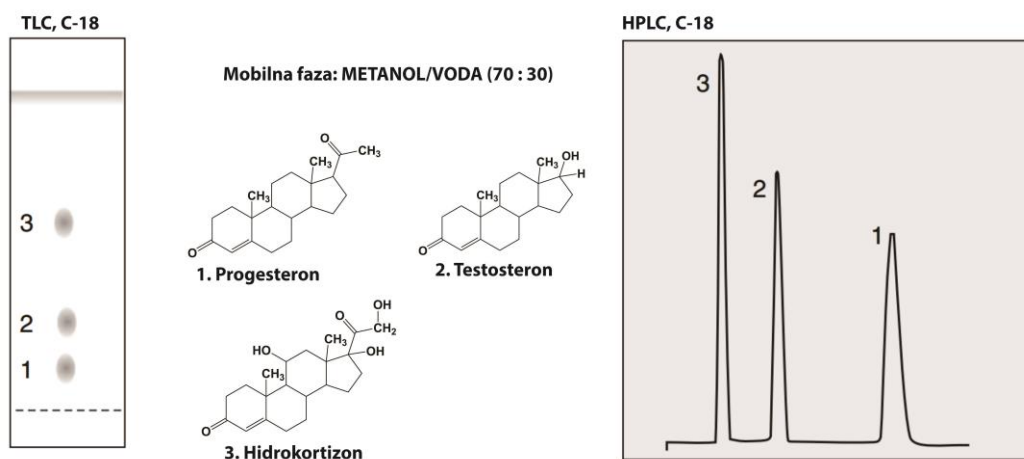
Prisustvo raznolikih funkcionalnih grupa u molekulima steroida značajno utiče na njihovo hromatografsko ponašanje, te je odabir odgovarajućih hromatografskih uslova od velikog značaja za njihovo razdvajanje iz smeše. Jedna od često primenjivanih hromatografskih tehnika za detekciju i separaciju steroida i analizu njihove lipofilnosti je tankoslojna hromatografija (TLC, engl. *Thin-Layer Chromatography*), odnosno tankoslojna hromatografija visokih performansi (HPTLC, engl. *High Performance TLC*). Petrović i saradnici [5] su ispitali retenciono ponašanje derivata estrana i sekoestrana u HPTLC sistemu na C-18 stacionarnoj fazi i mobilnoj fazi metanol/voda pri različitim udelima metanola. Takođe, ispitana je i lipofilnost derivata dehidroepiandrosterona pomoću HPTLC sa C-18 stacionarnom fazom i sistemima rastvarača aceton/voda i dioksan/voda [6]. U novije vreme, izvedena je analiza retencionog ponašanja različitih žučnih kiselina u TLC sistemu sa silika gelom kao stacionarnom fazom i smešom dihlormetan/acetona kao mobilnom fazom, prilikom čega je ispitan i uticaj temperature na njihovu retenciju u primenjenom hromatografskom sistemu [7]. Takođe, HPTLC je našla primenu i kod utvrđivanja lipofilnosti nosintetisanih sekoandrostenskih derivata na normalnim fazama (NP, engl. *Normal Phase*) i obrnutim fazama (RP, engl. *Reversed Phase*) pomoću različitih binarnih mobilnih faza [8]. Osim silika gela i C-18 faze, za razdvajanje steroida u TLC sistemu koriste se i celuloza, aluminijum(III)-oksid, magnezijum-oksid, kao i polarne faze koje sadrže diol, cijano (nitrilo) i amino funkcionalne grupe koje su preko nepolarnog kratkog lanca vezane za silika gel [4,9]. Androstani se uspešno razdvajaju na tankom sloju celuloze impregnirane sa 1,2-propandiol grupama, dok se određene grupe sterola i acetata sterola uspešno razdvajaju na sloju aluminijum(III)-oksida i magnezijum-oksida [9].

Visokopritisna (visokoefikasna) tečna hromatografija (HPLC, engl. *High Pressure (Performance) Liquid Chromatography*) je pored TLC najprimenjenija hromatografska tehnika za razdvajanje steroidnih jedinjenja iz smeše i određivanje njihove lipofilnosti. Kod analize lipofilnosti HPLC tehnikom najčešće se primenjuju obrnute faze (RP-HPLC) [10]. Retenciono ponašanje estrogena i njegovih derivata ispitano je primenom RP-HPLC sa C-18 stacionarnom fazom i mobilnom fazom heksan/tetrahidrofuran/izopropanol, i NP-HPLC sa cijanopropil stacionarnom fazom i mobilnom fazom acetonitril/voda [11]. Ispitivanje lipofilnosti kortikosterona, deksametazona, estrona, estradiola, hidrokortizona, hidrokortizon-21-acetata, progesterona i testosterona obavljeno je pomoću RP-HPLC na amid C-16 stacionarnoj fazi sa metanol/voda i metanol/voda/1-oktanol mobilnim fazama [12]. Slično ispitivanje uticaja 1-oktanol na hromatografsku lipofilnost nekoliko steroidnih hormona sproveli su Benhaim i Grushka [13] koristeći RP-HPLC tehniku sa C-18 stacionarnom fazom. RP-HPLC sa C-18 stacionarnom fazom i mobilnom fazom metanol/fosfatni pufer primenjena je kod definisanja hidrofobnosti i retencionog koeficijenta serije okso



derivata žučnih kiselina [14]. Ista tehnika korišćena je i kod određivanja uticaja temperature na retenciju serije derivata žučnih kiselina [15].

Razdvajanje steroida iz smeše može se postići različitim hromatografskim tehnikama. Na odabir tehnike (metode) analize utiče više faktora, kao što su priroda analita, koncentracija, prisustvo interferencija i dr. Na *slici 2.2.* prikazana je razlika u efikasnosti razdvajanja tri steroida primenom TLC i HPLC sa istim stacionarnim fazama i pri korišćenju iste mobilne faze. Kao što se može uočiti sa slike, razdvajanje testosterona i hidrokortizona je izraženije na TLC ploči, nego na HPLC hromatogramu, dok se obrnut slučaj javlja kod razdvajanja progesterona i testosterona.



Slika 2.2. Poređenje hromatografskog razdvajanja progesterona, testosterona i hidrokortizona na TLC ploči i u HPLC koloni [16]

Pored HPLC i TLC, za razdvajanje i identifikaciju steroida i njihovih derivata u različitim biološkim uzorcima mogu da se koriste tečna hromatografija sa masenom spektrometrijom [17], micelarna elektrokinetička hromatografija [18], hromatografija fluidima u subkritičnom stanju [19], tečna hromatografija ultravisokih performansi sa masenom spektrometrijom [20], gasna hromatografija sa masenom spektrometrijom [21] i druge odgovarajuće analitičke tehnike.



2.3. Biološka aktivnost steroidnih jedinjenja

Biološka uloga mnogih steroidnih jedinjenja u životinjskom i biljnom organizmu do sada je obimno ispitivana. Ustanovljeno je da steroidi učestvuju u izgradnji ćelijskih membrana (holesterol), imaju hormonsku ulogu (testosteron, estrogen, progesteron), pospešuju varenje hrane (soli žučnih kiselina). Organskim sintezama novih steroidnih jedinjenja došlo se do derivata koji imaju širok spektar biološkog dejstva i koji su predmet sadašnjih i budućih ispitivanja. U tabeli 2.2. prikazani su odabrani novosintetisani biološki aktivni steroidi.

Tabela 2.2. Neka novosintetisana steroidna jedinjenja i njihova biološka aktivnost

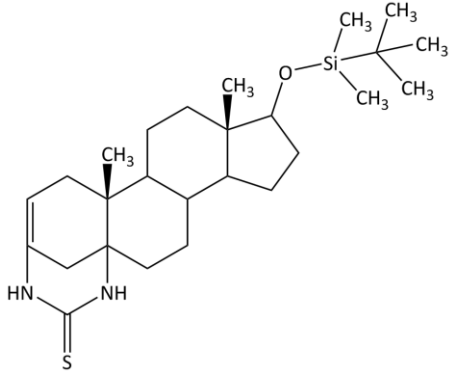
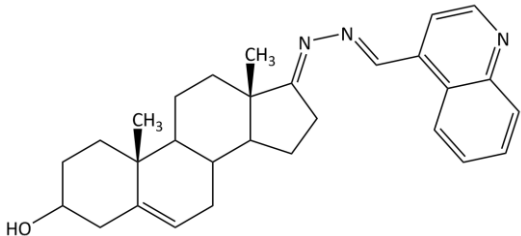
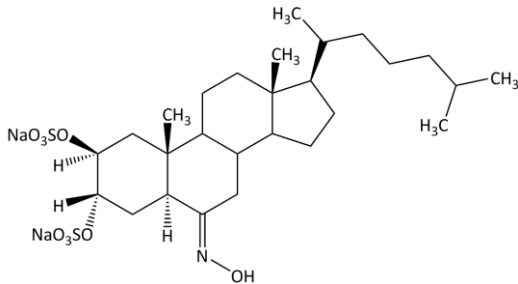
Jedinjenje ili serija jedinjenja (primeri)	Biološka aktivnost	Referenca
 <p>3-(<i>tert</i>-butil-dimetil-silaniloksil)-5β,3α-dimetil-oktahidroindeno[4,5-d]10,12-diazatriciklo[7.3.1.01,6]tridekan-9-tion</p>	Antibakterijska aktivnost <i>in vitro</i> prema <i>Escherichia coli</i> i <i>Vibrio cholerae</i> .	[22]
 <p>Dehidroepiandrosteron-hinolin-4-metanilidenhidrazon</p>	Antiproliferativna aktivnost <i>in vitro</i> prema ćelijama kancera grlića materice (HeLa), debelog creva (HT-29), jetre (Bel 7404) i želuca (SGC 7901).	[23]
 <p>Serija novih 6E-hidroksiminosteroida</p>	Antiproliferativna aktivnost <i>in vitro</i> prema ćelijskim linijama kancera prostate (AR-neg. PC-3 i AR+ LNCaP).	[24]



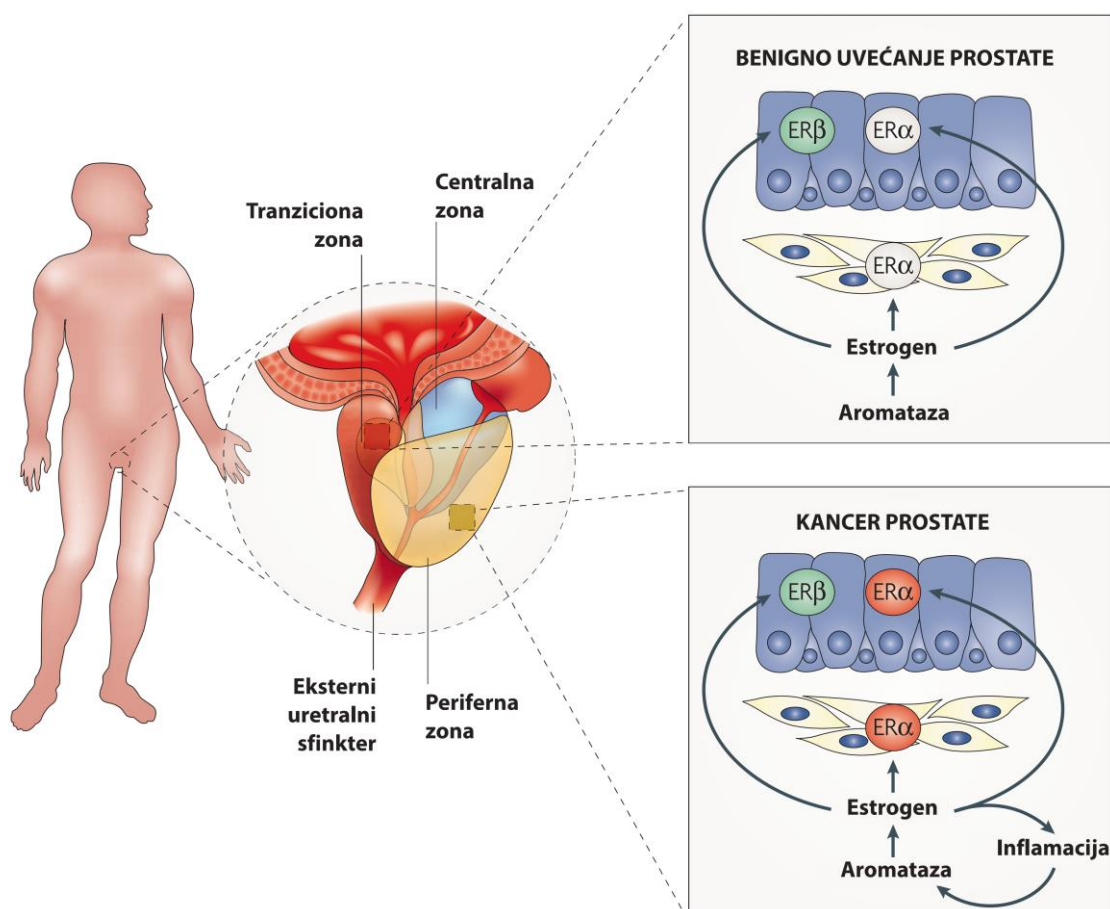
Tabela 2.2. Novosintetisana steroidna jedinjenja i njihova biološka aktivnost (nastavak)

Jedinjenje ili serija jedinjenja (primeri)	Biološka aktivnost	Referenca
<p>A,B-modifikovani D-homo lakton androstanski derivati</p>	Antiproliferativna aktivnost <i>in vitro</i> prema ćelijama kancera prostate (AR-neg. PC-3), debelog creva (HT-29) i grlića materice (HeLa).	[25]
<p>Seriya novih oksisterola izvedenih iz lanosterola</p>	Značajna <i>in vitro</i> antifungalna i umerena antibakterijska aktivnost.	[26]
<p>[1,2,4]triazolo[1,5-a]pirimidin-fenil povezani steroidni dimer</p>	Jaka antiproliferativna aktivnost <i>in vitro</i> prema ćelijama kancera jednjaka (EC-109), želuca (MGC-803), prostate (AR-neg. PC-3), dojke (MCF-7) i jetre (SMMC-7721).	[27]
<p>Seriya novih 17-fluorovanih steroida</p>	Jaka <i>in vivo</i> antiprogestaciona aktivnost, <i>in vitro</i> antitumorska aktivnost prema ćelijama kancera dojke (T47D).	[28]
<p>Seriya novih steroidnih halkona</p>	Baktericidno dejstvo prema <i>Escherichia coli</i> i <i>Bacillus subtilis</i> i fungicidno dejstvo prema <i>Candida albicans</i> i <i>Aspergillus niger</i> (<i>in vitro</i>).	[29]



Na osnovu primera navedenih u *tabeli 2.2.* može se videti da su mogućnosti primene steroidnih derivata u medicini i farmakologiji značajne i da ove supstance predstavljaju veoma dobru osnovu i za QSAR ispitivanja. Mnoga steroidna jedinjenja se u medicini primenjuju i kao antiinflamatorne supstance, anabolici, antitumorski i kontraceptivni agensi [30]. Pored toga, anaboličko-androgeni steroidi su našli primenu kod povećanja muskulatorne mase kod profesionalnih sportista, iako mogu da ispoljavaju brojna neželjena dejstva (infarkt miokarda, srčana oboljenja, arterijska tromboembolija) [31]. Stoga je Svetska antidoping agencija sportistima najstrože zabranila upotrebu steroidnih anabolika, kortikosteroida i antiestrogena. Pored navedenih bioloških efekata, pojedini steroidi, kao što su derivati 5 α -androst-16-ena i androstadiena, ponašaju se i kao aroma komponente mesa divljih svinja, dok mošusni miris 5 α -androst-16-en-3 α -ola doprinosi tipičnoj aromi tartufa [32].

Kancer je trenutno jedan od najčešćih uzroka smrti kod ljudi, sa skoro sedam miliona smrtnih slučajeva godišnje i prognozama da će se do 2020. godine godišnji broj smrtnih slučajeva uzrokovanih kancerom povećati na 10 miliona [33]. Kancer prostate se smatra jednim od češće dijagnostikovanih malignih oboljenja i kao drugi najčešći uzrok smrti od malignih oboljenja kod muškaraca [34]. Androgeni hormoni i androgeni receptori su esencijalni za razvoj prostate, međutim androgeni igraju i veoma važnu ulogu kod razvoja kancera prostate [35]. Pokazalo se da je za uznapredovali kancer prostate hemijska ili hirurška kastracija, radi supresije delovanja androgena, vrlo efikasna terapija [36].

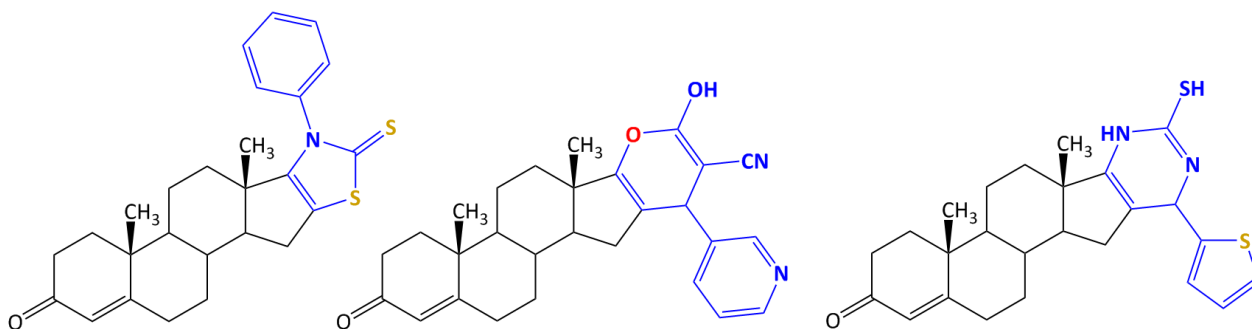


Slika 2.3. Mogući mehanizmi razvoja kancera i benignog uvećanja prostate kod ljudi [34]



Pored utvrđenog uticaja androgena na razvoj kancera prostate, postoje pretpostavke o direktnoj ili indirektnoj ulozi estrogena u pojavi ovog oboljenja i benignom uvećanju prostate (BPH) [34]. Mehanizmi uticaja estrogena na ove procese u ljudskom organizmu prikazani su na *slici 2.3*. Kancer prostate se tipično javlja u periferalnoj zoni, dok je BPH karakteristična za tranzicionu zonu prostate. Aktivacija lokalnog estrogen-receptora ER β , bilo sistemski ili preko lokalno aromatizovanih estrogena, neophodna je da bi se sprečio početak BPH, dok ER α nema jasnu ulogu u razvoju ovog oboljenja. Kod kancera prostate, aktivacija ER β receptora može sprečiti razvoj bolesti, dok se aktivacijom ER α postiže suprotan efekat [34]. Aktivacija ER α receptora može izazvati inflamaciju koja stimuliše aktivnost aromataze dovodeći do dalje aktivacije ovog receptora, a samim tim i do razvoja kancera. Iako su neophodni činioci razvoja kancera prostate, androgeni sami po sebi nisu dovoljni za njegovu pojavu. Ovo maligno oboljenje i BPH su česti kod starijih muškaraca, kod kojih se nivo testosterona sa godinama smanjuje, dok nivo estrogena ostaje nepromenjen ili raste. Stoga, promena odnosa koncentracije ova dva steroidna hormona dovodi se u vezu sa početkom razvoja kancera prostate i BPH [34].

Mnogim ispitivanjima zavisnosti strukture od aktivnosti steroidnih jedinjenja došlo se do zaključka da heterociklični sistem (piridin, pirimidin, imidazol, pirazol, tiazol i dr.) povezan sa steroidnim jezgrom preko A ili D prstena može da igra važnu ulogu u biološkoj aktivnosti i da predstavlja molekularni okvir za razvoj farmaceutskih agenasa široke primene [37]. Mnoga od ovih jedinjenja su našla primenu u terapiji kancera, kao npr. abirateron u lečenju kancera prostate. Novijim sintezama [38] se došlo do derivata androstendiona sa heterocikličnim proširenjem u vidu pirana, pirimidina i tiazola sa optimalnom *in vitro* citotoksičnom aktivnošću prema kanceru želuca (NUGC), debelog creva (DLD1), jetre (HA22T i HEPG2), dojke (MCF) i nazofarinksa (HONE1). Strukture pojedinih jedinjenja iz ove serije prikazane su na *slici 2.4*.



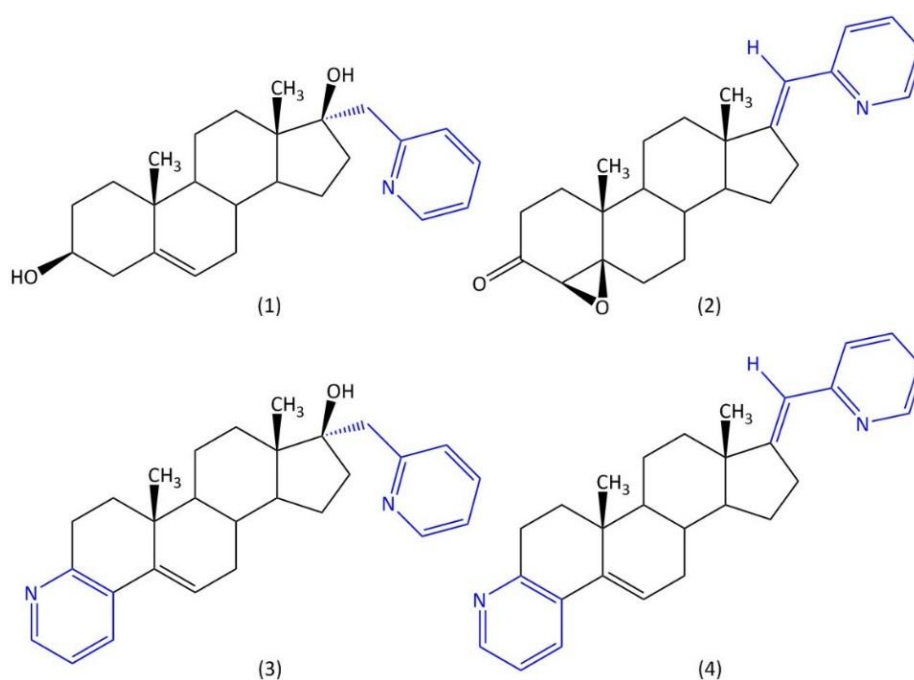
Slika 2.4. Novosintetisani citotoksični steroidni derivati sa različitim heterocikličnim sistemima na D-prstenu [38]

U grupu derivata sa heterocikličnim sistemom na D-prstenu steroidnog jezgra spadaju i 17-pikolil i 17-pikoliniliden derivati androstana [37]. Njihov citotoksični potencijal je testiran na nekoliko vrsta ćelijskih linija kancera, prema kojima su neki od derivata ove serije pokazali značajnu antiproliferativnu aktivnost [37,39-44].



2.4. 17-Pikolil i 17-pikoliniliden androstanski derivati

S obzirom na to da je utvrđeno da se uvođenjem heterocikličnih sistema u steroidno jezgro poboljšava biološki potencijal jedinjenja, sintetisana je nova serija androstanskih derivata koji sadrže 17-pikolil i 17-pikoliniliden supstituente na prstenu D i piridinski sistem vezan za prsten A steroidnog jezgra [37,39-44]. Osim ovih modifikacija, u steroidno jezgro su uvođene i acetoksi, hidroksi, keto, hidroksimetilen, hidroksimino, dihidroksimino funkcionalne grupe, kao i A-homo laktamska funkcija, A-seko sistem i 3-en, 3,5-dien i 4,6-dienska funkcija [37]. Strukture odabranih predstavnika 17-pikoli i 17-pikoliniliden androstanskih derivata date su na slici 2.5.



Slika 2.5. Strukture odabranih novosintetisanih 17-pikolil i 17-pikoliniliden androstanskih derivata: **(1)** 17 α -pikolil-androst-5-en-3 β ,17 β -diol, **(2)** 4 β ,5 β -epoksi-17(*E*)-pikoliniliden-androstan-3-on, **(3)** 17 α -pikolil-piridino[2',3':3,4]-androst-5-en-17 β -ol, **(4)** 17(*E*)-pikoliniliden-piridino[2',3':3,4]-androst-5-en

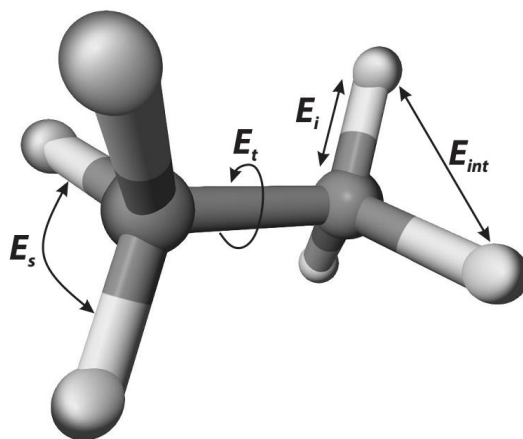
Antiproliferativna aktivnost 17-pikolil i 17-pikoliniliden androstanskih derivata testirana je *in vitro* prema humanim ćelijskim linijama kancera prostate (AR-neg. PC-3), grlića materice (HeLa), debelog creva (HT-29), pluća (A549), dojke (ER+ MCF-7 i ER-neg. MDA-MB-231) i prema zdravoj ćelijskoj liniji [37,39-44]. Rezultati ukazuju na njihov značajan antikancerogeni potencijal. Takođe, sprovedena je doking analiza ovih derivata kojom se predviđa njihov afinitet prema vezivanju za 17 α -hidroksilazu (CYP17A1) i koja istovremeno sugeriše da 17(*E*)-pikoliniliden androstanski derivati mogu biti dobra polazna osnova za razvoj novih supstanci za tretman kancera prostate [44]. Doking analiza često se koristi za predviđanje afiniteta i aktivnosti ispitivanog molekula (liganda). Računarsko (*in silico*) molekulska modelovanje može da pruži značajan uvid u različite fizičko-hemijske osobine jedinjenja, te nalazi široku primenu u QSAR analizama.



2.5. Računarsko modelovanje molekula

Teorijska hemija je grana hemije koja primenjuje matematičke metode u kombinaciji sa osnovnim zakonima fizike sa ciljem analize različitih karakteristika molekula. Ona može da pruži informacije o geometrijskom rasporedu atoma u molekulu koji odgovara stabilnosti i relativnoj energiji molekula, stepenu konverzije jednog stabilnog oblika molekula u drugi, interakciji dva različita molekula i pojedinim osobinama molekula (dipolni moment, polarizabilnost, elektronska gustina i dr.) [45]. Računarska hemija (engl. *Computational chemistry*) predstavlja granu hemije koja koristi metode teorijske hemije pri računarskim simulacijama molekulske strukture i molekulske interakcije. Jedan od glavnih problema računarske hemije je odabir odgovarajućeg teorijskog pristupa rešavanju zadatog problema, te je pri računarskom modelovanju molekula neophodno imati dovoljno teorijskog znanja kako bi se izbegli pogrešni rezultati, ili pak pogrešna interpretacija dobrih rezultata [45].

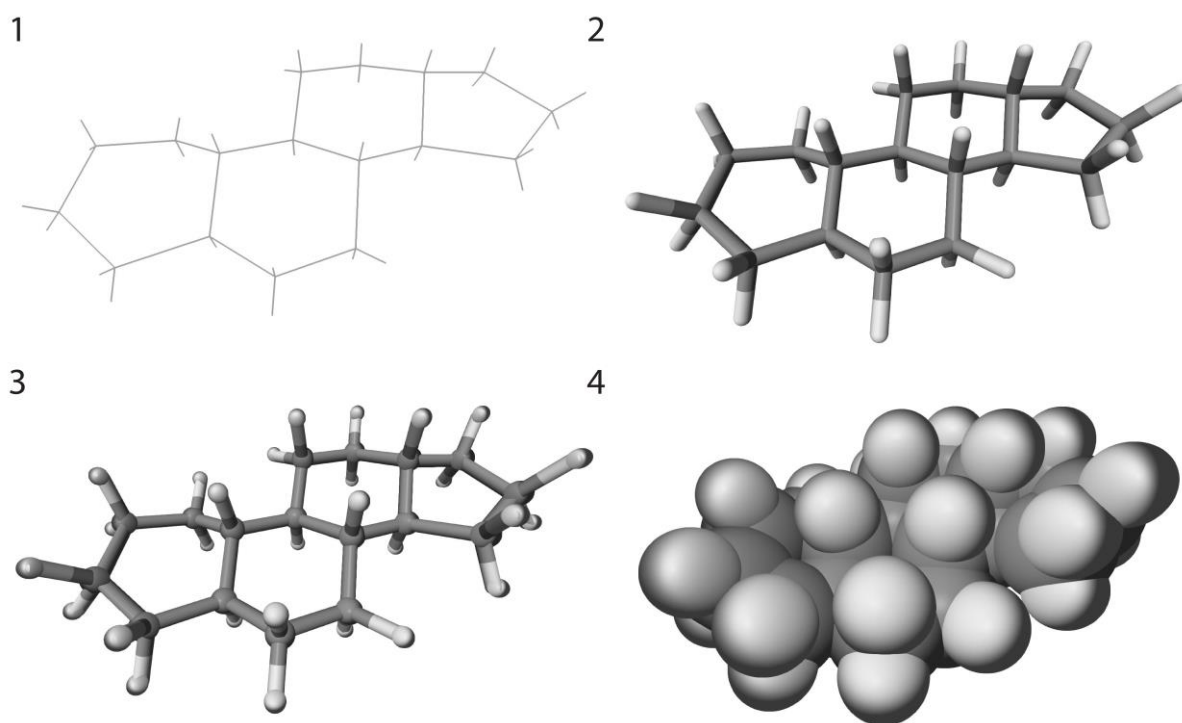
Pri određivanju karakteristika nekog molekula uglavnom se koriste dvodimenzionalne (2D) i trodimenzionalne (3D) molekulske strukture. Formiranje 2D struktura je prilično jednostavno i može se izvesti pomoću bilo kog softvera za crtanje molekulske strukture. Međutim, kod formiranja 3D modela, neophodno je pretpostaviti relativne pozicije atoma u prostoru koje su u skladu sa principom minimalne energije [46]. Jedan od češće korišćenih pristupa za optimizaciju geometrije malih organskih molekula je metod polja sila, tzv. FF metod (engl. *Force Field*), kao metod molekulske mehanike (MM, engl. *Molecular Mechanics*). Molekulska mehanika je zasnovana na matematičkom modelu koji posmatra molekul kao skup kuglica (atomi) povezanih oprugama (hemijske veze), što podrazumeva da se energija celog molekula menja sa promenom rastojanja između atoma i uglova između susednih hemijskih veza [47]. Alingerov MM metod (verzija 2 i verzija 3) je jedan od najčešće korišćenih FF metoda za minimizaciju energije malih organskih molekula [45], dok se za minimizaciju energije proteina i drugih bioloških struktura uglavnom upotrebljava MMFF94 metod. Energija molekula je prema FF metodi (E_{FF}) izražena kao funkcija torzione energije (E_t), energije istezanja veza (E_i), energije promene ugla između susednih veza (E_s) i energije interakcija između atoma koji nisu povezani hemijskom vezom (E_{int}) (slika 2.6.).



Slika 2.6. Osnovne FF energije molekula



Optimizacijom geometrije formira se stabilna 3D struktura molekula koja odgovara minimumu potencijalne energije. Stabilna 3D konfiguracija se može dobiti definisanjem minimalne E_{FF} kao funkcije koordinata atomskih nukleusa (FF metod zanemaruje elektrone u atomu). Optimizovana 3D struktura može biti prikazana na više načina, od kojih su neki dati na slici 2.7.



Slika 2.7. Žičani model (1), model cilindričnih veza (2) model kuglica-cilindar (3) i kalotni model (4) 3D strukture steroidnog jezgra optimizovane pomoću MM2 metode

Eksperimentalno određivanje 3D strukture molekula može se obaviti primenom kristalografije zasnovane na difrakciji X-zraka, nuklearne magnetne rezonance i mikrotalasne spektroskopije. Najveća baza podataka eksperimentalno određenih 3D struktura je *Cambridge Crystallographic Database (CSD)*, koja trenutno broji preko 750 000 struktura malih molekula [48], dok *Brookhaven Protein Data Bank (PDB)* raspolaže sa preko 100 000 struktura proteina i drugih makromolekula [49]. Međutim, broj jedinjenja kojima je do sada određena 3D struktura je relativno mali u odnosu na broj poznatih hemijskih jedinjenja (preko 90 miliona), te je doprinos računarskog modelovanja određivanju 3D struktura veoma značajan. Računarske tehnike često služe i za modelovanje još uvek nesintetisanih ili nestabilnih jedinjenja (prelazna stanja u hemijskim reakcijama), kojima struktura eksperimentalnim putem ne može biti određena [50].

Nakon optimizacije strukture, može se pristupiti računanju numeričkih karakteristika molekula, takozvanih *molekulskih deskriptora*. Molekulski deskriptori se definišu kao konačni rezultat logičko-matematičke procedure koja transformiše hemijsku informaciju, kodiranu u simboličkom predstavljanju molekula, u koristan broj ili rezultat nekog standardizovanog eksperimenta [51]. Molekulski deskriptor,



kao brojna vrednost, može biti upotrebljen u svrhu interpretacije neke karakteristike molekula ili sa ciljem formiranja matematičkog modela za predviđanje neke druge karakteristike molekula. Pri interpretaciji nekog fizičko-hemijskog parametra molekula dobijenog računskim putem (računarskim modelovanjem), neophodno je imati u vidu činjenicu da je modelovanje obavljeno na jednom izolovanom idealizovanom molekulu. Ukoliko se isti fizičko-hemijski parametar određuje eksperimentalno, radi se sa ravnotežnom smešom veoma velikog broja molekula pri nekoj temperaturi različitoj od nule. U tom slučaju izmerena vrednost se odnosi na smešu molekula u dinamičkoj ravnoteži, odnosno na grupu mogućih stereoizomera ili tautomera strukturno različitih od idealizovane strukture [46]. Stoga, molekulski deskriptori se više koriste u analizi kvantitativne veze između strukturnih karakteristika molekula i njihovih fizičko-hemijskih osobina i biološkog ponašanja. Jedna od mogućih klasifikacija molekulskih deskriptora data je u *tabeli 2.3*.

Tabela 2.3. Podela i tipovi molekulskih deskriptora

Klasifikacija molekulskih deskriptora prema:	
Dimenzionalnosti	<ol style="list-style-type: none">1. bezdimenzionalni (0D)2. jednodimenzionalni (1D)3. dvodimenzionalni (2D)4. trodimenzionalni (3D)5. četvorodimenzionalni (4D)
Načinu generisanja	<ol style="list-style-type: none">1. empirijski (eksperimentalni)2. teorijski (računski)
Delu molekula na koji se odnose	<ol style="list-style-type: none">1. lokalni2. globalni
Prirodi osobina koje opisuju	<ol style="list-style-type: none">1. fizičko-hemijski2. topološki3. kvantno-mehanički4. biološki (ADMET)

Podela deskriptora prema dimenzionalnosti zasnovana je na prikazu molekula iz koga se izvode proračuni. Bezdimezionalni prikaz podrazumeva molekulsku formulu, koja daje informacije o broju i vrsti atoma, ali ne daje nikakve informacije o strukturi molekula. Informacije koje daje jednodimezionalni prikaz molekula odnose se na prisustvo određenih molekulskih fragmenata, funkcionalnih grupa, supstituenata i hemijskih veza, dok molekulska struktura nije u potpunosti poznata. Dvodimezionalna struktura sadrži informacije o povezivanju atoma u molekulu, ne uzimajući u obzir dinamičku ravnotežu, rastojanje između nukleusa atoma koji grade molekul, ugao veze itd. Na osnovu 2D strukture računaju se topološki deskriptori. Trodimenzionalni (topografski) deskriptori zavise od



geometrijskih koordinata u molekulu, koji se posmatra kao rigidni geometrijski objekat, i obuhvataju elektronske i sterne deskriptore. Četvorodimenzionalni molekulski deskriptori (energetski deskriptori) zasnivaju se na interakcijama između ispitivanog molekula sa atomima ili molekulima iz njegovog okruženja.

Teorijski (računski) molekulski deskriptori imaju mnogobrojne prednosti u odnosu na deskriptore dobijene eksperimentom. Teorijski deskriptori ne sadrže statističku grešku merenja, često su u direktnoj vezi sa fizičko-hemijskim svojstvima molekula i većina njih je lako dostupna. Međutim, greške koje prate računanje teorijskih deskriptora leže u određenim pretpostavkama i aproksimacijama prilikom izvođenja proračuna. Proračun deskriptora može da obuhvati celokupnu strukturu molekula, kada se kao rezultat dobijaju globalni deskriptori, dok se za dobijanje lokalnih (fragmentnih) deskriptora u obzir uzimaju određeni delovi strukture (supstituenti, funkcionalne grupe).

Fizičko-hemijski deskriptori (lipofilnost, rastvorljivost, temperatura ključanja, temperatura topljenja, zapremina molekula, kritična temperatura, kritični pritisak, konstanta disocijacije i dr.) našli su veliku primenu u predviđanju fizičko-hemijskih osobina molekula [52], njihovog hromatografskog ponašanja [53] i bioloških karakteristika [54]. Jedan od najviše korišćenih fizičko-hemijskih deskriptora je deskriptor lipofilnosti neutralnog molekula predstavljen kao logaritam podeonog koeficijenta (P) za sistem 1-oktanol/voda – $\log P$ [55]. Postoje brojni pristupi računanju $\log P$ parametra, kao što su proračuni zasnovani na molekulskoj dinamici, topologiji molekula, elektrotopološkim stanjima, fragmentacionoj metodi, te je razlika između pojedinačnih $\log P$ vrednosti istog molekula, računatih različitim metodama ili računarskim programima, očekivana. Neki od najčešće korišćenih deskriptora lipofilnosti u QSAR analizama su $A\log P$, $C\log P$, $\log P_{Kow}$, $X\log P$, $m\log P$, $CS\log P$, $\log P_{Hyper}$ i $\log P_{ChDr}$ [54]. Pored $\log P$ parametra, neophodno je istaći i značaj logaritma distribucionog koeficijenta ($\log D$), koji je specifičan po tome što u obzir uzima raspodelu jonizovanih molekula između vode i 1-oktanola. Stoga, prilikom prikazivanja $\log D$ parametra neophodno je naglasiti pri kojoj vrednosti pH je računat. Rastvorljivost (S) je još jedan značajan fizičko-hemijski deskriptor. Izražava se u logaritamskom obliku ($\log S$) kao rastvorljivost molekula u vodi ili nekom drugom rastvaraču.

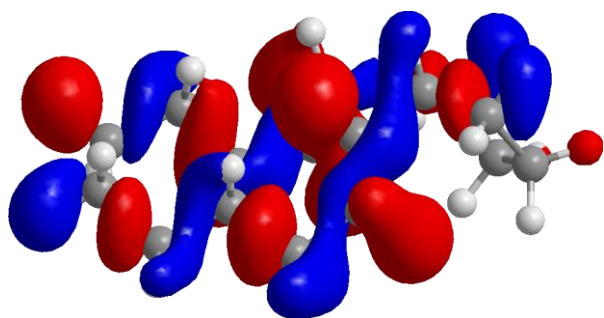
Topološki deskriptori proizilaze iz 2D strukture molekula. Formulirani su na osnovu teorije grafova [56]. Opisivanje molekula teorijom grafova zanemaruje dužinu hemijskih veza, uglove između veza, torzione uglove i ostale informacije o konformaciji strukture, ali zato uzima u obzir grananje, susedstvo i relativne distance između određenih komponenata molekula, dajući numeričku vrednost koja određuje široki opseg fizičko-hemijskih i bioloških karakteristika [57]. Neki od deskriptora koji pripadaju ovoj grupi su WI (*Wiener index*), BI (*Balaban index*), MTI (*Molecular topological index*), TC (*Total connectivity*), TVC (*Total valence connectivity*) i mnogi drugi.

Kvantno-mehanički ili kvantno-hemijski molekulski deskriptori izražavaju sve elektronske i geometrijske osobine molekula i njihove međusobne interakcije [58]. U osnovi se razlikuju od eksperimentalno izmerenih vrednosti, iako između njih postoji prirodno poklapanje. U ovu grupu deskriptora ubrajaju se deskriptori atomskih naelektrisanja, energije najviših zauzetih molekulskih orbitala (engl. *Highest Occupied Molecular Orbitals – HOMO*), energije najnižih slobodnih molekulskih orbitala (engl. *Lowest Unoccupied Molecular Orbitals – LUMO*), deskriptori elektronske gustine, superdelokalizabilnost, atom-atom polarizabilnost, molekulska polarizabilnost, dipolni moment, indeksi polarnosti i mnogi drugi [51,59]. Neki od računarskih programa za molekulsko modelovanje imaju



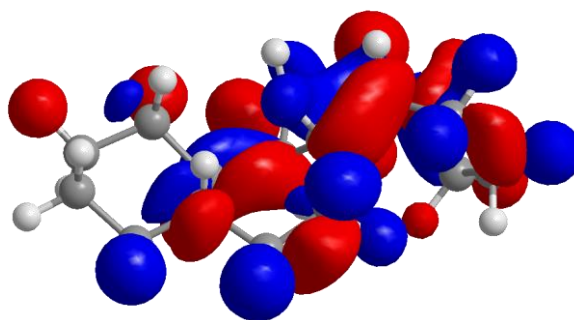
moгуćnost vizuelizacije HOMO i LUMO atomskih orbitala i elektrostatičkog potencijala (slika 2.8), što može da pruži bolji uvid u relativnu polarnost, kao i uticaj supstituenata i funkcionalnih grupa na ukupnu polarnost molekula. Mora se imati u vidu da su kvantno-hemijske kalkulacije izvedene na jednom molekulu na energetskom minimumu koji odgovara hipotetičkom fizičkom stanju gasa na 0 K pri beskonačno niskom pritisku, te ovi deskriptori nisu pogodni za korelaciju sa osobinama zavisnim od entropije i temperaturnih efekata [58].

1



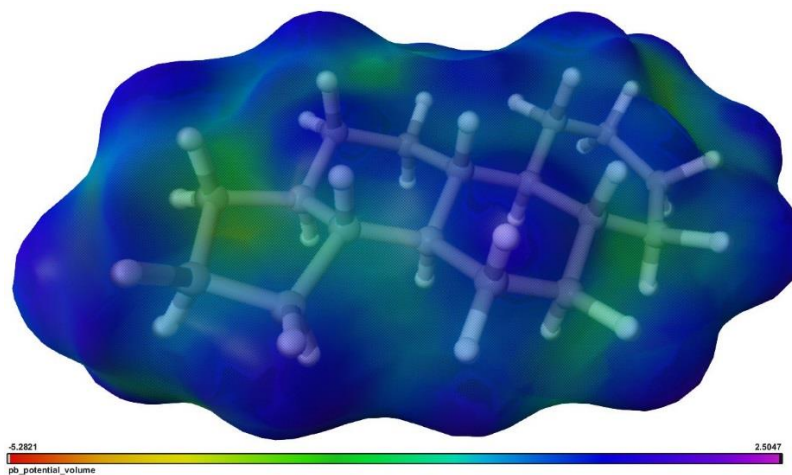
$$E_{HOMO} = -13,989 \text{ eV}$$

2



$$E_{LUMO} = 24,043 \text{ eV}$$

3

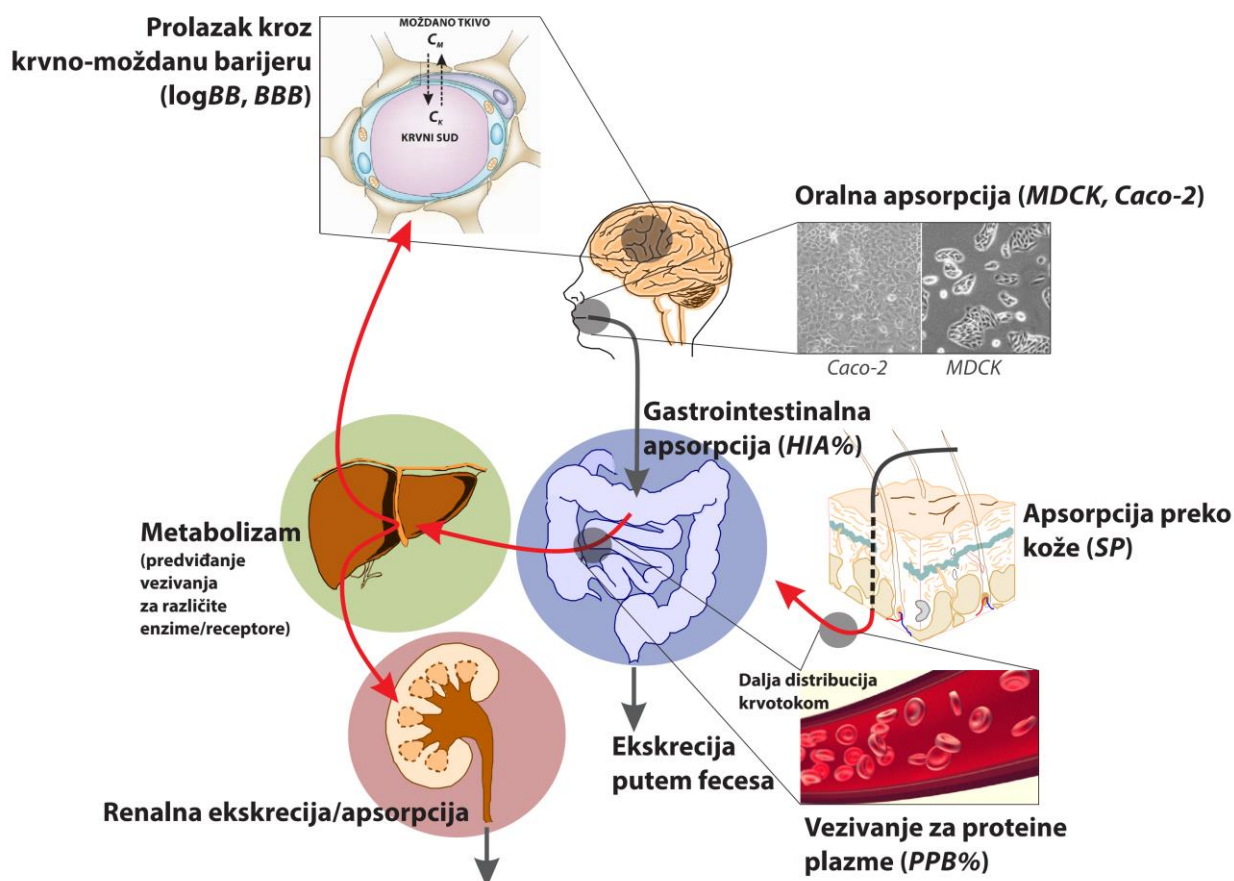


Slika 2.8. HOMO (1) i LUMO (2) orbitale i Poason-Bolcmanova površina elektrostatičkog potencijala (3) steroidnog jezgra modelovane pomoću programa ChemBio3D [60] i BioLuminate® [61]

Bioški ili ADMET (engl. *Absorption, Distribution, Metabolism, Excretion and Toxicity*) deskriptori numerički opisuju biološke osobine nekog jedinjenja u pogledu njegove apsorpcije, distribucije, metabolizma, izlučivanja (ekskrecije) i toksičnosti (farmakokinetički parametri). ADMET deskriptori se najčešće primenjuju kao eliminacioni faktor pri dizajniranju novih farmakološki aktivnih jedinjenja doprinoseći smanjenju polaznog skupa i odabiru supstanci koje će biti predmet daljih *in vitro* i *in vivo*



ispitivanja. Softveri koji omogućavaju predviđanje ADMET karakteristika, a samim tim i predviđanje sudbine biološki aktivne supstance u organizmu, zasnovani su na matematičkim modelima formiranim na osnovu velike baze podataka. *Slika 2.9.* prikazuje neke od mogućih puteva distribucije supstanci u organizmu, kao i neke najčešće modelovane ADMET deskriptore.



Slika 2.9. Neki od mogućih puteva jedinjenja u organizmu nakon administracije sa najčešće korišćenim *in silico* ADMET deskriptorima u QSAR modelovanju

Pri dizajniranju novih terapeutika, mora se voditi računa o mogućim metaboličkim putevima, načinima apsorpcije, dopremanju do mesta delovanja i drugim mnogobrojnim faktorima. *In silico* ADMET pristup omogućava predviđanje ponašanja molekula u organizmu i time skraćuje vreme potrage za terapeutcima sa najefikasnijim dejstvom. U ovu grupu deskriptora ubrajaju se :

- deskriptori oralne apsorpcije, koja je često izražena preko permeabilnosti Madin-Darby epitelnih ćelija psećeg bubrega (*MDCK*) i humanih epitelnih ćelija kolorektalnog adenokarcinoma (*Caco-2*) [62,63],
- deskriptor apsorpcije u gastrointestinalnom traktu (*HIA%*, engl. *Human Intestinal Absorption*) [64],



- deskriptor prolaska kroz krvno-moždanu barijeru, izražen kao odnos koncentracije jedinjenja u krvnom sudu (c_K) i moždanom tkivu (c_M), *BBB* (engl. *Blood Brain Barrier permeability*), ili u logaritamskom obliku, $\log BB$ [65,66],
- deskriptor kožne apsorpcije/propustljivosti (*SP*, engl. *Skin Permeability*) koja je izražena kao logaritam distribucionog koeficijenta supstance ($\log K_p$) između *stratum corneum*-a i krvnog suda [67],
- deskriptori renalne apsorpcije (*ROCT*, engl. *Renal Organic Cation Transporter*) i ekskrecije [68],
- deskriptor vezivanja za proteine krvne plazme (*PPB%*, engl. *Plasma Protein Binding*) [69],
- deskriptori inhibicije, indukcije ili aktivacije određenih enzima i vezivanja za određene receptore (*EI* – inhibicija enzima, *KI* – inhibicija kinaze, *PI* – inhibicija proteaze, *GPCR* – vezivanje za G-protein spregnuti receptor, *NRL* – vezivanje za nuklearni receptor, *ICM* – modulacija jonskih kanala i dr.) [70],
- deskriptori toksičnosti jedinjenja (akutna oralna toksičnost, kancerogenost, mutagenost i dr.) [71].

Generalno, apsorpcija jedinjenja u organizmu je manja ukoliko je molarna masa molekula veća od 500 g/mol, ako molekul ima više od 10 akceptora i više od 5 donora za građenje vodonične veze i ako je izračunata $\log P$ vrednost veća od 5 (pravilo Lipinskog) [72].

Računanjem molekulskih deskriptora često se izbegavaju vrlo skupi laboratorijski eksperimenti. *In silico* molekulski deskriptori su našli široku primenu u definisanju osobina novih jedinjenja ili pri definisanju osobina velikog skupa jedinjenja čija su eksperimentalna određivanja prilično komplikovana i dugotrajna. Savremeni softveri za modelovanje jednostavnih i složenih molekulskih struktura omogućavaju računanje preko 5000 molekulskih deskriptora, čiji broj kontinualno raste. Razvoj novih molekulskih deskriptora prate i novi pristupi posmatranju i opisivanju molekulske strukture [73-75]. Da bi novi molekulski deskriptor bio opšteprihvaćen, on mora da nosi hemijsku informaciju, da se dobija tačno definisanim procedurama koje omogućavaju njegovo računanje za bilo koji set molekula i da se njegova vrednost menja sa promenama u molekulskoj strukturi [51].

Molekulski deskriptori su neizostavni činilac osnovnih metoda hemometrijskog pristupa, kojima se uspostavlja korelacija između strukturnih karakteristika definisanih molekulskim deskriptorima (teorijskim ili eksperimentalnim) i fizičko-hemijskih ili bioloških karakteristika molekula. U prvom slučaju primenjuje se QSPR pristup (engl. *Quantitative Structure–Property Relationship approach*), dok je u drugom slučaju u pitanju QSAR pristup.



2.6. Osnovne hemometrijske metode

2.6.1. Hemometrija kao multidisciplinarni naučni pristup

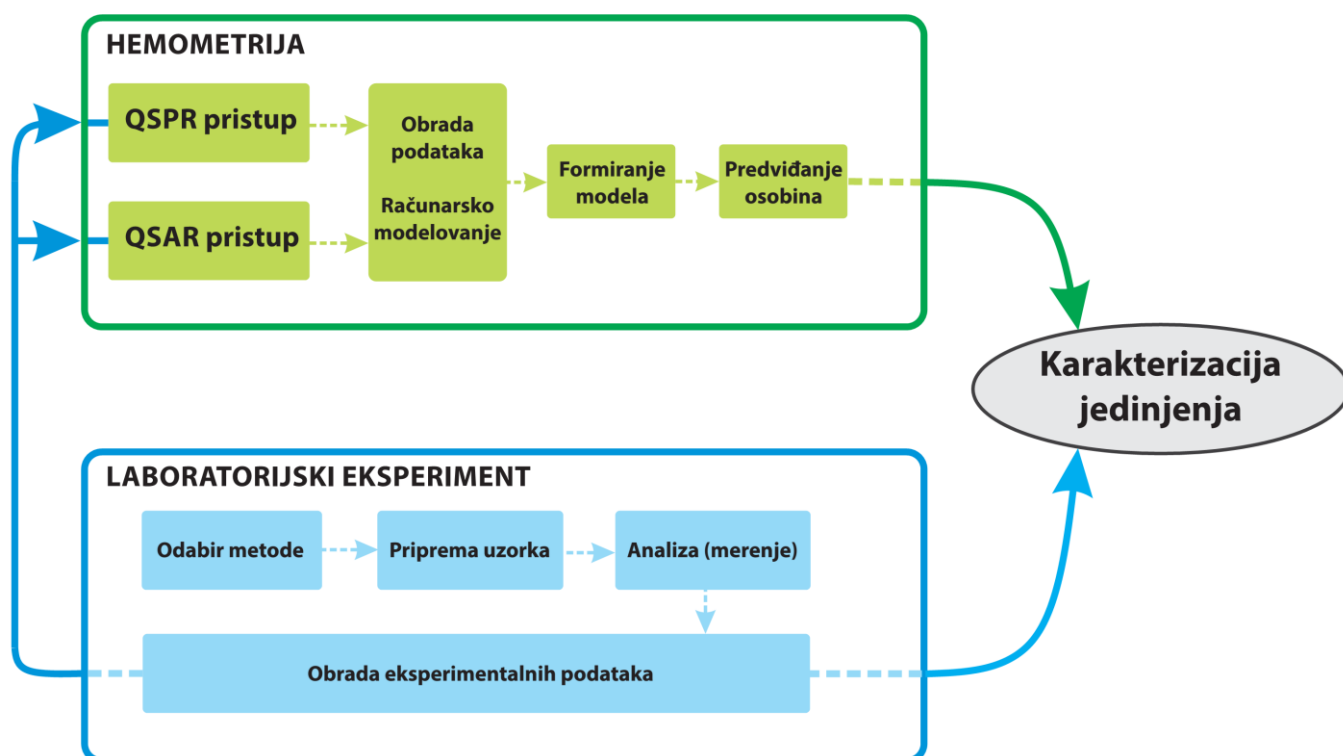
Rezultati hemijske, fizičko-hemijske ili biološke analize ponekad ne otkrivaju mnogo o sistemu iz kog su proistekli ukoliko se ne analiziraju na adekvatan način, statistički obrade i prezentuju. Razvojem instrumentalnih metoda analize stvorile su se mogućnosti dobijanja ogromnog broja rezultata, a klasični statistički pristupi su bili često nedovoljni da opišu sve moguće varijacije i uticaje različitih promenljivih na konačan ishod eksperimenta. Svante Vold (*Svante Wold*) je 1974. godine prvi put uveo pojam „hemometrija“, definišući ga kao „umetnost ekstrahovanja hemijski relevantnih informacija iz podataka dobijenih hemijskim eksperimentima“ [76]. Osamdesetih godina XX veka hemometrija se izdvojila kao posebna naučna disciplina [77]. Hemometrija se može predstaviti kao multidisciplinarni pristup obradi podataka u oblasti hemije, zasnovan na primeni matematičkih i statističkih metoda, sa ciljem izdvajanja korisnih informacija. Na samom početku razvoja bila je primenjivana uglavnom u oblasti analitičke hemije [77], međutim hemometrijske metode su do danas našle široku primenu i u mnogim drugim oblastima, kao što su organska hemija [78], mikrobiologija [79-82], forenzička hemija [83], prehrambeno i farmaceutsko inženjerstvo [78,84-87], biotehnologija [88]. Hemometrija se takođe oslanja i na razvoj računarskih programa omogućavajući brzo izvođenje kompleksnih izračunavanja, ali i bolji vizuelni prikaz rezultata. Dizajniranje eksperimenata, prepoznavanje oblika, predviđanje osobina i različite vrste simulacija su zadaci pri čijem rešavanju hemometrija daje izuzetno veliki doprinos.

Jedan od najznačajnijih domena primene hemometrije je predviđanje osobina molekula. Tu se uočava tesna veza između molekulskog modelovanja i hemometrije. U zavisnosti od prirode osobine koja se predviđa, hemometrija nudi dva glavna pristupa – QSAR i QSPR. Njihova pozicija u celokupnom toku analize i predviđanja karakteristika novosintetisanih i još uvek nesintetisanih jedinjenja prikazana je na slici 2.10. Na datoj slici se može uočiti da je formiranje QSPR i QSAR modela povezano sa eksperimentalnim radom, s obzirom na to da se eksperimentalni podaci koreliraju sa određenom osobinom molekula radi formiranja stabilnog matematičkog modela. Formirani model služi za predviđanje tražene varijable novih jedinjenja koja nisu učestvovala u formiranju modela. Time se izbegava eksperiment, štede vreme i finansijska sredstva. Obrada podataka nakon odabira hemometrijskog pristupa zavisi od željenog cilja, što može biti predviđanje osobina jedinjenja ili njihova diskriminacija. Hemometrija u tom pogledu obuhvata regresione i klasifikacione tehnike.

Osnovna razlika između QSPR i QSAR analize je u vrsti izlazne informacije, odnosno osobini molekula koja se predviđa. Kao zavisno promenljive varijable u QSPR analizi javljaju se fizičko-hemijske karakteristike jedinjenja (tačka ključanja, tačka topljenja, polarizabilnost, hromatografska retencija itd.). QSAR modelovanje ima za cilj predviđanje bioloških karakteristika, uključujući antifungalnu, antibakterijsku i antikancerogenu aktivnost, prolazak kroz krvno-moždanu barijeru, stepen oralne apsorpcije, sposobnost inhibicije različitih enzima i mnoge druge [78-83].



Kao deo QSPR pristupa, razvila se QSRR metoda (engl. *Quantitative Structure–Retention Relationships*), koja je našla široku primenu u predviđanju retencionog ponašanja i lipofilnosti velikog broja različitih jedinjenja [89-95]. QSRR metoda omogućava i karakterizaciju stacionarnih i mobilnih faza, kao i analizu efikasnosti razdvajanja analita u različitim hromatografskim sistemima [94].



Slika 2.10. Osnovni hemometrijski pristupi predviđanju molekulske osobine i karakterizaciji jedinjenja i njihova veza sa laboratorijskim eksperimentom

U naučnim publikacijama se javljaju i modifikacije QSAR akronima, kojima se u nazivu metode naglašava specifična karakteristika koja se predviđa, kao u slučaju QSTR modela (engl. *Quantitative Structure–Toxicity Relationships*) za predviđanje toksičnosti jedinjenja [96], QSAAR modela (engl. *Quantitative Structure–Antioxidant Activity Relationships*) za predviđanje antioksidativne aktivnosti [97], ali i modela za predviđanje biološke aktivnosti na osnovu retencionih parametara (QRAR, engl. *Quantitative Retention–Activity Relationships*) [98]. Kao u slučaju QSRR analize koja suštinski spada u QSPR pristup, tako i navedene metode koriste QSAR osnove, te se ne posmatraju kao posebni hemometrijski pristupi.



2.6.2. Osnove QSRR analize i pojam hromatografske lipofilnosti

Krajem 70-ih godina XX veka objavljene su naučne studije koje su predstavljale začetke metodologije poznate kao QSRR [99]. Vremenom, QSRR postaje veoma popularna hemometrijska metoda u analizi širokog spektra jedinjenja različitim hromatografskim tehnikama. Osnovu QSRR metode čini matematički model koji povezuje nezavisno promenljive varijable koje kodiraju molekulsku strukturu (molekulski deskriptori) sa retencionim parametrom, kao zavisno promenljivom varijablom. Glavna svrha ovih matematičkih modela je predviđanje retencionog ponašanja serije strukturno sličnih jedinjenja za koju su definisani (ne postoji opšti QSRR model koji važi za sva jedinjenja). Stoga, QSRR omogućava predviđanje retencionog ponašanja novih, još uvek nesintetisanih, jedinjenja na osnovu izračunatih molekulskih deskriptora, kao i optimizaciju odvajanja komponenata iz kompleksnih smeša i dizajniranje hromatografskih eksperimenata [100]. Fizičko-hemijski deskriptori pokazuju najznačajnije korelacije sa retencionim parametrima i najčešće služe za njihovo predviđanje, dok kvantno-mehanički deskriptori omogućavaju uvid u mehanizam hromatografske retencije na molekularnom nivou [100]. Broj molekulskih deskriptora koji mogu biti izračunati za svaki analit je izuzetno velik, međutim problem može nastati pri tumačenju fizičkog smisla određenih deskriptora, pogotovo u slučaju kada deskriptori sa nejasnim značenjem formiraju statistički stabilan QSRR model [101]. QSRR, kao ekstratermodinamički pristup [101], može biti primenjen na retencione podatke dobijene pomoću različitih vidova hromatografije (adsorpciona i podeona hromatografija) i različitih hromatografskih tehnika (kolonska i planarna).

U tačnoj podeonoj hromatografiji raspodelu analita između mobilne i stacionarne faze možemo posmatrati kao termodinamičku ravnotežu, koja je detaljno opisana u revijalnom radu o QSRR analizi [99]. Ponašanje analita u tačnoj hromatografiji zavisi od njegove raspodele između mobilne (m) i stacionarne faze (s). Termodinamički ravnotežni parametar koji definiše retenciju analita u hromatografskom procesu je koeficijent raspodele analita između stacionarne i mobilne faze (K):

$$K = c_s / c_m \quad (2.1.)$$

(c_s – koncentracija analita u stacionarnoj fazi, c_m – koncentracija analita u mobilnoj fazi)

Veza između Gibsove slobodne energije (ΔG_r^0) prelaza jednog mola analita iz standardnog stanja pri beskonačnom razblaženju u nepokretnoj fazi u standardno stanje pri beskonačnom razblaženju u pokretnoj fazi, pri temperaturi T i konstantnom pritisku, i koeficijenta raspodele analita (K) između stacionarne i mobilne faze data je relacijom:

$$-\Delta G_r^0 = RT \ln K \quad (2.2.)$$

(R – univerzalna gasna konstanta)

Retencioni parametar analita koji se meri u tačnoj hromatografiji naziva se *kapacitetni faktor analita (rastvorka)*, k_a . Kapacitetni faktor predstavlja odnos molova analita u stacionarnoj (n_s) i mobilnoj fazi (n_m):

$$k_a = n_s / n_m \quad (2.3.)$$



iz čega sledi relacija:

$$k_a = (c_s / c_m) \cdot (V_s / V_m) = K \cdot (V_s / V_m) \quad (2.4.)$$

Dakle, direktna veza između K i k_a u podeonoj i adsorpcionoj hromatografiji [4] data je relacijom:

$$k_a = K \cdot \varphi \quad (2.5.)$$

(φ – zapreminski odnos)

U podeonoj kolonskoj hromatografiji kapacitetni faktor se najčešće izražava preko retencionog vremena analita (t_a) i mrtvog retencionog vremena (t_m) koji se očitavaju sa hromatograma:

$$k_a = (t_a - t_m) / t_m \quad (2.6.)$$

Mrtvo retenciono vreme se definiše kao vreme prolaska supstance kroz hromatografsku kolonu bez zadržavanja na stacionarnoj fazi. Ono se može odrediti upotrebom tzv. „markera mrtvog vremena“, kao što su izotopski ili radioaktivno obeležene komponente eluenta, jonske i nejonske komponente ili, generalno, supstance čije su interakcije sa stacionarnom fazom minimalne (zanemarljive) [99,102,103]. Još jedan od načina definisanja mrtvog vremena zasniva se na očitavanju retencionog vremena prvog signala (pika) na hromatogramu [104].

Kapacitetni faktor analita u HPLC često se izražava u logaritamskom obliku – $\log k_a$. Pored vrste stacionarne i mobilne faze i prirode analita, na $\log k_a$ utiču i različite interakcije između analita i rastvarača, kao što su dipol-dipol, dielektrične i disperzione interakcije i formiranje vodoničnih veza [16]. Dipol-dipol interakcije su zastupljene ukoliko analit i rastvarač imaju permanentne dipolne momente, dok se disperzione sile javljaju u slučaju interakcija nepolarnih susednih molekula bez permanentnih dipolnih momenata. Ukoliko između molekula analita i rastvarača postoji elektrostatičko privlačenje tipa vodonik – nukleofilni atom sa slobodnim elektronskim parom, dolazi do formiranja vodonične veze. Dielektrične interakcije stimulišu rastvaranje jonskih jedinjenja u polarnom rastvaraču. Generalno, u tečnoj hromatografiji kapacitetni faktor analita zavisi od ravnotežnih interakcija [105]:

- analit – stacionarna faza,
- analit – mobilna faza,
- stacionarna faza – mobilna faza,
- solvent – solvent (modifikator – modifikator, voda – voda, modifikator – voda),
- stacionarna faza – stacionarna faza (u slučaju interakcije fleksibilnih fragmenata),
- analit – analit (u slučaju visokih koncentracija).

U hromatografiji na obrnutim fazama retencija rastvorka je dominantno uzrokovana interakcijama nepolarnog dela molekula analita sa polarnim rastvaračem (voda) u mobilnoj fazi. Ove solvofobne (hidrofobne) interakcije zavise od površinskog napona i gustine rastvarača i površine analita [4]. Molekulski mehanizam retencionog ponašanja analita u RP-HPLC može se predstaviti modelom linearne korelacije slobodne energije solvatacije (LSER, engl. *Linear Solvation Energy Relationship*) [106,107]. Solvatohromna LSER jednačina korelira retenciju analita u tečnoj hromatografiji ($\log k_a$) sa molekulskom zapreminom, dipolarnosti/polarizabilnosti i sposobnosti donacije/primanja protona u vodoničnoj vezi [106]. Brojnim ispitivanjima je pokazano da na vrednost $\log k_a$ parametra u RP-HPLC



značajan uticaj ima vrednost podeonog koeficijenta analita u dvofaznom sistemu 1-oktanol/voda, $\log P$ [99,100,108,109]. Ovaj uticaj se može predstaviti relacijom [4,12,105]:

$$\log P = A \log k_a + B \quad (2.7.)$$

(A, B – nagib i odsečak linearne zavisnosti)

Izraz 2.7. je ključna spona između podeonog koeficijenta $\log P$, kao klasičnog merila lipofilnosti nekog jedinjenja, i tzv. *hromatografske lipofilnosti* definisane RP-HPLC tehnikom [12,13,105]. Iako su utvrđene odlične korelacije između $\log P$ dobijenog metodom mućkanja i $\log k_a$, razlika u njihovom načinu dobijanja je evidentna. Glavna prednost hromatografske lipofilnosti je mogućnost upotrebe organskih modifikatora polarnosti vodene faze. Različiti udeli organskog modifikatora u vodenoj fazi mogu dati različite vrednosti parametra hromatografske lipofilnosti, odnosno mogu uticati na retenciju analita u hromatografskom sistemu. Zato je neophodno naglasiti sastav mobilne faze, tj. udeo organskog modifikatora u binarnoj mobilnoj fazi, pri kome je definisana hromatografska lipofilnost u slučaju izokratskog postupka hromatografisanja [105]. Loša korelacija $\log P$ i retencije u RP-HPLC može biti posledica jonizacije (disocijacije) analita, koja se suzbija dodavanjem pufera u mobilnu fazu [4]. Uticaj zapreminskog udela modifikatora (φ) na vrednosti $\log k_a$ u RP-HPLC sistemu definisan je Soczewinski-Snyder-ovom jednačinom [110,111]:

$$\log k_a = \log k_w - S \varphi \quad (2.8.)$$

(S – nagib, $\log k_w$ – odsečak linearne zavisnosti)

Odsečak, $\log k_w$, odgovara retenciji analita u mobilnoj fazi bez modifikatora, odnosno čistoj vodi. Nagib prikazane zavisnosti (S) je konstanta za dati analit i HPLC sistem. Neophodno je naglasiti da se $\log k_w$ vrednost razlikuje od eksperimentalno određene retencije u čistoj vodi (kada ju je moguće odrediti) i da zavisi od vrste primenjenog organskog modifikatora pri njenom određivanju, što može da navede na zaključak da je određivanje $\log k_w$ manipulacija i da u nekim slučajevima ovaj parametar nema fizičko značenje [105]. Njegova interpretacija kao retencionog faktora analita u sistemu bez organskog modifikatora u mobilnoj fazi (čistoj vodi) je pogrešna, te se treba smatrati samo kao način standardizacije retencionih podataka [105]. U nekim slučajevima zavisnost između $\log k_a$ i φ može se opisati i polinomskom funkcijom drugog reda [4,105]:

$$\log k_a = A \varphi^2 + B \varphi + \log k_w \quad (2.9.)$$

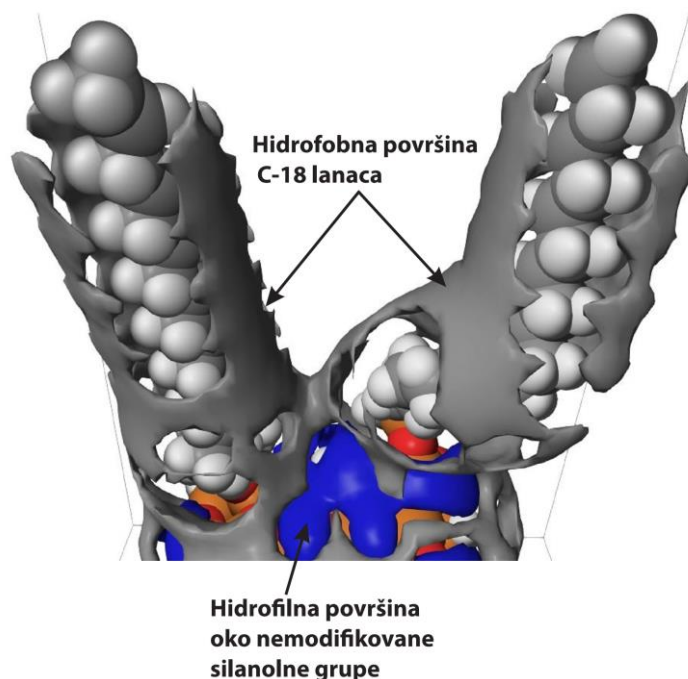
(A, B – konstante)

Analizom korelacije je ustanovljeno da se hromatografska lipofilnost seta test analita, dobijena pomoću RP-HPLC sistema sa mobilnom fazom metanol/voda, i $\log P$ parametar najbolje koreliraju po linearnom modelu [112]. U poređenju sa mobilnom fazom acetonitril/voda, mobilna faza metanol/voda pokazuje više sličnosti sa sistemom 1-oktanol/voda [112], te se određivanje hromatografske lipofilnosti jedinjenja često definiše upravo korišćenjem metanola kao modifikatora.

Kod primene hromatografije na obrnutim fazama za određivanje hromatografske lipofilnosti uglavnom se koristi modifikovani silika gel sa n -oktadecil (C-18) i n -oktil (C-8) ugljovodoničnim nizom. Kod analize steroidnih jedinjenja najpopularniji je C-18 modifikovani silika gel [113]. Pri korišćenju



modifikovanog silika gela, na retenciju jedinjenja utiču i zaostale nemodifikovane silanolne grupe koje su polarnije u odnosu na ugljovodonični lanac (silanofilne interakcije) [4]. Računarskim modelovanjem hidrofилnih i hidrofobnih površina segmenta C-18 modifikovanog silika gela programom BioLuminate® [61] može se uočiti značajan hidrofилni karakter na mestu nemodifikovanih silanolnih grupa između C-18 ugljovodoničnih lanaca (slika 2.11.).



Slika 2.11. Hidrofилne i hidrofobne površine segmenta C-18 modifikovanog silika gela sa preostalim nemodifikovanim silanolnim grupama

Značajan broj QSRR modela definisan je na osnovu retencionih parametara dobijenih pomoću HPLC tehnike [84,104,114-118], međutim potrebno je naglasiti i da je veoma veliki broj QSRR studija zasnovan na primeni TLC tehnike sa ciljem analize retencionog ponašanja različitih jedinjenja, analize karakteristika stacionarnih faza i ispitivanja uticaja pojedinih fizičko-hemijskih parametara na retenciju [53,54,90-95,118-120]. Retencija jedinjenja u TLC sistemu se najčešće izražava u obliku R_M vrednosti, računata na osnovu R_f parametra:

$$R_M = \log((1/R_f) - 1) \quad (2.10.)$$

(R_f – odnos pređenog puta supstance i fronta mobilne faze)

Ekstrapolacijom zavisnosti $R_M - \varphi$ na 0% φ , dobija se R_M^0 retencioni parametar jedinjenja u čistoj vodi, analogno $\log k_w$ parametru koji je definisan primenom HPLC tehnike.

Neophodno je naglasiti da se pojam hromatografske lipofilnosti može podjednako odnositi na parametre $\log k_a$, $\log k_w$, R_M i R_M^0 , u zavisnosti od primenjene hromatografske tehnike, ali i na alternativne parametre lipofilnosti (hidrofobnosti), φ_0 za HPLC [121] i C_0 za TLC tehniku [122]:



$$\varphi_0 = -\log k_w / S \quad (2.11.)$$

$$C_0 = -R_M^0 / S \quad (2.12.)$$

(S – nagib zavisnosti $\log k_a - \varphi$, odnosno $R_M - \varphi$)

Navedeni parametri takođe mogu biti upotrebljeni kao zavisno promenljive u QSRR modelima.

Najjednostavniji primer QSRR modela mogao bi se predstaviti modifikovanom jednačinom 2.7. u kojoj bi zavisno promenljiva bila retencija ($\log k_a$), a nezavisno promenljiva parametar lipofilnosti ($\log P$), dobijen eksperimentalno ili izračunat pomoću računarskog programa. Međutim, QSRR relacije mogu biti kompleksne, uključujući više od jedne nezavisno promenljive veličine. Često su zasnovane na primeni multivarijantnih regresionih tehnika koje mogu da predvide retencioni faktor ispitivanih jedinjenja u primenjenom hromatografskom sistemu na osnovu više varijabli (faktora). Tako je u slučaju serije estranskih derivata formiran QSRR model pomoću multiple linearne regresije, namenjen predviđanju retencionog parametra (R_M^0) strukturno sličnih jedinjenja u RP-TLC sistemu pomoću izračunatih molekulskih deskriptora [53]:

$$R_M^0 = 0,002038 GE + 0,146148 MR - 0,017372 MP + 1,571067 \quad (2.13.)$$

(R_M^0 – retencioni parametar jedinjenja u RP-TLC pri 0% modifikatora (metanola) u mobilnoj fazi (voda); GE – Gibsova slobodna energija, MR – molarna refraktivnost, MP – temperatura topljenja)

QSRR modeli koji definišu retenciju serije jedinjenja u određenom hromatografskom sistemu neophodno je da budu statistički validni i definisani na osnovu velikog broja jedinjenja, kako bi se mogli pouzdano koristiti za predviđanje retencionog ponašanja. QSRR relacije definisane na manjem broju jedinjenja mogu biti pokazatelji uticaja određenih varijabli (deskriptora) na retenciju jedinjenja upotrebljenih za formiranje modela, ali je njihova sposobnost predviđanja ograničena.

2.6.3. Osnove QSAR analize

Pre više od pet decenija, Korvin Hanš (*Corwin Hansch*) i saradnici [123] postavili su temelje QSAR analize ispitujući korelaciju između biološke aktivnosti i fizičkih osobina molekula primenom regresione analize. Danas, QSAR modelovanje je jedan od glavnih računskih pristupa korišćenih u medicinskoj hemiji i farmakologiji, čija je glavna namena predviđanje bioloških osobina jedinjenja na osnovu njihove hemijske strukture. QSAR pristup, pored definisanja matematičkih relacija, obuhvata i računarsko modelovanje molekulske strukture kako jednostavnih organskih molekula, tako i složenih molekulskih celina (enzimi, ćelijski receptori, DNA itd.). QSAR metoda je u začetku bila smatrana za oblast fizičke organske hemije, međutim razvila se u zasebnu celinu, od primene jednostavnih regresionih modela na manjim molekulskim serijama, pa sve do analize veoma velikih setova podataka složenim statističkim i regresionim tehnikama [124].



Unoseći niz olakšica u analizu biološki aktivnih jedinjenja i potragu za potencijalnim terapeutima u lečenju različitih vrsta oboljenja, QSAR analiza je postala neizostavni deo procesa dizajniranja novih lekova. Njeno mesto u tzv. virtuelnom skriningu [125] biološki aktivnih jedinjenja prikazano je *slikom* 2.12. Virtuelnim skriningom se obavlja identifikacija hemijskih jedinjenja iz celokupne baze jedinjenja za koja se pretpostavlja da će pokazati željenu biološku aktivnost. S obzirom na to da baza sadrži preko milion jedinjenja, doprinos QSAR pristupa njihovoj selekciji je od velikog značaja.



Slika 2.12. Mesto QSAR pristupa u virtuelnom skriningu baze hemijskih jedinjenja [124]

Procenjuje se da je broj komercijalno dostupnih hemijskih jedinjenja oko 22 miliona, dok se broj prirodnih jedinjenja kreće oko 160 000 [126]. Međutim, primena prirodnih jedinjenja u dizajniranju farmakološki aktivnih supstanci je ograničena, delom zbog tehničkih prepreka koje onemogućavaju kompletan skrining njihovog delovanja prema datim metama [126]. Međutim, mogućnosti organske sinteze novih biološki aktivnih jedinjenja su skoro neograničene. S obzirom na složenost organskih sinteza i finansijske troškove koji ih prate, QSAR modelovanje može da pruži značajne smernice pri organskim sintezama i skрати vreme potrage za jedinjenjima sa željenom biološkom aktivnošću [78].

QSAR metodologija se bazira na činjenici da razlike u biološkoj aktivnosti jedinjenja iste molekulske serije mogu biti kvantitativno korelirane sa razlikama u njihovoj strukturi. Ove razlike se mogu odnositi na:

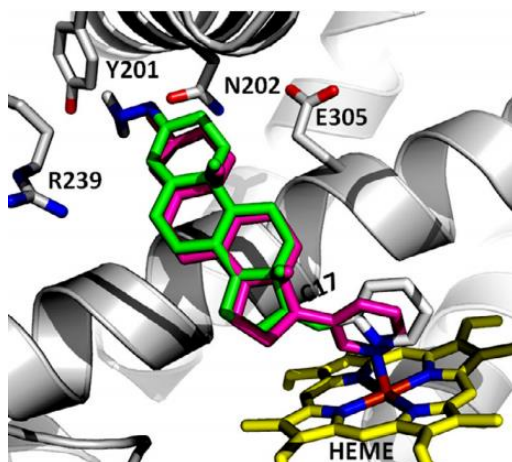


- veličinu i oblik ugljeničnog skeleta;
- prirodu i stepen supstituisanja;
- stereohemiju.

U klasičnom QSAR pristupu razmatraju se samo 2D strukture molekula (2D-QSAR), pri čemu se uglavnom analiziraju varijacije supstituenata na delu molekula identičnom za sve članove iste molekulske serije, dok savremeni 3D-QSAR pristup povezuje biološku aktivnost sa 3D molekulskom strukturom [54]. Stoga, kada je analiza 3D molekulske strukture postala rutina sa paralelnim razvojem nekoliko računarskih tehnika molekuskog modelovanja u 80-im godinama prošlog veka, započet je novi pristup dizajniranju novih biološki aktivnih supstanci, potencijalnih terapeutika, poznat kao kompjuterski asistirano molekulske modelovanje (CAMD, engl. *Computer-Aided/Assisted Molecular Design*), odnosno kompjuterski asistirano dizajniranje lekova (CADD, engl. *Computer-Aided/Assisted Drug Design*) [127]. U zavisnosti od načina računanja i odabira molekulskih deskriptora i specifičnih statističkih algoritama za konstituisanje QSAR modela, predloženo je nekoliko QSAR metodologija [127]:

- 0D-QSAR (zasnovana na broju i vrsti atoma u molekulskoj strukturi);
- 1D-QSAR (zasnovana na broju molekulskih fragmenata, funkcionalnih grupa, supstituenata);
- 2D-QSAR (zasnovana na topologiji molekula);
- 3D-QSAR (zasnovana na geometriji molekula, odnosno koordinatama atoma u prostoru);
- 4D-QSAR (zasnovana na ispitivanju višestrukih geometrijskih orijentacija atoma u molekulima liganada);
- 5D-QSAR (zasnovana na analizi prilagođavanja aktivnog mesta receptora topologiji pojedinačnog liganada);
- 6D-QSAR (uzima u obzir različite solvatazione modele)[128].

Savremeni računarski programi za molekulski dizajn omogućavaju formiranje 3D modela sternih interakcija različitih jedinjenja (liganada) sa aktivnim mestima odgovarajućih proteina (enzima). Ovaj pristup je poznat kao molekulski doking (engl. *Molecular Docking*) [44,129]. Na slici 2.13. prikazana je simulacija vezivanja androstanskog derivata za aktivno mesto enzima CYP17A1 [44].



Slika 2.13. QSAR doking studija koja predviđa prostornu orijentaciju 17(E)-pikoliniliden androstanskog derivata (zelena struktura) pri vezivanju za aktivno mesto CYP17A1 enzima [44]



Jedan od najpopularnijih metoda 3D-QSAR analize je komparativna analiza molekuskog polja (CoMFA, engl. *Comparative Molecular Field Analysis*) [130], zasnovana na primeni 3D molekuskog modelovanja i metodi parcijalnih najmanjih kvadrata (PLS). Cilj CoMFA je analiza korelacije između biološke aktivnosti serije jedinjenja i njihove 3D konformacije, uzimajući u obzir elektrostatičke i sterne osobine i mogućnost formiranja vodonične veze.

Analizom povezanosti strukture i biološke aktivnosti brojnih jedinjenja, utvrđeno je da na biološku aktivnost najznačajniji uticaj imaju elektronski, sterni i lipofilni (hidrofobni) efekti [54]. Prva ispitivanja uticaja ovih efekata na reaktivnost jedinjenja sprovedi su Luis Hamet (*Louis Plack Hammett*) i Robert Taft (*Robert W. Taft*). Hamet je diskutovao uticaje elektronskih efekata supstituenata na reaktivnost benzoevih kiselina, dok je Taft kvantifikovao uticaj sternih efekata koristeći reakciju hidrolize estara [54,131]. Raspodela elektrona u molekulu ispoljava značajan efekat na biološku aktivnost, s obzirom na to da utiče na formiranje veze između jedinjenja i mesta delovanja (npr. aktivnog mesta enzima) [132]. Elektronski uticaji su u QSAR analizi definisani Hametovom konstantom (σ), dok je Taftov sterni parametar (E_s) prvi parametar kojim je definisan uticaj oblika i veličine molekula na njegovu aktivnost [132]. Još jedan važan parametar biološki aktivnog jedinjenja je molarna refraktivnost (MR), koja obuhvata zapreminu molekula i njegovu sposobnost polarizacije. Lipofilnost (P ili $\log P$) je važno merilo sposobnosti molekula da savlada prolazak kroz lipidne membrane kako bi dostigao mesto delovanja. Uticaj pojedinih supstituenata na ukupnu lipofilnost molekula definisan je Hanšovom konstantom lipofilnosti supstituenta (π). Prema Hanšovoj pretpostavci, biološka aktivnost jedinjenja odvija se u dve faze:

1. transport jedinjenja do mesta delovanja;
2. vezivanje jedinjenja za mesto delovanja.

Svaka od navedenih faza zavisi od fizičko-hemijskih osobina jedinjenja i njegovog mesta delovanja. Uzimajući u obzir prethodno navedene parametre, Hanš je definisao jednostavan matematički model koji predviđa biološku aktivnost jedinjenja ($\log 1/C$) [132]:

$$\log(1/C) = k_1 P - k_2 P^2 + k_3 \sigma + k_4 E_s + k_5 \quad (2.14)$$

(C – minimalna koncentracija jedinjenja neophodna da izazove specifični biološki odgovor, k_1 – k_5 – regresione konstante)

Savremena QSAR ispitivanja uključuju primenu *in silico* molekulkih deskriptora u predviđanju biološke aktivnosti, koji u nekim slučajevima predstavljaju efekte koje su definisali Hamet i Taft. Tako je, na primer, u slučaju QSAR modelovanja antifungale aktivnosti serije derivata benzimidazola prema *Saccharomyces cerevisiae* utvrđen značajan uticaj lipofilnosti jedinjenja ($\log P$), dipolnog momenta (DM) i površine molekula (SAG , engl. *Surface Area Grid*) na vrednost minimalne inhibitorne koncentracije (C_{MIC}) prema relacijama [133]:

$$\log(1/C_{MIC}) = 0,9134 \log P + 0,0289 DM + 0,6758 \quad (2.15)$$

$$\log(1/C_{MIC}) = 0,8491 \log P + 0,0022 SAG + 0,0174 \quad (2.16)$$



Odabir prediktorskih varijabli za formiranje QSAR modela zavisi od veličine koja se modeluje. Određeni deskriptori su pogodni za predviđanje prolaska molekula kroz lipidne membrane ($\log P$, pK_a , $\log D$, MW , SAG i dr.), dok neki mogu da budu dobri pokazatelji ponašanja molekula u farmakokinetičkim fazama (BBB , SP , $PPB\%$, $HIA\%$ i dr.). Bez obzira na veličinu koja se predviđa, bitno je obezbediti dovoljan broj validnih eksperimentalnih podataka visokog kvaliteta, kako bi se dobio ispravan QSAR model visoke prediktivnosti. Eksperimentalni rezultati ispitivanja biološke aktivnosti velikog broja jedinjenja dostupni su u internet bazama podataka, od kojih je najpoznatija ChEMBL internet baza [134].

Naučna literatura obiluje radovima baziranim na primeni QSAR analize u sferama medicinske hemije, mikrobiologije, toksikologije i hemije životne sredine. U tabeli 2.4. dat je kratak pregled literature sa primerima primene QSAR analize različitih jedinjenja. Iako je QSAR metoda već decenijama bitan faktor u analizi i predviđanju biološke aktivnosti molekula, primenom novih statističkih i matematičkih pristupa i računarskih programa, olakšava se formiranje novih QSAR modela, kao i interpretacija dobijenih rezultata.

Tabela 2.4. Kratak pregled primene QSAR analize u savremenoj naučnoj literaturi

Seriya jedinjenja	Aktivnost / osobina koja se predviđa	Referenca
Derivati benzimidazola	Antifungalna aktivnost <i>in vitro</i> prema <i>Saccharomyces cerevisiae</i> ; Antibakterijska aktivnost <i>in vitro</i> prema <i>Pseudomonas aeruginosa</i> .	[133,135,136]
Seriya benzoksazola i oksazolo[4,5- <i>b</i>]piridina	Antifungalna aktivnost <i>in vitro</i> prema <i>Candida albicans</i>	[79,137]
Seriya alkantiola	Prag mirisa	[138]
Seriya derivata maleimida	Inhibicija humane monoglicerid lipaze; molekulski doking	[139]
Set od 850 strukturno različitih jedinjenja	Biokoncentracioni faktor (BCF)	[140]
Seriya 1-[2-hidroksietoksi-metil]-6-(feniltio)timin (HEPT) derivata	Anti-HIV-1 aktivnost	[141]
Seriya derivata ksantenona	Antikancerogena aktivnost <i>in vitro</i> prema malignom gliomu (U251) i kanceru bubrega (NCI-H460)	[142]



2.7. Hemometrijska statistika

Budući da je zasnovana na statističkim i matematičkim metodama, pravilna primena hemometrijske analize zahteva poznavanje osnova ovih oblasti. Hemometrijske tehnike se načelno dele na klasifikacione i regresione [143]. Klasifikacione tehnike se primenjuju za utvrđivanje sličnosti ili razlika između jedinjenja ili uzoraka, kao i u otkrivanju određenih šablona među podacima (tehnike raspoznavanja oblika/šablona, engl. *Pattern Recognition Techniques*), dok se regresione tehnike koriste za analizu povezanosti između zavisnih i nezavisnih promenljivih. Neke od češće korišćenih klasifikacionih i regresionih hemometrijskih tehnika prikazane su u tabeli 2.5.

Tabela 2.5. Klasifikacione i regresione hemometrijske tehnike.

Klasifikacione tehnike (Tehnike raspoznavanja oblika)	Regresione tehnike
Klaster analiza	Linearna regresija
Analiza glavnih komponentata	Multipla linearna regresija
Veštačke neuronske mreže	Regresija glavnih komponentata
Faktorska analiza	Regresija metodom parcijalnih najmanjih kvadrata
Diskriminaciona analiza	Stepenasta regresija
Analiza sume razlika rangova (SRD)	Veštačke neuronske mreže

U QSRR i QSAR analizi jednostavni regresioni modeli koji predviđaju biološku aktivnost ili retenciju na osnovu jedne varijable zanemaruju složenost ovih procesa. Biološka aktivnost i hromatografska retencija nekog jedinjenja posledica su uticaja više od jednog faktora, te se za njihovo modelovanje najčešće primenjuju multivarijantne hemometrijske tehnike. Osnovni preduslov za primenu multivarijantnih tehnika je da raspoloživi podaci nose korisne informacije o veličini koja je cilj modelovanja. Pri QSRR i QSAR modelovanju mnogo je važnije obaviti pravilnu selekciju korisnih varijabli, nego fokusirati se na ukupan broj raspoloživih varijabli. U suprotnom, bez obzira na mogućnosti koje pruža primenjena hemometrijska tehnika, rezultati koji se dobiju neće sadržati traženu informaciju ili mogu da navedu na pogrešne zaključke.

Za analizu podataka u ovoj doktorskoj disertaciji od posebnog značaja su klaster analiza (CA, engl. *Cluster Analysis*), analiza glavnih komponentata (PCA, engl. *Principal Components Analysis*), analiza sume razlika rangova (SRD, engl. *Sum of Ranking Differences*), linearna i polinomska regresija, multipla linearna regresija (MLR, engl. *Multiple-Linear Regression*), regresija glavnih komponentata (PCR, engl. *Principal Components Regression*), regresija metodom parcijalnih najmanjih kvadrata (PLS, engl. *Partial Least Squares*), stepenasta regresija (SR) i regresija pomoću veštačkih neuronskih mreža (ANN, engl. *Artificial Neural Networks*).



2.7.1. Klasifikacione hemometrijske tehnike

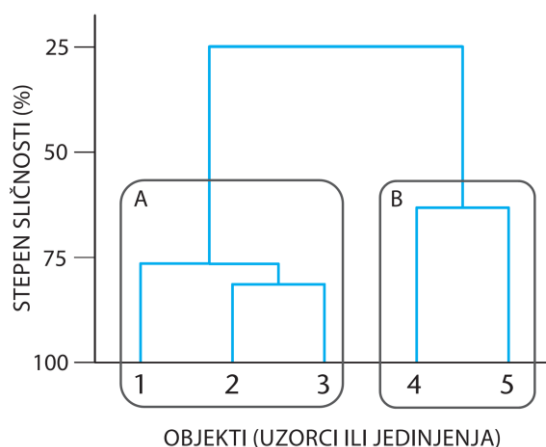
Klaster analiza (CA) je metoda multivarijantne hemometrijske analize koja služi za grupisanje sličnih objekata u klaster. Grupisanje objekata se ostvaruje na osnovu njihovih karakteristika (varijabli). Klasifikacioni pristup je numerički, a broj klastera unapred nije poznat. Rastojanje (d) između dva objekta u n -dimenzionalnom prostoru varijabli sa koordinatama x_1, x_2, \dots, x_n i y_1, y_2, \dots, y_n naziva se Euklidova distanca i definisana je jednačinom [144]:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2.17.)$$

Klasterovanje može da se obavi aglomerativnim i divizionim pristupima. Aglomerativni hijerarhijski pristup polazi od toga da svaki objekat pripada zasebnom klasteru. Pri svakom narednom koraku analize, klasteri koji pokazuju najveću sličnost udružuju se u zajednički klaster. Klasterovanje se zaustavlja kada svi objekti budu svrstani u jedan klaster. Divizionim pristupom ostvaruje se razdvajanje jedne grupe od n objekata na manje grupe. Postupak se ponavlja dok svaki od objekata ne formira zaseban klaster (suprotan smer u odnosu na aglomerativni pristup). Postoje brojni načini računanja distanci između dva klastera koja sadrže više od jednog člana. Najjednostavniji pristup je računanje rastojanja između dva najbliža objekta koja pripadaju različitim klasterima (pristup pojedinačnog povezivanja, engl. *single linkage*) [144]. Takođe, postoje i metode potpunog povezivanja (engl. *complete linkage*), prosečnog povezivanja (engl. *average linkage*), metoda centroida (engl. *centroid method*) i Vardova (*Ward*) metoda.

Rezultati klaster analize se najčešće prikazuju u vidu dendrograma, prikazanog na slici 2.14. Dendrogram ilustruje spajanje ili deljenje izvedeno u svakom koraku analize i na taj način uspešno organizuje skup podataka. Čvorovi dendrograma predstavljaju klasterne, dok je rastojanje na kojima se klasteri udružuju označeno dužinom grane. Kompletan raspored čvorova i grana čini topologiju dendrograma. Na y-osi dendrograma može da se predstavi stepen sličnosti ili različitosti objekata.

Klaster analiza je našla primenu u QSRR i QSAR metodama pri analizi sličnosti jedinjenja na osnovu njihovog retencionog ponašanja ili biološke aktivnosti [89,90,94].



Slika 2.14. Dendrogram kao vizuelni prikaz rezultata klaster analize (formirani klasteri jedinjenja ili uzoraka obeleženi su sa A i B)



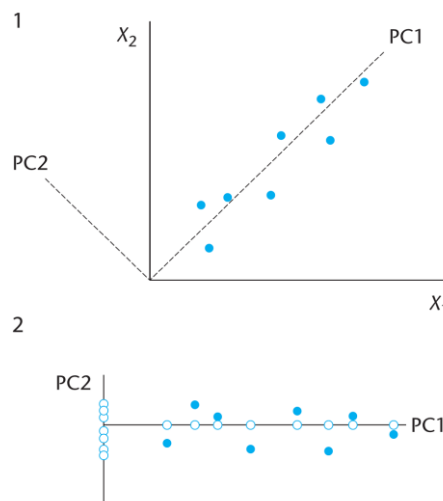
Analiza glavnih komponentata (PCA) jedna je od najčešće primenjivanih multivarijantnih tehnika za smanjenje dimenzionalnosti skupa podataka sastavljenog od velikog broja međusobno koreliranih varijabli [144,145]. PCA omogućava ekstrakciju najbitnijih informacija iz skupa podataka, olakšava opisivanje skupa podataka i pruža uvid u sličnosti i razlike između analiziranih objekata. PCA nije korisna tehnika ukoliko ne postoji korelacija između varijabli [144].

Glavni cilj PCA je definisanje glavnih komponentata ($PC1, PC2, \dots, PCn$), kao kombinacije originalnih varijabli (X_1, X_2, \dots, X_n), koje opisuju svaki od objekata (hemijska jedinjenja, uzorci):

$$PC1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n \quad (2.18.)$$

$$PC2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2n}X_n \quad (2.19.)$$

Kada se u koordinatnom sistemu, sačinjenom od onoliko osa koliko ima varijabli, rasporede objekti, PCA omogućava definisanje pravca koji na optimalan način odražava relativno rastojanje između objekata i ujedno prezentuje maksimalnu varijansu projekcije podataka na dati pravac (skorovi). Ovaj pravac se definiše kao prva glavna komponenta (PC1). Druga glavna komponenta (PC2) predstavlja pravac normalan na PC1, koji opisuje maksimum preostalog varijabiliteta. Dakle, PCA je ekvivalentna rotaciji originalnih osa tako da je PC1 postavljena u pravcu maksimalne varijabilnosti (slika 2.15.). Ukupan broj glavnih komponentata jednak je broju analiziranih varijabli. Kako broj glavnih komponentata raste, tako opada značaj informacija koje nose u datoj analizi. Glavne komponente koje se uzimaju za dalju analizu treba da objašnjavaju najmanje 80% ukupne varijanse podataka [146].



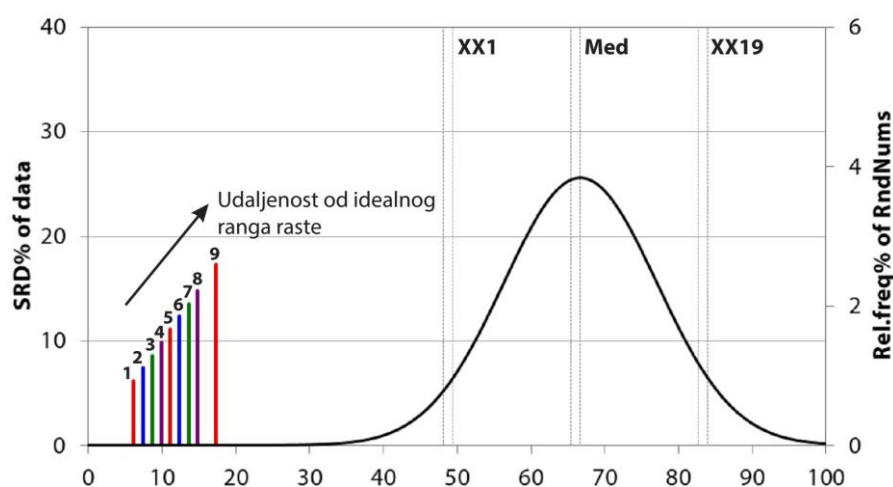
Slika 2.15. Grafik koji ilustruje dve glavne komponente (PC1 i PC2) za dve varijable (X_1 i X_2) i raspodelu objekata u prostoru ● (1) i njihovu projekciju ○ na PC1 i PC2 (2) [144]

Rezultati PCA se prikazuju pomoću grafika skorova (*scores plot*) i grafika koeficijenata latentnih varijabli (*loadings plot*). Grafik koeficijenata latentnih varijabli pokazuje koliko svaka varijabla doprinosi rasporedu (projekciji) objekata duž PC ose.

PCA nalazi široku primenu u analizi retencionih parametara i retencionih mehanizama na različitim stacionarnim fazama, kao i u analizi biološki aktivnih jedinjenja [79,90].



Analiza sume razlika rangova (SRD analiza) predstavlja relativno novu hemometrijsku tehniku. Zasnovana je na rangiranju matematičkih modela, analitičkih tehnika ili različitih uzoraka u odnosu na neki unapred postavljeni idealni referentni rang, tzv. „zlatni standard“ [147-149]. Podatke je neophodno složiti tako da se objekti (npr. jedinjenja) na osnovu kojih se izvodi rangiranje nalaze u redovima, dok kolone predstavljaju varijable (npr. matematičke modele) koje se porede. Suma razlika između idealnog referentnog ranga i ranga datog objekta računa se za svaki objekat koji se poredi i predstavlja se grafički (slika 2.16.) i tabelarno. U idealnom slučaju, odnosno u slučaju da objekat koji se poredi ima isti rang kao referentni rang, suma razlike rangova (SRD) jedanaka je nuli. Ostali objekti se rangiraju na osnovu porasta SRD vrednosti. Što je veća SRD vrednost, veća je udaljenost od idealnog ranga [148]. Kao idealni rang mogu biti zadate prosečne vrednosti (engl. *average*), minimalne ili maksimalne vrednosti varijabli za svaki objekat, eksperimentalne vrednosti ili neka drugi referentni niz.



Slika 2.16. Grafički prikaz rezultata SRD analize (1-9 – objekti (matematički modeli), XX1 – prvi ikosail (5%), Med – medijana, XX19 – poslednji ikosail)

Validacija SRD analize može se izvesti metodom poređenja rangova sa nasumičnim brojevima (CRRN, engl. *Comparison of Ranks by Random Numbers*), kao i unakrsnom validacijom [148,149]. Unakrsna validacija (CV, engl. *Cross-Validation*) se sprovodi deljenjem seta podataka na sedam celina, prilikom čega se analiza obavlja u sedam ponavljanja izostavljajući po jednu sedminu podataka pri svakom ponavljanju. Ovaj način validacije je poznat kao *7-fold unakrsna validacija*. CRRN validacija se u slučaju manjeg broja objekata ($n < 14$) izvodi primenom teorijske diskretne raspodele [148]. Normalna raspodela se koristi pri aproksimaciji teorijske (nasumične) funkcije SRD raspodele za set koji sadrži veći broj objekata ($n > 13$) [148].

SRD metoda je našla primenu kod poređenja QSRR i QSAR modela, analizi hromatografskih retencionih podataka i parametara lipofilnosti [150-153]. SRD analiza može biti primenjena i kod rangiranja molekulskih deskriptora, hemijskih jedinjenja ili različitih tipova uzoraka. Poređenjem SRD vrednosti može se obaviti i grupisanje i utvrđivanje sličnosti i razlika između rangiranih varijabli (slične varijable imaju bliske SRD vrednosti) [148].



2.7.2. Regresione hemometrijske tehnike

Cilj regresione hemometrijske analize je pronalaženje zavisnosti između zavisno i nezavisno promenljivih varijabli. Kada se ustanovi određena zavisnost, jačina te veze utvrđuje se korelacionom analizom.

Linearna regresija (LR), odnosno regresija metodom najmanjih kvadrata, jedan je od najjednostavnijih regresionih pristupa u hemometriji. Opšti oblik linearne zavisnosti može se predstaviti jednačinom:

$$y = \beta_0 + \beta_1 x \quad (2.20.)$$

(β_0 – odsečak, β_1 – nagib)

Iako je y linearna funkcija x , u većini slučajeva stvarna posmatrana vrednost y se ne nalazi tačno na pravoj liniji koja predstavlja ovu zavisnost [154]. Stoga, generalizovan linearni model uključuje i slučajnu grešku (ϵ):

$$y = \beta_0 + \beta_1 x + \epsilon \quad (2.21.)$$

Vrednosti regresionih koeficijenata u linearnom modelu (β_0 i β_1) definišu se tako da greška modela bude svedena na minimum, što se postiže metodom najmanjih kvadrata. Ovom metodom se definiše prava sa nagibom i pravcem pomoću kojih se postiže najmanja vrednost sume kvadrata ostataka, odnosno razlika između x i y vrednosti [154].

Multipla linearna regresija (MLR) uključuje više od jedne nezavisno promenljive veličine. Opšti MLR model može se prikazati jednačinom:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (2.22.)$$

Koeficijenti β_1 , β_2 se nazivaju parcijalni regresioni koeficijenti, s obzirom na to da npr. koeficijent β_1 meri očekivane promene y -vrednosti po jedinici promene x_1 -vrednosti dok se x_2 drži konstantnom, odnosno β_2 meri očekivane promene y -vrednosti po jedinici promene x_2 -vrednosti dok se x_1 drži konstantnom [154]. Generalno, koeficijent β_n predstavlja očekivanu promenu y -vrednosti po jedinici promene x_n -vrednosti dok se ostali regresori x drže konstantnim. 3D prikaz MLR modela sa dve prediktorske varijable predstavljen je na slici 2.17.

Modeli koji su kompleksniji u strukturi od modela 2.21. često mogu biti razmatrani kao oblik multiple linearne regresije. Ukoliko je u pitanju polinomski model drugog stepena:

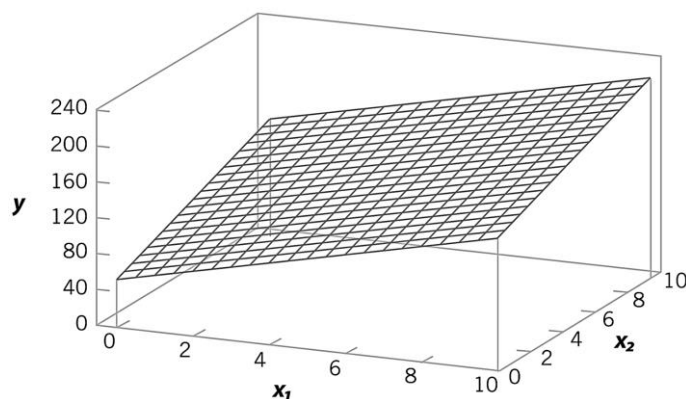
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon \quad (2.23.)$$

i ako se u njega uvrsti smena: $x_1^2 = x_2$, sledi oblik:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad (2.24.)$$



koji predstavlja klasičan oblik MLR jednačine sa dve nezavisno promenljive veličine. Stoga, polinomski model se može smatrati oblikom MLR jednačine [154].



Slika 2.17. Grafička interpretacija MLR modela sa dve nezavisno promenljive varijable (x_1 i x_2)

Prilikom formiranja QSRR i QSAR modela primenom MLR, neophodno je voditi računa o odnosu broja prediktorskih varijabli i broju podataka (broj jedinjenja, broj uzoraka, broj objekata) na osnovu kojih se formira model. Ovaj odnos je definisan pravilom Topliss-Kostelo (*Topliss-Costello*) [155]:

$$N / n \geq 5 \quad (2.25.)$$

(N – broj podataka, n – broj prediktorskih varijabli)

Velik broj prediktorskih varijabli može da destabilizuje model i da umanja njegovu sposobnost predviđanja povećavajući grešku predviđanja.

Glavno ograničenje primene MLR u definisanju QSAR i QSRR modela je *multikolinearnost*, odnosno visoka korelacija između prediktorskih varijabli. Multikolinearnost ima značajan uticaj na definisanje regresionih koeficijenata MLR modela i na njegovu potencijalnu primenu. Kod međusobno koreliranih prediktorskih varijabli veoma je teško ustanoviti koja od njih zapravo ispoljava najjači efekat na zavisno promenljivu. Multikolinearnost se može detektovati na osnovu faktora inflacije varijanse (*VIF*, engl. *Variance Inflation Factor*), koji je definisan jednačinom:

$$VIF(x_j) = 1 / (1 - R_j^2) \quad j = 1, 2, 3, \dots, k \quad (2.26.)$$

(R_j^2 – koeficijent determinacije između x_j i ostalih prediktorskih varijabli)

Ukoliko je $VIF > 1$, multikolinearnost je prisutna u modelu u određenom stepenu. U literaturi se tolerancija multikolinearnosti razlikuje, te pojedini autori smatraju da problem multikolinearnosti postaje ozbiljan ukoliko je $VIF > 5$, dok neki tolerišu VIF vrednosti koje ne prelaze vrednost 10 [154].

Postoje određene multivarijacione regresione tehnike kod kojih je multikolinearnost osnovni uslov za njihovu primenu. U ovu grupu spadaju regresija glavnih komponenata i regresija pomoću metode parcijalnih najmanjih kvadrata.



Regresija glavnih komponentata (PCR) predstavlja tehniku pogodnu za regresiono modelovanje podataka koji su korelirani u velikom stepenu. Ukoliko je prisutna multikolinearnost kod MLR modela, veoma je verovatno da će procenjene vrednosti regresionih koeficijenata biti pogrešne. Međutim, u slučaju PCR multikolinearnost je preduslov za dobijanje ispravnog modela [144].

PCR je zasnovana na primeni skorova glavnih komponentata nezavisno promenljivih varijabli i predstavlja kombinaciju MLR i PCA [156]. Suštinski, PCR predstavlja način određivanja koeficijenata u MLR jednačini. Nezavisno promenljive su u slučaju PCR linearna kombinacija originalnih (x) varijabli. PCR ima dve glavne prednosti [156]:

1. skorovi su ortogonalni, tako da se problem koreliranih varijabli anulira;
2. broj relevantnih glavnih komponentata je uglavnom manji od broja originalnih varijabli, te se na taj način postiže redukcija broja nezavisno promenljivih.

Prvi korak PCR analize čini PCA, koja transformiše originalnu matricu podataka \mathbf{X} u matricu \mathbf{T} . U drugom koraku, matrica \mathbf{T} se koristi u MLR modelu [145]:

MLR model izražen preko originalne matrice \mathbf{X} :

$$y = \beta \mathbf{X} + \varepsilon \quad (2.27.)$$

PCR model izražen preko matrice \mathbf{T} :

$$y = \beta \mathbf{T} + \varepsilon \quad (2.28.)$$

Broj glavnih komponentata koje se primenjuju u PCR modelu bira se tako da one na najbolji mogući način predviđaju zavisno promenljivu veličinu. Pri odabiru njihovog broja procenat objašnjene varijanse u nezavisno promenljivim varijablama nije presudan, te se kao značajne glavne komponente mogu odabrati i one koje ne doprinose varijabilnosti u značajnom procentu [145].

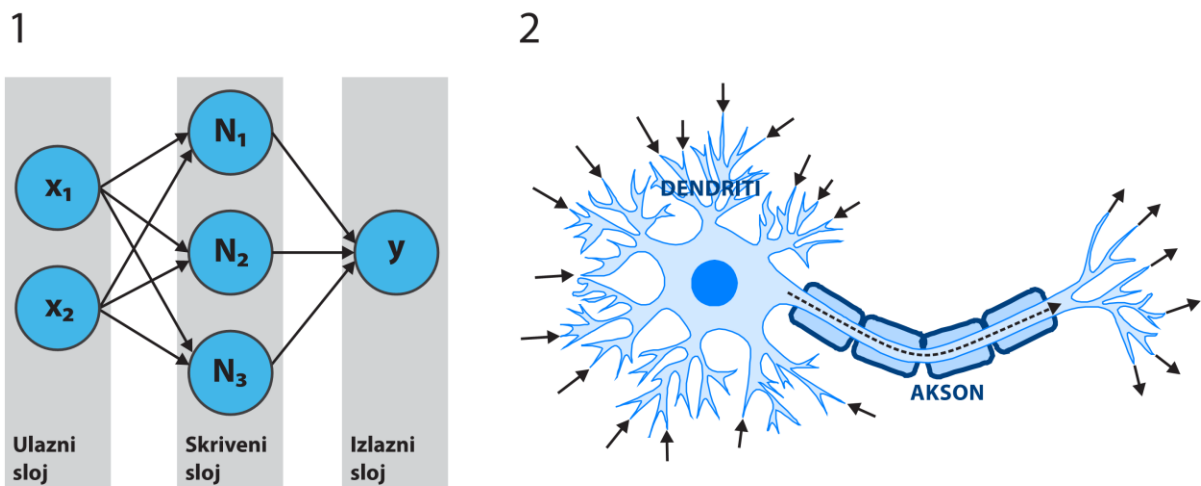
Regresija metodom parcijalnih najmanjih kvadrata (PLS) je vrlo slična PCR metodi. PLS regresija takođe primenjuje linearnu kombinaciju prediktorskih varijabli koje se koriste za MLR regresiju sa nezavisno promenljivom, međutim PCR i PLS se razlikuju po načinu odabira varijabli koje figurišu u modelu. U PCR analizi glavne komponente su odabrane tako da opisuju što je više moguće varijacija između nezavisno promenljivih varijabli, bez obzira na jačinu njihove korelacije sa zavisno promenljivom, dok u slučaju PLS analize varijable koje su visoko korelirane sa zavisno promenljivom dodatno dobijaju na značaju [144,157]. Na ovaj način linearne kombinacije prediktorskih varijabli su kod PLS regresije odabrane tako da su visoko korelirane sa zavisno promenljivom i da istovremeno objašnjavaju što više varijacija između nezavisno promenljivih [144]. Dok se u PCR primenjuju skorovi glavnih komponentata dobijenih na osnovu matrice \mathbf{X} , u PLS regresiji se koriste glavne komponente visoko korelirane sa y . Kod oba pristupa, PCR i PLS, očekuje se da će svega nekoliko linearnih kombinacija nezavisno promenljivih varijabli biti dovoljno da se opiše što više varijacija. Jedan od najčešće korišćenih algoritama za formiranje PLS regresionog modela je NIPALS (engl. *Non-linear iterative partial least squares*), uključujući i neke njegove modifikacije i proširenja.

PLS regresiona metoda na neki način predstavlja kompromis između MLR i PCR, s obzirom na to da obuhvata veći deo varijanse matrice \mathbf{X} i značajne vrednosti koeficijenta korelacije sa y . MLR, PCR i PLS



su linearne regresione tehnike koje se međusobno razlikuju u regresionim koeficijentima i u sposobnosti predviđanja zavisno promenljive. MLR, PCR i PLS su multivarijantne tehnike koje se najčešće primenjuju pri formiranju QSRR i QSAR modela [79-81,158,159].

Regresija pomoću veštačkih neuronskih mreža (ANN) je hemometrijska tehnika pogodna za modelovanje kompleksnih veza između varijabli. Veštačka neuronska mreža se sastoji od veštačkih neurona sa implementiranom suštinom biološkog neurona. Ona može biti sastavljena iz nekoliko slojeva (slika 2.18.(1)). Ulazni sloj čine nezavisno promenljive varijable, koje su povezane sa neuronima u skrivenom sloju. Mreža se može sastojati iz jednog ili više skrivenih slojeva. Izlazni sloj čini jedna ili više izlaznih varijabli (zavisno promenljivih). Svaki neuron, koji se još naziva i skrivena jedinica, kao činilac mreže prima signal od prethodnog neurona, transformiše ga datom aktivacionom (transfer) funkcijom i prenosi na sledeći neuron, analogno putu električnog signala kroz biološki neuron, od dendrita (ulaza) preko aksona do izlaza, odnosno sinapsi, što je šematski prikazano na slici 2.18.(2). Aktivaciona funkcija kontroliše propagaciju neuronskog signala do sledećeg sloja, tako da pozitivne težine simuliraju ekscitatorni stimulus, a negativne težine inhibitorni stimulus [160]. Veštački neuron sadrži regresionu jednačinu koja procesuirala ulaznu informaciju u nelinearni izlazni podatak. Stoga, neuronske mreže se smatraju nelinearnom regresionom metodom [78,160]. Aktivaciona funkcija je za skrivene slojeve uglavnom nelinearna (sinusoidna, hiperbolična, prag funkcija i dr.), dok je za izlazni sloj karakteristična linearna funkcija [83]. Sličnost veštačkih neurona sa biološkim ogleda se u njihovoj sposobnosti da uče na osnovu iskustva, zaključuju i izdvajaju bitne informacije.



Slika 2.18. Jednostavni model veštačke neuronske mreže sastavljene od ulaznog sloja sa dve ulazne varijable (x_1 i x_2), skrivenog sloja sa tri neurona (N_1 , N_2 i N_3) i izlaznog sloja sa jednom izlaznom varijablom (y) (1) i struktura biološkog neurona kao jedinice građe nervnog sistema (strelice pokazuju smer kretanja električnog impulsa kroz neuron) (2)



Postoji nekoliko tipova neuronskih mreža. Na osnovu arhitekture dele se na jednoslojne (SLP, engl. *Single Layer Perceptron*) i višeslojne (MLP, engl. *Multi-Layer Perceptron*). Postoje modeli ANN arhitekture sa propagacijom signala unapred (nerekurentne neuronske mreže, engl. *feedforward*) ili sa povratnom spregom (rekurentne neuronske mreže, engl. *feedback*).

Neophodno je istaći da je ponekad pre primene ANN metode potrebno obaviti normalizaciju (skaliranje) podataka [161]. Normalizacija se najčešće primenjuje na ulazne podatke, ukoliko se sastoje od nekoliko različitih varijabli čije se vrednosti razlikuju u opsegu koji obuhvataju. Normalizacijom se izbegava slučaj da jedna varijabla, koja ima širok opseg vrednosti, potpuno zamaskira varijacije u drugoj varijabli sa mnogo manjim opsegom vrednosti. Takođe, u nekim slučajevima pogodno je normalizovati i izlazne podatke kojima se obavlja treniranje mreže [162-164]. Jedan od načina normalizacije podataka je tzv. *min-max* normalizacija, koja se izvodi primenom relacije:

$$y_{\text{norm}} = (1 - \Delta^U - \Delta^L) \cdot (y - y_{\text{min}}) / (y_{\text{max}} - y_{\text{min}}) + \Delta^L \quad (2.29.)$$

gde su Δ^U i Δ^L vrednosti margina koje ograničavaju ekstrapolacionu sposobnost mreže, y_{norm} , y_{max} i y_{min} normalizovana, maksimalna i minimalna vrednost (ne)zavisno promenljive. Ukoliko podaci nisu normalizovani, proces treniranja (učenja) mreže može biti veoma spor. Treniranje mreže može se izvesti primenom različitih algoritama, kao što su Brojden-Flečer-Goldfarb-Shanno (*Broyden-Fletcher-Goldfarb-Shanno*) algoritam, algoritam opadanja gradijenta (engl. *Gradient Descent Algorithm*), Levenberg-Markvard (*Levenberg-Marquardt*) algoritam i dr. Prilikom ANN modelovanja pomoću programa Statistica, podaci se dele na tri seta – kalibracioni, validacioni i test set. Kalibracioni set podataka se koristi za formiranje modela. Test set se koristi za definisanje opšte greške modela, dok validacioni set služi za pronalaženje najbolje konfiguracije mreže poređenjem greške validacije i greške treniranja mreže [162].

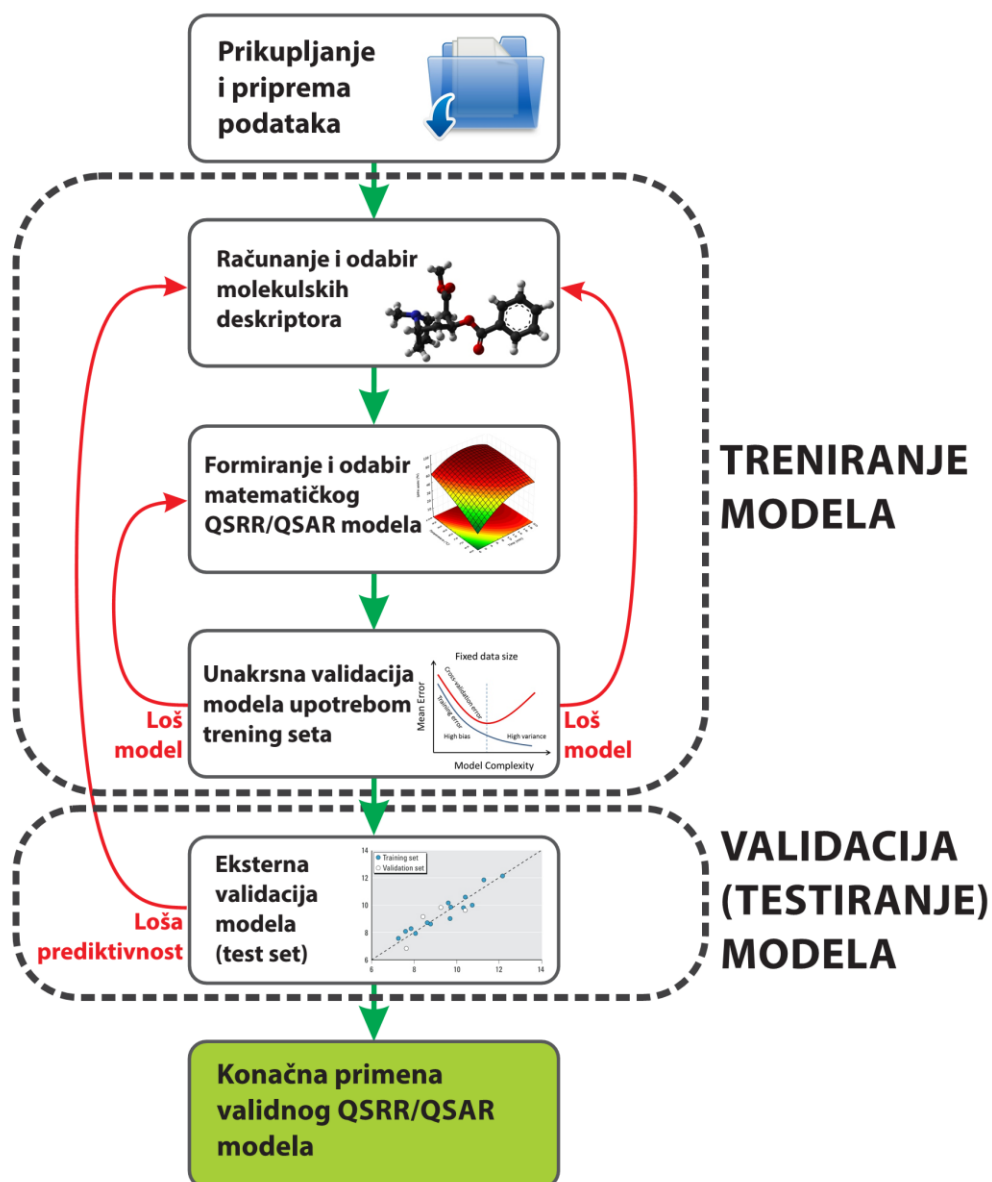
Uprkos širokoj primeni u mnogim oblastima, ANN regresiona metoda se i dalje smatra „crnom kutijom“, s obzirom na to da je prilično teško razumeti tačnu matematičku vezu između ulaznih i izlaznih podataka [160]. Kako bi se prevazišao ovaj problem, primenjujane su različite numeričke metode za analizu međusobnog odnosa varijabli. Analiza globalne osetljivosti (GSA, engl. *Global Sensitivity Analysis*) je jedna od najčešće primenjivanih metoda u tu svrhu [165,166]. Ovom metodom se analizira uticaj svake ulazne varijable na varijacije vrednosti parametara ANN modela, koje imaju uticaj na vrednosti izlazne varijable. Visoka vrednost GSA koeficijenta ulazne varijable ukazuje na to da se učinak ANN modela može značajno promeniti sa malim varijacijama njenih vrednosti. Ukoliko je vrednost GSA koeficijenta određene varijable manja od 1, varijabla treba da bude izuzeta iz modela [78].

Budući da je ANN regresija specijalni slučaj neparametarske regresije, ne postoji mogućnost definisanja parametarske jednačine, te se u QSRR i QSAR analizama ANN model prikazuje u vidu arhitekture mreže sa osnovnim statističkim podacima. Kao metoda hemometrijskog pristupa, neuronske mreže su našle primenu u predviđanju antifungalne aktivnosti [137], antikancerogene aktivnosti [78] i retencionog ponašanja različitih jedinjenja [167,168].



2.7.3. Formiranje i validacija QSRR i QSAR modela

Tok hemometrijske regresione analize se sastoji od nekoliko važnih koraka prikazanih šematski na slici 2.19. Prvi korak podrazumeva prikupljanje i pripremu eksperimentalnih podataka, dok se u drugom koraku izvodi modelovanje molekula i računanje molekulskih deskriptora. Savremeni računarski programi omogućavaju računanje izuzetno velikog broja molekulskih deskriptora, koji se koriste kao prediktorske varijable u QSRR i QSAR modelima. Najčešći problem u ovom koraku je odabir adekvatnih molekulskih deskriptora, koji se prevazilazi matematičkim i statističkim metodama.



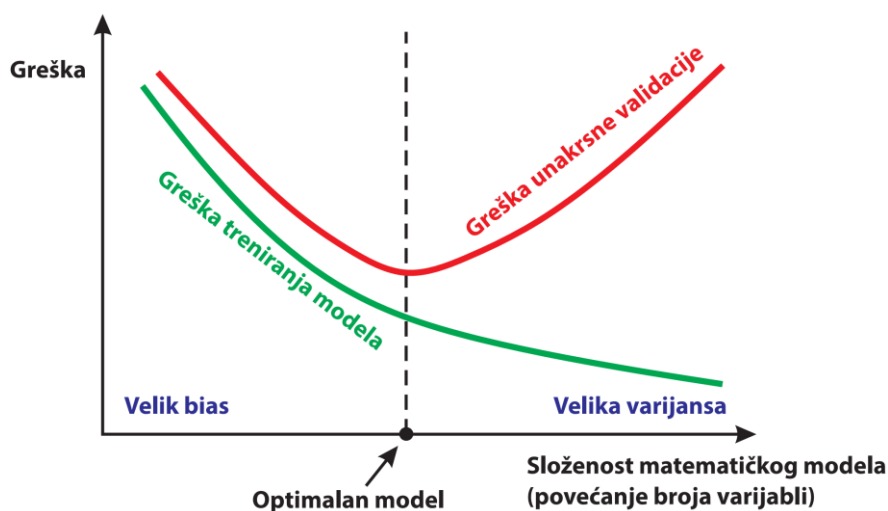
Slika 2.19. Glavni tok QSRR i QSAR modelovanja [169]



Neke od najčešće korišćenih metoda sa selekciju nezavisno promenljivih varijabli (molekulskih deskriptora) u QSRR i QSAR modelovanju su [170-172]:

- regresija svih promenljivih (engl. *all possible regression*) – metoda zasnovana na proveravanju korelacije između zavisno promenljive i jedne ili više nezavisno promenljivih, a odabir varijabli se obavlja na osnovu vrednosti koeficijenta korelacije i standardne greške;
- selekcija unapred (engl. *forward selection*) – prvi model se formira sa varijablom koja je najviše korelirana sa zavisno promenljivom (najviši koeficijent korelacije), dok svaka naredna dodata varijabla doprinosi povećanju koeficijenta korelacije, sve dok se ne uključi varijabla koja ne doprinosi značajno njegovoj promeni;
- eliminacija unazad (engl. *backward elimination*) – u prvom koraku se formira model koji sadrži sve varijable, nakon čega se eliminišu varijable sa najmanjim doprinosom vrednosti koeficijenta korelacije;
- postepena selekcija (engl. *stepwise selection*) – ovaj metod kombinuje selekciju unapred i eliminaciju unazad, tako što započinje dodavanjem varijabli u model, dok u svakom narednom stepenu postoji mogućnost eliminacije varijable iz modela (neophodno je postaviti dva kriterijuma, od kojih se jedan odnosi na dodavanje varijabli, a drugi na njihovu eliminaciju);
- PLS-VIP selekcija zasnovana na odabiru varijabli značajnih za PLS projekciju (*VIP*, engl. *Variable Importance in Projection*) – analiza započinje izvođenjem PLS analize, koja kao jedan od rezultata daje *VIP* vrednosti svih analiziranih varijabli. Varijable čija *VIP* vrednost prelazi definisanu granicu U , $U \in [0, \infty)$, ulaze u model;
- SRD metoda – zasnovana je na izdvajanju deskriptora koji nose slične (iste) informacije, npr. pri odabiru adekvatnog parametara lipofilnosti iz grupe od nekoliko izračunatih.

Broj varijabli koje će formirati model može da ima uticaj na sposobnost predviđanja modela, pogotovo u slučaju MLR. Na slici 2.20. može se uočiti da sa porastom kompleksnosti MLR modela greška treniranja kontinualno opada, međutim nakon dodatka određene varijable greška validacije počinje da raste, a samim tim i greška predviđanja.



Slika 2.20. Promena greške MLR modela sa porastom broja varijabli



Nakon uspešno izvedenog QSRR ili QSAR modelovanja, potrebno je odabrati modele koji ispunjavaju određene statističke kriterijume. Odabir modela se u prvom koraku najčešće obavlja na osnovu klasičnih statističkih parametara, kao što su:

- Pirsonov koeficijent korelacije (R);
- koeficijent determinacije (R^2) i koeficijent alijenacije (R^2_a);
- podešeni koeficijent determinacije (R^2_{adj});
- F -vrednost (Fišerov test);
- koren srednje kvadratne greške ($RMSE$, engl. *Root Mean Square Error*);
- p -vrednost (verovatnoća).

Nakon odabira modela sa odgovarajućim statističkim karakteristikama, neophodno je obaviti njihovu validaciju. Unakrsna validacija najčešće je primenjena metoda interne validacije [157]. Zasniva se na iterativnom izdvajanju različitih proporcija objekata izvan kalibracionog seta upotrebljenog za formiranje modela. Formirani model potom koristi izdvojeni set objekata, kako bi se omogućila procena njegove sposobnosti predviđanja. U zavisnosti od broja objekata izdvojenih iz kalibracionog seta, unakrsna validacija se obeležava kao:

- unakrsna validacija „izostavi jednog iz modela“ (LOO-CV, engl. *Leave-One-Out Cross-Validation*);
- unakrsna validacija „izostavi više iz modela“ (LMO-CV, engl. *Leave-More-Out Cross-Validation*).

Parametri unakrsne validacije su:

- koeficijent determinacije unakrsne validacije (R^2_{CV});
- ukupna suma kvadrata (TSS , engl. *Total Sum of Squares*);
- suma kvadrata odstupanja od predviđenih vrednosti ($PRESS$, engl. *Predicted Residual Error Sum of Squares*);
- $PRESS/TSS$ odnos;
- standardna devijacija predviđanja (SD_{PRESS}).

Jedan od najrelevantnijih načina procene prediktivnosti modela je eksterna validacija [173-177]. Eksterna validacija se izvodi tako što se model primeni na određeni broj objekata (eksterni test set) koji nisu učestvovali u njegovom formiranju. Eksterni test set se uglavnom formira izdvajanjem 15-30% objekata iz polaznog seta podataka. Statistički parametri eksterne validacije ($RMSE_{test}$, R_{test} , R^2_{test}) su često dobar pokazatelj validnosti formiranog modela [175-177].

Kod PCR i PLS regresione analize, najznačajniji indikatori kvaliteta modela su R^2Y_{kumul} (kumulativna suma kvadrata zavisno promenljive objašnjena pomoću izdvojenih komponenti) i Q^2Y_{kumul} (kumulativni deo ukupne varijacije zavisno promenljive koji može biti predviđen pomoću izdvojenih komponenti) [92]. U tabeli 2.6. date su idealne i prihvatljive vrednosti određenih validacionih parametara QSRR i QSAR modela.

**Tabela 2.6.** Idealne i prihvatljive vrednosti validacionih parametara QSRR i QSAR modela [157,173-180]

Parametar	Idealna vrednost	Prihvatljiva vrednost
R	1,00	$R > 0,80$ ukazuje na veoma jaku korelaciju
R^2	1,00	$R^2 > 0,64$ ukazuje na veoma jaku korelaciju
R^2_a	0,00	$R^2_a < 0,36$ ukazuje na veoma jaku korelaciju
R^2_{adj}	1,00	$R^2_{adj} > 0,70$
F -vrednost	Velika vrednost ukazuje na dobro slaganje (fitovanje) podataka	
$RMSE$	0,00	Zavisí od prirode podataka
p -vrednost	0,000000	Zavisí od nivoa značajnosti
R^2_{cv}	1,00	$R^2_{cv} > 0,60$
$ R^2 - R^2_{cv} $	0	$ R^2 - R^2_{cv} < 0,10$
TSS	Zavisí od prirode podataka	
$PRESS$	Zavisí od prirode podataka	
$PRESS/TSS$	0 ($PRESS \ll TSS$)	$PRESS/TSS < 0,40$
SD_{PRESS}	0	Zavisí od prirode podataka
$R^2 Y_{kumul}$	1,00	$R^2 Y_{kumul} > 0,80$
$Q^2 Y_{kumul}$	1,00	$Q^2 Y_{kumul} > 0,60$

Još jedan važan aspekt validacije QSRR i QSAR modela je poređenje eksperimentalnih i predviđenih vrednosti i analiza njihovih razlika (ostataka). Idealan slučaj poređenja eksperimentalnih i predviđenih vrednosti podrazumeva linearnu zavisnost oblika $y_{(eksperimentalno)} = y_{(predviđeno)}$, što podrazumeva da su svi ostaci jednaki nuli. Dakle, model je kvalitetniji ukoliko je nagib linearne zavisnosti eksperimentalno–predviđeno bliži jedinici, odnosno odsečak bliži nuli. Posebna pažnja se mora obratiti na nasumičnost distribucije ostataka oko $y = 0$ ose zavisnosti ostataka od predviđenih vrednosti. U idealnom slučaju vrednost koeficijenta korelacije ove zavisnosti jednak je nuli, jer vrednosti ostataka treba da budu nepredvidive. Još jedan korak u analizi ostataka zasnovan je na utvrđivanju normalnosti njihove raspodele. Ukoliko ostaci nisu normalno distribuirani, tada zavisna varijabla ili najmanje jedna prediktorska varijabla može imati pogrešnu funkcionalnu formu ili su bitne varijable izostavljene iz modela, što uzrokuje sistematsku grešku. Normalna raspodela ostataka i odstupanja od normalne raspodele lako se uočavaju pomoću grafika verovatnoće normalne raspodele (engl. *normal probability plot*). Nasumičnost ostataka se može testirati Vald-Volfovic testom (*Wald-Wolfowitz Run test – WWR*) [144]. Ovaj test se zasniva na definisanju blokova susednih ostataka istog predznaka na grafiku zavisnosti ostataka od predviđenih vrednosti. Ukoliko je broj blokova manji od tabelarno ili računski definisane vrednosti, hipoteza o nasumičnoj raspodeli podataka se odbacuje. Ovaj test se može primeniti i pri analizi pripadnosti objekata istoj populaciji. Formule za računanje parametara unakrsne validacije QSRR i QSAR modela date su u *prilogu (jednačine P-7.1. – P-7.4.)*. Detaljniji opis navedenih statističkih metoda dostupan je u literaturi [77,144-146,154,157,160,162].



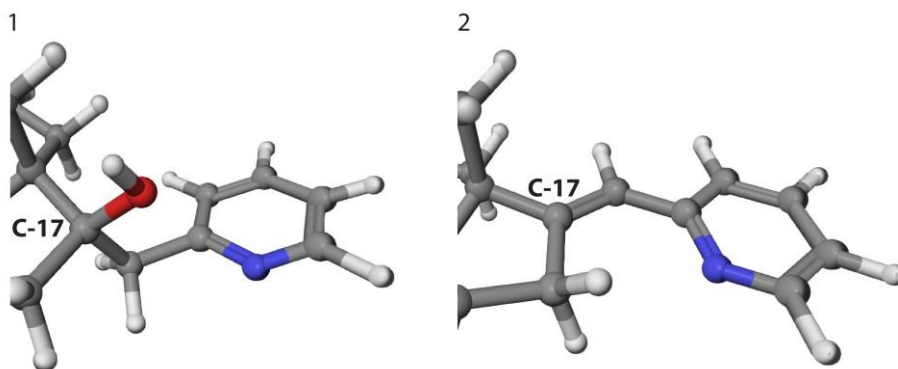
3. Eksperimentalni deo

3.1. Serija ispitivanih androstanskih derivata

Ispitivana jedinjenja su derivati androstana sa modifikacijama na A, B i D prstenovima steroidnog jezgra. U zavisnosti od vrste funkcionalne grupe na C-17 atomu, jedinjenja su podeljena u dve grupe:

1. grupa 17-pikolil androstanskih derivata (jedinjenja **1-11**);
2. grupa 17-pikoliniliden androstanskih derivata (jedinjenja **12-24**).

Njihove strukturne formule i IUPAC nazivi dati su u *tabeli 3.1*. Prstenovi A i B su modifikovani hidroksidnim, acetoksi, epoksi, okso, nitro i metoksi funkcionalnim grupama, kao i izolovanim ili konjugovanim dvostrukim vezama. Kod nekih jedinjenja prisutna je i N-oksida funkcija u piridinskom prstenu (piridin-N-oksida). Pikolil grupa se kod svih ispitivanih derivata nalazi u α -položaju, dok je hidroksidna grupa koja je vezana za isti C-17 atom postavljena u β -položaj. Pikoliniliden funkcija je vezana za C-17 atom u *E*-konformaciji. Stoga, ispitivane grupe jedinjenja se mogu približnije definisati kao 17 α -pikolil i 17(*E*)-pikoliniliden derivati androstana (*slika 3.1*).



Slika 3.1. Položaj 17 α -pikolil (1) i 17(*E*)-pikoliniliden (2) grupe u molekulima ispitivanih derivata

Ispitivani androstanski derivati su sintetisani na Katedri za organsku hemiju, Departmana za hemiju, biohemiju i zaštitu životne sredine, Prirodno-matematičkog fakulteta, Univerziteta u Novom Sadu. Postupci sinteze i karakterizacija jedinjenja detaljno su opisani u literaturi [37,40-43,181]. U literaturi je predstavljena i analiza reaktivnosti određenih 17-pikolil derivata androstana [182], kao i analiza njihovih interakcija sa različitim rastvaračima [183].

**Tabela 3.1.** Hemijske strukture ispitivanih 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata i njihovi IUPAC nazivi

Oznaka jedinjenja	Hemijska struktura	IUPAC naziv
1		17α -pikolil-androst-5-en- 3β , 17β -diol
2		17β -hidroksi- 17α -pikolil-androst-5-en- 3β -il acetat
3		17β -hidroksi- 17α -pikolil-androst-4-en-3-on
4		4β , 5β -epoksi- 17β -hidroksi- 17α -pikolil-androstan-3-on
5		5α , 6α -epoksi- 17α -pikolil-androstan-N-oksid- 3β , 17β -diol
6		17α -pikolil-androst-4-en- 3β , 17β -diol



Tabela 3.1. Hemijske strukture ispitivanih 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata i njihovi IUPAC nazivi (*nastavak*)

Oznaka jedinjenja	Hemijska struktura	IUPAC naziv
7		17 β -hidroksi-17 α -pikolil-androst-4-en-3 β -il acetat
8		17 β -hidroksi-17 α -pikolil-androst-4-en-3,6-dion
9		4,17 β -dihidroksi-17 α -pikolil-androsta-4,6-dien-3-on
10		17 α -pikolil-androsta-3,5-dien-17 β -ol
11		17 α -pikolil-androst-5-en-3 β ,4 β ,17 β -triol
12		17(<i>E</i>)-pikoliniliden-androst-5-en-3 β -il acetat

**Tabela 3.1.** Hemijske strukture ispitivanih 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata i njihovi IUPAC nazivi (*nastavak*)

Oznaka jedinjenja	Hemijska struktura	IUPAC naziv
13		17(<i>E</i>)-pikoliniliden-androst-5-en-3β-ol
14		17(<i>E</i>)-pikoliniliden-androst-4-en-3-on
15		4β,5β-epoksi-17(<i>E</i>)-pikoliniliden-androstan-3-on
16		17(<i>E</i>)-pikoliniliden-androst-4-en-3-on-N-oksid
17		4-nitro-17(<i>E</i>)-pikoliniliden-androst-4-en-3β-il acetat
18		3β-acetoksi-17(<i>E</i>)-pikoliniliden-androst-5-en-N-oksid



Tabela 3.1. Hemijske strukture ispitivanih 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata i njihovi IUPAC nazivi (*nastavak*)

Oznaka jedinjenja	Hemijska struktura	IUPAC naziv
19		4-metoksi-17(<i>E</i>)-pikoliniliden-androst-4-en-3-on
20		17(<i>E</i>)-pikoliniliden-androst-4-en-3 β -ol
21		17(<i>E</i>)-pikoliniliden-androst-4-en-3 β -il acetat
22		17(<i>E</i>)-pikoliniliden-androst-4-en-3,6-dion
23		4-hidroksi-17(<i>E</i>)-pikoliniliden-androsta-4,6-dien-3-on
24		17(<i>E</i>)-pikoliniliden-androsta-3,5-dien



3.2. Hromatografska analiza androstanskih derivata

Analiza hromatografskog ponašanja androstanskih derivata u hromatografiji na obrnutim fazama obavljena je u Laboratoriji za tehnologiju, kvalitet i bezbednost hrane – FINSLab, Naučnog instituta za prehrambene tehnologije u Novom Sadu (FINS).

Definisanje hromatografske lipofilnosti serije ispitivanih jedinjenja izvedeno je korišćenjem hromatografske kolone *ZORBAX Eclipse XDB-C18, 4,6 x 50 mm, 1,8 micron*, pomoću uređaja za visokoefikasnu tečnu hromatografiju *Agilent Technologies 1200 Series HPLC* sa *Diode Array* detektorom (DAD) i *Evaporative Light Scattering* detektorom (ELSD) (slika 3.2.). Uređaj je opremljen binarnom pumpom, degazerom i automatskim injektorom. Sistem je povezan sa računarom pomoću koga je obavljena obrada podataka uz korišćenje *AgilentChemStation* programa.



Slika 3.2. RP-HPLC-DAD-ELSD aparatura pomoću koje je izvedeno definisanje hromatografske lipofilnosti androstanskih derivata (Naučni institut za prehrambene tehnologije u Novom Sadu, FINS)

Pre hromatografske analize, pripremljeni su rastvori prečišćenih uzoraka androstanskih derivata koncentracije 2 mg/cm^3 u metanolu. Rastvori su potom filtrirani pomoću filtera *Captiva Econofilter, nylon membrane, 25 mm diameter, 0,45 μm pore size, 1000/pk*. Hromatografska analiza je izvedena izokratskim postupkom uz upotrebu mobilne faze metanol/voda sa zapreminskim udelima komponenata 90 : 10 i 70 : 30, pri protoku kroz kolonu od $0,600 \text{ cm}^3/\text{min}$. Vrednost pH mobilne faze održavana je na 7 pomoću 0,01 M fosfatnog pufera (Na_2HPO_4 , KH_2PO_4). Temperatura kolone je tokom analize održavana konstantnom na $25 \text{ }^\circ\text{C}$. Injekciona zapremina uzorka iznosila je $10 \text{ }\mu\text{L}$. Detekcija



jedinjenja pomoću DAD detektora obavljena je na 210 i 230 nm. Temperatura i pritisak ELSD detektora održavani su konstantnim na 40 °C i 4,5 bar, respektivno. Merenje retencije ispitivanih jedinjenja u primenjenom hromatografskom sistemu izvedeno je u tri ponavljanja. Retenciono vreme, koje je izmereno pomoću DAD detektora na 210 nm, upotrebjeno je za računanje kapacitetnog faktora, dok je signal dobijen pomoću ELSD detektora korišćen kao kontrolni signal.

Pri izvođenju hromatografske analize korišćene su hemikalije:

- metanol (BAKER HPLC Analyzed®, HPLC gradient grade);
- Na₂HPO₄ · 12H₂O (*p.a.* Lach-Ner, Češka);
- KH₂PO₄ (*p.a.* Lach-Ner, Češka);
- destilovana voda.

Retenciono vreme (t) supstanci je očitano sa hromatograma, a njihov kapacitetni faktor (k) je računat pomoću *jednačine 2.6*. Mrtvo retenciono vreme (t_m) definisano je prvim registrovanim signalom. Dobijeni retencioni podaci ($\log k$) su upotrebljeni kao zavisno promenljiva veličina pri formiranju QSRR modela.

3.3. Biološka aktivnost androstanskih derivata

Ispitivanje citotoksične (antiproliferativne) aktivnosti androstanskih derivata prema različitim ćelijskim linijama kancera predstavljeno je u literaturi [41-43]. Ovim ispitivanjima je utvrđeno da sintetisani derivati ispoljavaju značajnu citotoksičnu aktivnost.

U ovom radu su pri formiranju QSAR modela kao zavisno promenljive veličine korišćeni podaci za citotoksičnu aktivnost 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata prema ćelijskim linijama androgen-receptor negativnog kancera prostate (AR-neg. PC-3). Njihova citotoksična aktivnost prema AR-neg. PC-3 ćelijskoj liniji izražena je pomoću IC₅₀ vrednosti, koja se definiše kao polovina maksimalne inhibitorne koncentracije neke supstance, odnosno koncentracija pri kojoj se ispoljava polovina maksimalnog inhibitornog dejstva nekog jedinjenja. IC₅₀ vrednost se koristi kao mera efikasnosti jedinjenja u inhibiciji specifične biološke funkcije. Prema opšte usvojenim kriterijumima, citotoksična aktivnost nekog jedinjenja se klasifikuje kao [37]:

- izrazito jaka (IC₅₀ < 10 μ M);
- zadovoljavajuća (10 μ M < IC₅₀ < 20 μ M);
- umerena ili slaba (20 μ M < IC₅₀ < 100 μ M);
- zanemarljiva (IC₅₀ > 100 μ M).

Kompletna procedura analize antikancerogene aktivnosti 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata prema AR-neg. PC-3 ćelijskoj liniji opisana je u literaturi [41-43]. S obzirom na to



da je citotoksična aktivnost jedinjenja **2**, **3**, **12**, **13**, **17** i **23** zanemarljiva, ova jedinjenja nisu uzeta u obzir pri formiranju QSAR modela za predviđanje citotoksične aktivnosti. QSAR modelovanje je zasnovano na devet 17 α -pikolil i devet 17(*E*)-pikoliniliden derivata androstana, čije su IC₅₀ vrednosti date u *tabeli 3.2*.

Tabela 3.2. Rezultati *in vitro* ispitivanja citotoksičnosti 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata prema ćelijskoj liniji AR-neg. PC-3 izražene pomoću IC₅₀ vrednosti

Jedinjenje	IC ₅₀ (μ M)	Referenca
1	6,30	[41]
4	65,50	[41]
5	33,60	[43]
6	22,30	[43]
7	31,30	[43]
8	4,30	[43]
9	25,60	[43]
10	24,70	[43]
11	52,10	[43]
14	12,90	[41]
15	14,30	[41]
16	10,10	[41]
18	0,55	[41]
19	34,10	[43]
20	10,10	[42]
21	66,20	[42]
22	88,90	[42]
24	19,90	[43]



3.4. Računarsko modelovanje molekula i hemometrijska obrada podataka

Modelovanje hemijske strukture molekula i računanje molekulskih deskriptora izvedeno je odgovarajućim programima za molekulski dizajn:

- ChemBioDraw Ultra 14.0 [60];
- ChemBio3D Ultra 14.0 [60];
- MarvinSketch 6.1 [184];
- PreADMET online program [70];
- Simulation Plus ADMET Predictor™ [185];
- Molinspiration online program [186];
- Parameter Client Program [187];
- ADRIANA.Code 2.2.6 [188];
- ALOGPS 2.1 online program [189];
- HyperChem Professional 8.0 [190];
- Avogadro 1.0 [191];
- Bioluminate® 1.7 [61].

Računanje molekulskih deskriptora je obavljeno na osnovu 1D, 2D i 3D molekulskih struktura. Optimizacija 3D strukture postignuta je primenom MM2 metode minimizacije energije. Optimizacija je obavljena pomoću programa ChemBio3D Ultra (Austin Model 1 – AM1). Kao granica RMS (engl. *Root Mean Square*) greške zadata je vrednost od 0,0001 kcal/Å · mol uz primenu programskog dodatka MOPAC.

Kompletna hemometrijska analiza podataka izvedena je pomoću programa:

- Statistica v. 10 [192];
- NCCS 2007 [193];
- Microsoft Excel [194];
- OriginPro 9.1 [195];
- Minitab 16.1.1 [196];
- MATLAB R2013a [197].

Za SRD analizu primenjen je program u okviru Microsoft Excel softvera [198].



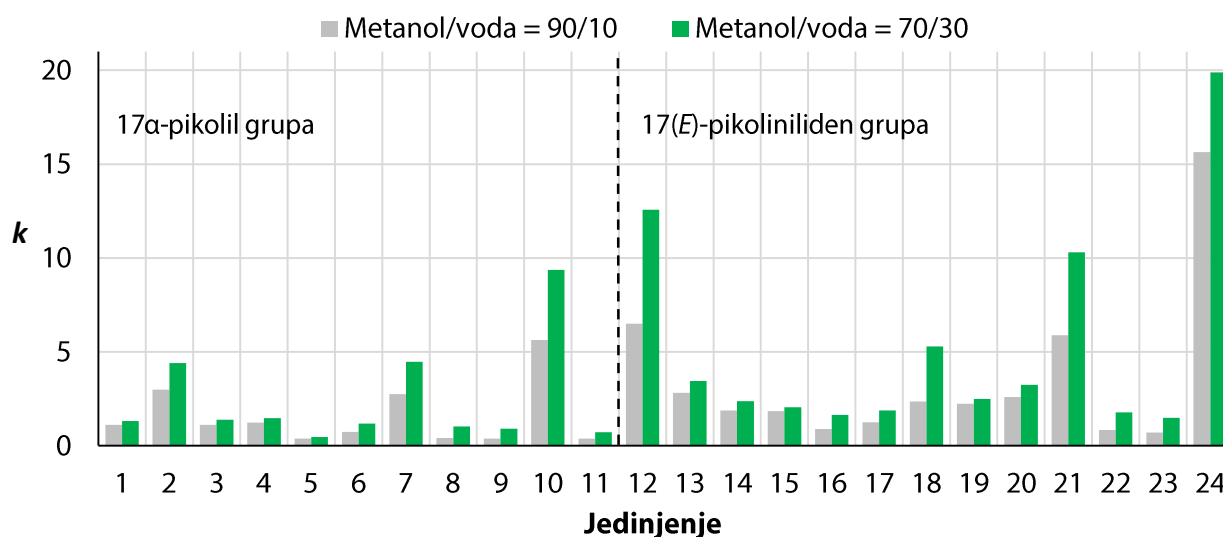


4. Rezultati i diskusija

4.1. Hromatografska analiza androstanskih derivata

Analiza retencionog ponašanja jedinjenja u hromatografiji na obrnutim fazama zasnovana je na pretpostavci da nepolarna jedinjenja ispoljavaju visok afinitet prema nepolarnoj stacionarnoj fazi i usled toga imaju i veću retenciju u odnosu na polarna jedinjenja. Modifikacije u strukturi jedinjenja koje vode povećanju njegove polarnosti smanjuju retenciju, dovodeći do pojave intenzivnijih interakcija sa komponentama mobilne faze, koja je u RP sistemu po pravilu polarnija od stacionarne faze. Modifikacija polarnosti mobilne faze postiže se menjanjem udela organskog rastvarača (modifikatora). U ovoj analizi, polarnija komponenta mobilne faze je voda, dok je kao modifikator upotrebljen metanol.

Rezultati određivanja retencionog ponašanja (hromatografske lipofilnosti) ispitivane serije androstanskih derivata u hromatografiji visokih performansi na obrnutim fazama (C-18 modifikovani silika gel) primenom dve mobilne faze sa različitim udelom metanola (φ) prikazani su na slici 4.1. Vrednosti k i log k parametara, uključujući njihove standardne devijacije, date su u prilogu (tabela P-7.1.).



Slika 4.1. Kapacitetni faktori (k) ispitivanih androstanskih derivata u hromatografiji na obrnutim fazama (RP-HPLC) dobijeni primenom dve mobilne faze sa različitim udelom metanola



Posmatranjem grafičkog prikaza kapacitetnih faktora ispitivanih jedinjenja (*slika 4.1.*) može se uočiti da je retencija obe grupe jedinjenja veća u sistemu sa mobilnom fazom koja sadrži manji udeo metanola. Takođe, može se primetiti da iz grupe 17 α -pikolil derivata, jedinjenje **10** ima najveću retenciju u primenjenom hromatografskom sistemu, uzimajući u obzir obe mobilne faze. Iz grupe 17(E)-pikoliniliden derivata, jedinjenje **24** je okarakterisano najvećim retencionim faktorom. Ono što je zajedničko ovim jedinjenjima je odsustvo supstituenata na A i B prstenovima, kao i prisustvo konjugovanih dvostrukih veza u položajima 3 i 5. Jedinjenje **10** ipak ispoljava manji afinitet prema C-18 stacionarnoj fazi u odnosu na jedinjenje **24**, a samim tim ima i manju retenciju, s obzirom na prisustvo hidroksidne grupe u položaju 17 β , koja je karakteristična za sve članove 17 α -pikolil grupe. Hidroksidna grupa, kao polaran supstituent, povećava ukupnu polarnost jedinjenja i time smanjuje njegov afinitet prema stacionarnoj fazi. Na grafiku je takođe uočljivo da značajnu retenciju imaju i jedinjenja **2, 7, 12, 18** i **21**. Zajednička karakteristika ovih jedinjenja je prisustvo acetoksi grupe u položaju 3. Iako jedinjenje **17** takođe poseduje acetoksi grupu u istom položaju, ono ima manju retenciju, zbog uticaja jako polarne nitro grupe u položaju 4. Jedinjenja **5, 11, 9, 8, 23, 22** i **16** odlikuju se relativno malom retencijom, s obzirom na prisustvo većeg broja polarnih funkcionalnih grupa u njihovoj strukturi.

Uticaj uvođenja ili zamene određenih supstituenata na retenciju može se prikazati na primerima jedinjenja čije se strukture razlikuju isključivo u prisustvu datog supstituenta, kao u slučaju parova jedinjenja prikazanih u *tabeli 4.1.* za koje su izračunate Δk vrednosti. Ukoliko se hidroksidna grupa u položaju 3 zameni acetoksi grupom, kao kod parova jedinjenja **1 i 2, 6 i 7, 13 i 12**, evidentno je da se retencija jedinjenja značajno povećava. Uvođenje N-oksid funkcije dovodi do smanjenja retencije (parovi jedinjenja **14 i 16, 12 i 18**), dok je intenzivniji efekat smanjenja retencije postignut uvođenjem nitro grupe u položaj 4 (par jedinjenja **21 i 17**). Prisustvo metoksi grupe u položaju 4 dovodi do blagog povećanja retencije (par jedinjenja **14 i 19**). Pomeranje dvostruke veze iz položaja 5 u položaj 4 u većini slučajeva je dovelo do smanjenja retencije jedinjenja (parovi jedinjenja **1 i 6, 12 i 21, 13 i 20**), izuzev u slučaju jedinjenja **2 i 7**, gde je zabeležen blagi porast retencije u sistemu sa manjim sadržajem metanola u mobilnoj fazi ($\varphi = 0,70$). Upoređivanjem retencije jedinjenja koja se strukturno razlikuju samo po funkciji prisutnoj u položaju 17, može se uočiti da zamena 17 α -pikolil i 17 β -hidroksi funkcionalnih grupa 17(E)-pikoliniliden grupom dovodi do značajnog povećanja retencije (parovi jedinjenja **10 i 24, 1 i 13, 2 i 12, 8 i 22, 4 i 15, 3 i 14**). Razlika u retencionim vremenima odabranih parova jedinjenja iz *tabele 4.1.* može se uočiti poređenjem njihovih hromatograma prikazanih u *prilogu (slika P-7.1. i P-7.2.)*. Vrednost za koju će se promeniti retencija jedinjenja nakon uvođenja određenog supstituenta, odnosno nakon određene strukturne promene, ne zavisi samo od polarnosti datog supstituenta, nego i od mogućih induktivnih i rezonacionih efekata koji se mogu javiti kao posledica uticaja supstituenta na osnovnu strukturu, odnosno na polarnost jedinjenja. Na ovaj način se može objasniti i uticaj položaja nezasićenih veza na retenciju jedinjenja. Takođe, na promenu retencije mogu ispoljiti uticaj i novoformirane vodonične veze sa komponentama mobilne faze, ukoliko supstituent poseduje elektronegativan atom.

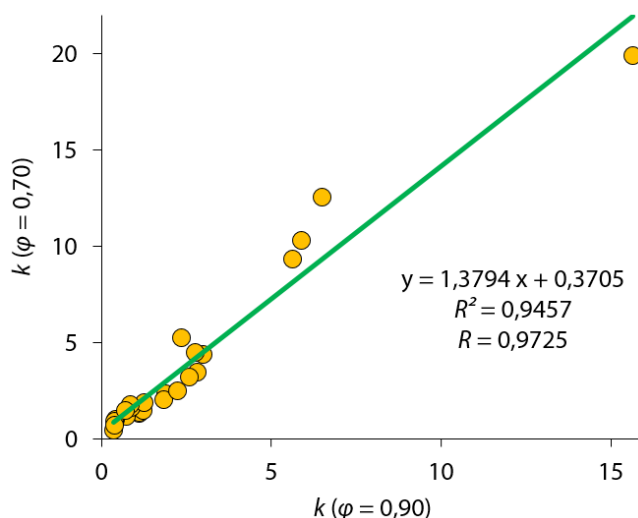
Neophodno je naglasiti da je određivanje retencije androstanskih derivata primenom čiste vode, kao mobilne faze, i C-18 stacionarne faze bilo praktično neizvodljivo za određivanje $\log k_w$ vrednosti kao hromatografskog parametra lipofilnosti. U tom slučaju ovaj parametar ne bi imao fizički smisao za ispitivanu seriju derivata androstana, čak i ukoliko se definiše ekstrapolacijom φ - $\log k$ zavisnosti na $\varphi = 0$. Stoga, hromatografska lipofilnost androstanskih derivata definisana je k i $\log k$ parametrima pri kontrolisanim hromatografskim uslovima, odnosno pri tačno definisanom sastavu mobilne faze.



Tabela 4.1. Δk vrednosti izmerene za uvođenje ili zamenu određenih funkcionalnih grupa, odnosno za promenu položaja dvostruke veze u steroidnom jezgru ispitivanih androstanskih derivata

Jedinjenja koja se porede	Zamena ili uvođenje supstituenta / razlika u položaju dvostruke veze	$\Delta k (\varphi = 0,90)$	$\Delta k (\varphi = 0,70)$	
Poređenje jedinjenja u okviru iste grupe	1 i 2	3-OH \rightarrow 3-AcO	1,880	3,074
	6 i 7	3-OH \rightarrow 3-AcO	2,003	3,301
	1 i 6	5-en \rightarrow 4-en	-0,372	-0,146
	2 i 7	5-en \rightarrow 4-en	-0,249	0,081
	13 i 12	3-OH \rightarrow 3-AcO	3,681	9,127
	14 i 16	N \rightarrow N \rightarrow O	-1,001	-0,728
	12 i 18	N \rightarrow N \rightarrow O	-4,146	-7,296
	21 i 17	4-NO ₂	-4,635	-8,425
	12 i 21	5-en \rightarrow 4-en	-0,612	-2,272
	13 i 20	5-en \rightarrow 4-en	-0,225	-0,213
	14 i 19	4-OCH ₃	0,350	0,127
Poređenje jedinjenja koja pripadaju različitim grupama	10 i 24		10,012	10,524
	1 i 13		1,701	2,129
	2 i 12	17 α -pikolil-17 β -hidroksi	3,502	8,182
	8 i 22	↓	0,435	0,756
	4 i 15	17(E)-pikoliniliden	0,607	0,572
	3 i 14		0,778	0,986

Korelacija između kapacitetnih faktora dobijenih primenom mobilnih faza sa različitim udelom metanola prikazana je na slici 4.2. Visok koeficijent korelacije ukazuje na odlično slaganje retencionih podataka dobijenih pomoću različitih mobilnih faza.



Slika 4.2. Korelacija između k vrednosti definisanih primenom mobilnih faza sa različitim udelom metanola



Da bi se utvrdile potencijalne razlike u retencionom ponašanju 17 α -pikolil grupe u odnosu na ponašanje 17(E)-pikoliniliden grupe androstanskih derivata na osnovu k vrednosti definisanih pomoću dve mobilne faze, primenjen je Vald-Volfovici statistički test. U prvom koraku analize, jedinjenja obe grupe su poređana po rastućoj vrednosti kapacitetnog faktora. Potom su formirani blokovi susednih jedinjenja, članova iste grupe. Broj formiranih blokova (r) je upoređen sa kritičnim brojem blokova (r_{kr}), koji je definisan na osnovu broja članova jedne i druge grupe (n_1 i n_2) pomoću tablice date u literaturi [199]. Ukoliko je $r < r_{kr}$, hipoteza H_0 o razlici retencionog ponašanja ove dve grupe jedinjenja se prihvata. Rezultati Vald-Volfovici testa primenjenog na retencionim podacima ispitivanih jedinjenja prikazani su u tabeli 4.2. Ovi rezultati ukazuju na to da u sistemu sa mobilnom fazom koja sadrži manji udeo metanola retenciono ponašanje 17 α -pikolil derivata androstana se značajno razlikuje od ponašanja 17(E)-pikoliniliden derivata. U tom slučaju kao diskriminacioni faktor ove dve grupe jedinjenja može da se koristi k vrednost.

Tabela 4.2. Rezultati Vald-Volfovici testa

Mobilna faza	n_1	n_2	r	r_{kr}	Hipoteza H_0
$\varphi = 0,90$	11	13	10	7	Odbacuje se ($r > r_{kr}$)
$\varphi = 0,70$	11	13	6	7	Prihvata se ($r < r_{kr}$)

Rezultati hromatografske analize daju značajne informacije o fizičko-hemijskim osobinama ispitivanih androstanskih derivata, najviše o njihovoj polarnosti i lipofilnosti. Polarnost i lipofilnost su faktori sa najvećim uticajem na retenciono ponašanje jedinjenja u hromatografiji na obrnutim fazama. S obzirom na složenost hromatografskih procesa, neophodno je ispitati i potencijalne uticaje drugih faktora na retenciju, kao što su zapremina molekula, molarna masa, molarna refraktivnost i dr. Analiza uticaja ovih faktora na retenciju može se izvesti primenom odgovarajućih korelacionih hemometrijskih tehnika, čijoj primeni često prethodi računarsko modelovanje molekula.



4.2. Računarsko modelovanje strukture androstanskih derivata

Struktura 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata je utvrđena rendgenostrukturnom analizom (RSA) i nuklearnom magnetnom rezonancom (NMR) [37], koje su dale važne podatke o prostornom rasporedu atoma. Pri računarskom modelovanju 3D strukture androstanskih derivata, definisanje rasporeda atoma u molekulu sa što manjom potencijalnom energijom izvedeno je tako da projektovana struktura što više odgovara strukturi definisanoj pomoću eksperimentalnih tehnika. S obzirom na to da su u pitanju mali organski molekuli, optimizacija 3D strukture je obavljena MM2 metodom pomoću programa ChemBio3D i Bioluminate[®]. Optimizovane 3D strukture androstanskih derivata prikazane su u *tabeli 4.3*.

Tabela 4.3. Energetski optimizovane 3D strukture 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata

Jedinjenje	3D struktura (model kuglica-cilindar)	3D struktura (kalotni model)
1		
2		
3		



Tabela 4.3. Energetski optimizovane 3D strukture 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata (nastavak)

Jedinjenje	3D strukutra (model kuglica-cilindar)	3D struktura (kalotni model)
4		
5		
6		
7		
8		



Tabela 4.3. Energetski optimizovane 3D strukture 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata (nastavak)

Jedinjenje	3D strukutra (model kuglica-cilindrar)	3D struktura (kalotni model)
9		
10		
11		
12		
13		



Tabela 4.3. Energetski optimizovane 3D strukture 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata (nastavak)

Jedinjenje	3D strukutra (model kuglica-cilindar)	3D struktura (kalotni model)
14		
15		
16		
17		
18		

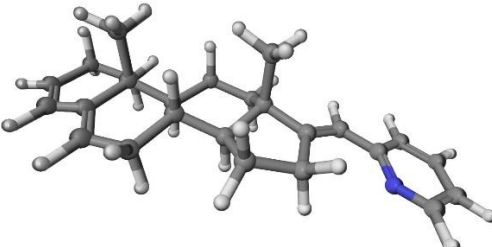
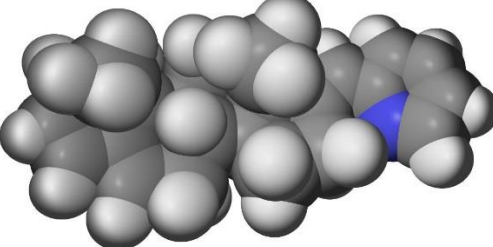


Tabela 4.3. Energetski optimizovane 3D strukture 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata (nastavak)

Jedinjenje	3D strukutra (model kuglica-cilindrar)	3D struktura (kalotni model)
19		
20		
21		
22		
23		



Tabela 4.3. Energetski optimizovane 3D strukture 17 α -pikolil i 17(*E*)-pikoliniliden androstanskih derivata (nastavak)

Jedinjenje	3D strukutra (model kuglica-cilindar)	3D struktura (kalotni model)
24		

Molekulskim modelovanjem dobijen je set od 1233 molekulska deskriptora, uključujući deskriptore lipofilnosti, elektrostatičke, topološke, fizičko-hemijske, kvantno-hemijske i ADMET deskriptore. S obzirom na obim podataka, u prilogu ove disertacije prikazani su najbitniji deskriptori, koji su korišćeni pri QSRR i QSAR modelovanju. Za hemometrijsko modelovanje korišćeni su relevantni deskriptori, isključujući one koji su pokazivali identične vrednosti za veći broj jedinjenja iz serije.

Procena lipofilnosti analiziranih jedinjenja je izvedena na osnovu izračunatih $\log P$ parametara. Pored toga, računarsko modelovanje hidrofилnih i lipofilnih (HILI) površina molekula pruža detaljniji uvid u afinitet određenih delova molekula prema vodenom (hidrofilnom) i/ili lipofilnom okruženju. Modelovanje HILI površina androstanskih derivata obavljeno je pomoću programa Bioluminate® na osnovu van der Valsove energije i energije orijentisanih dipola. Dobijeni 3D modeli HILI površina ispitivanih derivata dati su u tabeli 4.4. Hidrofilne površine su obeležene plavom, a lipofilne narandžastom bojom.

Površine (mape) elektrostatičkog potencijala (EP) omogućavaju uvid u potencijalne asocijacije molekula i u osobine malih molekula [200]. Definisane raspodele elektrostatičkog potencijala u molekulu omogućava procenu polarnosti molekula i potencijalnih međumolekulskih interakcija. Elektronski efekti se mogu kvantifikovati i na osnovu energija HOMO i LUMO molekulskih orbitala. HOMO energija (E_{HOMO}) je pokazatelj distribucije π elektrona u molekulu i objašnjava π - π interakcije prisutne kod jedinjenja sa nezasićenim vezama, dok je LUMO energija (E_{LUMO}) mera interakcija zasnovanih na prenosu naelektrisanja i efekata formiranja vodonične veze. Razlika u LUMO i HOMO energijama ($E_{\text{GAP}} = E_{\text{LUMO}} - E_{\text{HOMO}}$) naziva se energetska praznina (engl. *gap*) i predstavlja indikator stabilnosti, odnosno reaktivnosti molekula (visoka vrednost podrazumeva visoku stabilnost molekula). EP mape zasnovane na Poason-Bolcmanovoj jednačini (Poason-Bolcmanove mape elektrostatičkog potencijala) i HOMO/LUMO orbitale analiziranih androstanskih derivata prikazane su u tabeli 4.4. (zelena boja predstavlja pozitivnu fazu, dok ljubičasta predstavlja negativnu fazu talasne funkcije, odnosno orbitale), a vrednosti dobijenih kvantno-hemijskih deskriptora date su u prilogu u tabeli P-7.2.



Tabela 4.4. EP i HILI molekulske površine i HOMO/LUMO orbitale androstanskih derivata

Jed.	EP površine	HILI površine	HOMO i LUMO orbitale
1	<p>-32,8914 20,5245</p>		<p>HOMO</p> <p>LUMO</p>
2	<p>-28,9100 14,3180</p>		<p>HOMO</p> <p>LUMO</p>
3	<p>-23,9038 9,1332</p>		<p>HOMO</p> <p>LUMO</p>
4	<p>-29,4555 10,9479</p>		<p>HOMO</p> <p>LUMO</p>
5	<p>-27,9069 12,7736</p>		<p>HOMO</p> <p>LUMO</p>
6	<p>-25,7551 17,2959</p>		<p>HOMO</p> <p>LUMO</p>

Tabela 4.4. EP i HILI molekulske površine i HOMO/LUMO orbitale androstanskih derivata (*nastavak*)

Jed.	EP površine	HILI površine	HOMO i LUMO orbitale
7	 -24,7850 7,8921		 HOMO LUMO
8	 -23,1188 20,5366		 HOMO LUMO
9	 -23,5547 13,9145		 HOMO LUMO
10	 -23,9294 21,5052		 HOMO LUMO
11	 -31,2091 29,0187		 HOMO LUMO
12	 -19,9968 12,3674		 HOMO LUMO



Tabela 4.4. EP i HILI molekulske površine i HOMO/LUMO orbitale androstanskih derivata (*nastavak*)

Jed.	EP površine	HILI površine	HOMO i LUMO orbitale
13	<p>-22,3089 23,7128</p>		<p>HOMO</p> <p>LUMO</p>
14	<p>-19,7805 8,7498</p>		<p>HOMO</p> <p>LUMO</p>
15	<p>-19,2984 9,8107</p>		<p>HOMO</p> <p>LUMO</p>
16	<p>-19,7723 12,6601</p>		<p>HOMO</p> <p>LUMO</p>
17	<p>-18,4626 19,6741</p>		<p>HOMO</p> <p>LUMO</p>
18	<p>-17,8046 12,8184</p>		<p>HOMO</p> <p>LUMO</p>



Tabela 4.4. EP i HILI molekulske površine i HOMO/LUMO orbitale androstanskih derivata (*nastavak*)

Jed.	EP površine	HILI površine	HOMO i LUMO orbitale
19	<p>-20,6962 11,8608</p>		<p>HOMO</p> <p>LUMO</p>
20	<p>-27,8664 13,6236</p>		<p>HOMO</p> <p>LUMO</p>
21	<p>-21,1495 9,3695</p>		<p>HOMO</p> <p>LUMO</p>
22	<p>-13,5652 11,5387</p>		<p>HOMO</p> <p>LUMO</p>
23	<p>-17,0841 12,2078</p>		<p>HOMO</p> <p>LUMO</p>
24	<p>-19,5922 9,5346</p>		<p>HOMO</p> <p>LUMO</p>



Posmatranjem modelovanih HILL površina, može se uočiti da je kod većine jedinjenja lipofilni karakter zastupljeniji u odnosu na hidrofilni. Hidrofilne površine su locirane u okolini N atoma piridinskog prstena, kao i u okolini hidroksi, metoksi, acetoksi, nitro, okso i epoksi funkcionalnih grupa vezanih za steroidno jezgro. Kod velikog broja analiziranih derivata hidrofobna površina je znatno veća od hidrofilne, što govori u prilog umerenoj polarnosti ovih jedinjenja. Hidrofilne površine obeležavaju delove molekula androstanskih derivata koji stupaju u dipol-dipol interakcije sa molekulima komponenata mobilne faze (voda i metanol), dok hidrofobne površine markiraju molekulski prostor koji stupa u hidrofobne interakcije (van der Valsove sile) sa stacionarnom C-18 fazom.

Polarni centri ispitivanih molekula mogu se uočiti na EP mapama. Ovi centri su locirani na N atomu piridinskog prstena i na mestima uvedenih hidroksi, metoksi, acetoksi, nitro, okso i epoksi funkcionalnih grupa, na kojima je električni naboj najizraženiji. Negativan električni naboj se može uočiti i na mestima dvostrukih veza (π -elektroni).

Analizom HOMO orbitala, može se zaključiti da su elektroni najviše energije u molekulima analiziranih derivata locirani na 17α -pikolil i $17(E)$ -pikoliniliden grupama, izuzev kod jedinjenja **16**. HOMO orbitale ovog jedinjenja nalaze se na A prstenu steroidnog jezgra. LUMO orbitale su kod većine jedinjenja raspoređene takođe na 17α -pikolil i $17(E)$ -pikoliniliden funkcijama, dok se kod ostalih jedinjenja nalaze u zoni A i B prstenova. Jedinjenje **5** ima najveću energetska prazninu ($E_{\text{GAP}} = 12,9150$ eV) i u tom pogledu se može okarakterisati kao hemijski najstabilnije, za razliku od jedinjenja **8** i **22**, koja imaju najmanju stabilnost sa energetska prazninom od 0,8770 eV i 0,4650 eV, respektivno. HOMO i LUMO orbitale ukazuju na mesta u molekulu na kojima je transfer naelektrisanja najizvesniji. Energije HOMO i LUMO orbitala često se koriste kao molekulski deskriptori u QSRR studijama.

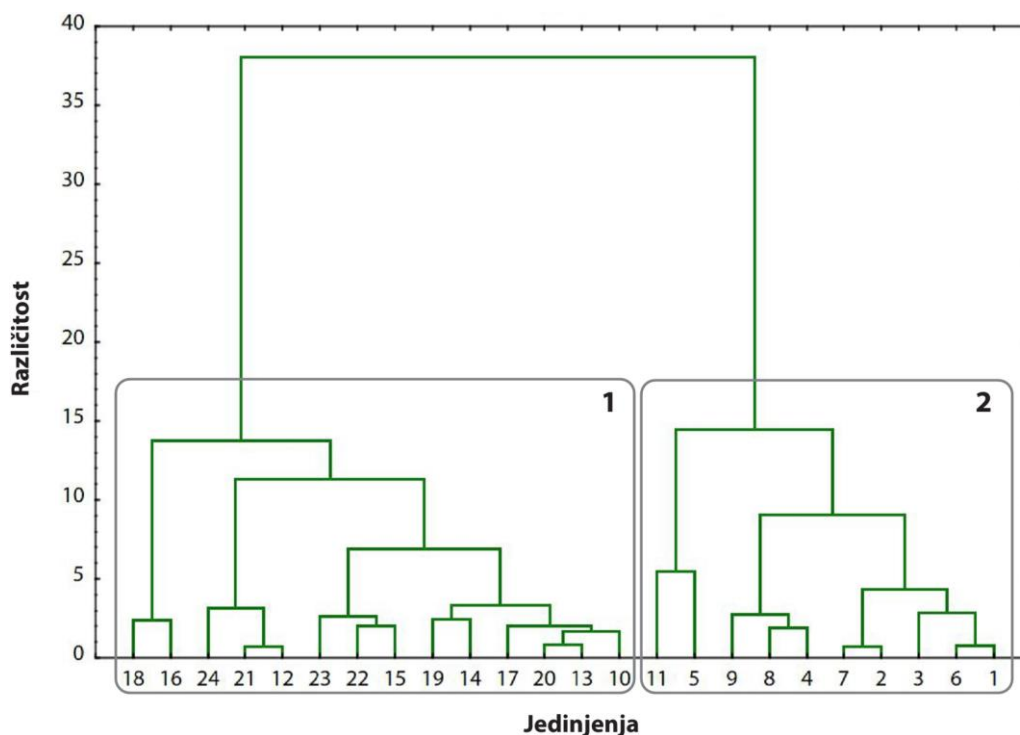
Izračunati kvantno-hemijski deskriptori će poslužiti kao nezavisno promenljive veličine u QSRR modelovanju 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata, kako bi se utvrdio potencijalni uticaj elektronskih efekata na njihovo retenciono ponašanje.



4.3. Analiza androstanskih derivata klasifikacionim hemometrijskim tehnikama

Klaster analiza je izvedena na osnovu tri seta deskriptora primenom Vardove metode pomoću programa Statistica 10, kako bi se utvrdile osobine po kojima su 17 α -pikolil i 17(E)-pikoliniliden androstanski derivati najslićniji, odnosno osobine po kojima se najviše razlikuju.

U prvom koraku analize izvedeno je klasterovanje jedinjenja na osnovu deskriptora lipofilnosti, datih u prilogu u tabeli P-7.3. Pored $\log P$ vrednosti, u analizu su uključeni i distribicioni koeficijent (bez i sa uzimanjem u obzir potencijalne tautomerije i rezonance, $\log D$ i $T\log D$) i rastvorljivost u vodi ($\log S$) kao merila lipofilnosti jedinjenja. S obzirom na to da se prosečne $\log P$ vrednosti androstanskih vrednosti kreću u opsegu 2–7, ova jedinjenja se mogu smatrati lipofilnim. Dendrogram klaster analize zasnovane na deskriptorima lipofilnosti dat je na slici 4.3. Na prikazanom dendrogramu uočavaju se dva jasno definisana klastera, sa podklasterima. U klasteru 1 nalaze se jedinjenja 17(E)-pikoliniliden grupe i jedinjenje 10 17 α -pikolil grupe, dok se u klasteru 2 nalaze jedinjenja 17 α -pikolil grupe. Dobijeni rezultat ukazuje na razliku u lipofilnosti između 17 α -pikolil grupe i 17(E)-pikoliniliden grupe androstanskih derivata, što navodi na pretpostavku da je razlika u lipofilnosti ove dve grupe jedinjenja zapravo posledica razlike u supstituentima u položaju 17.

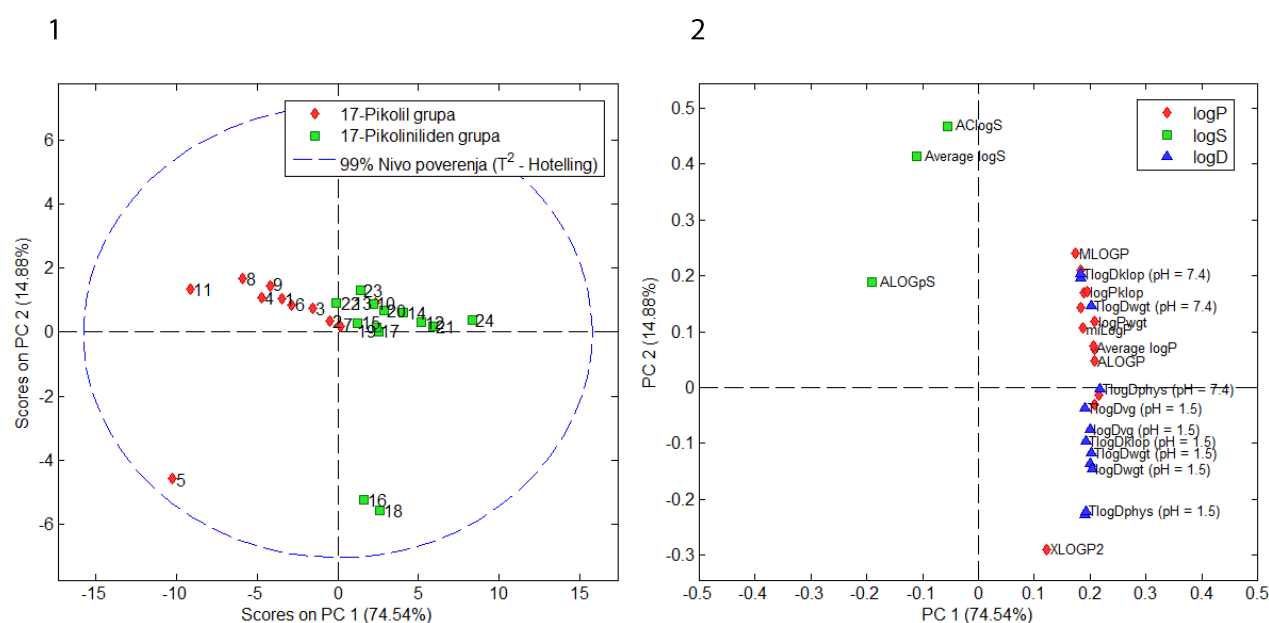


Slika 4.3. Dendrogram kao rezultat klasterovanja jedinjenja na osnovu vrednosti deskriptora lipofilnosti



Na osnovu predstavljenog dendrograma može samo da se konstatuje razdvajanje jedinjenja u dva klastera, međutim ne može da se dobije odgovor na pitanje koja je od te dve grupe jedinjenja okarakterisana manjom ili većom lipofilnošću. Stoga, primenjena je PCA metoda, kako bi se dobio detaljniji uvid u sličnosti i razlike između androstanskih derivata na osnovu lipofilnosti. PCA je izvedena pomoću algoritma dekompozicije singularnih vrednosti (SVD, engl. *Singular Value Decomposition*), uključujući autoskaliranje podataka, korišćenjem MATLAB programa sa PLS_Toolbox dodatkom.

PCA model zasnovan na deskriptorima lipofilnosti obuhvata tri glavne komponente, čija je *eigen* vrednost (EiV) veća od 1 ($EiV_{PC1} = 20,90$; $EiV_{PC2} = 4,17$; $EiV_{PC3} = 1,72$). Ove tri glavne komponente obuhvataju 95,56% varijanse, kojoj PC1 doprinosi sa 74,54%, PC2 sa 14,88% i PC3 sa 6,14%. Grafik skorova i grafik koeficijenata latentnih varijabli PCA predstavljeni su na slici 4.4.



Slika 4.4. Rezultati PCA zasnovani na deskriptorima lipofilnosti ($\log P$, $\log D$, $\log S$): grafik skorova (1) i grafik koeficijenata latentnih varijabli (*loadings plot*) (2) za prve dve glavne komponente (PC1-PC2)

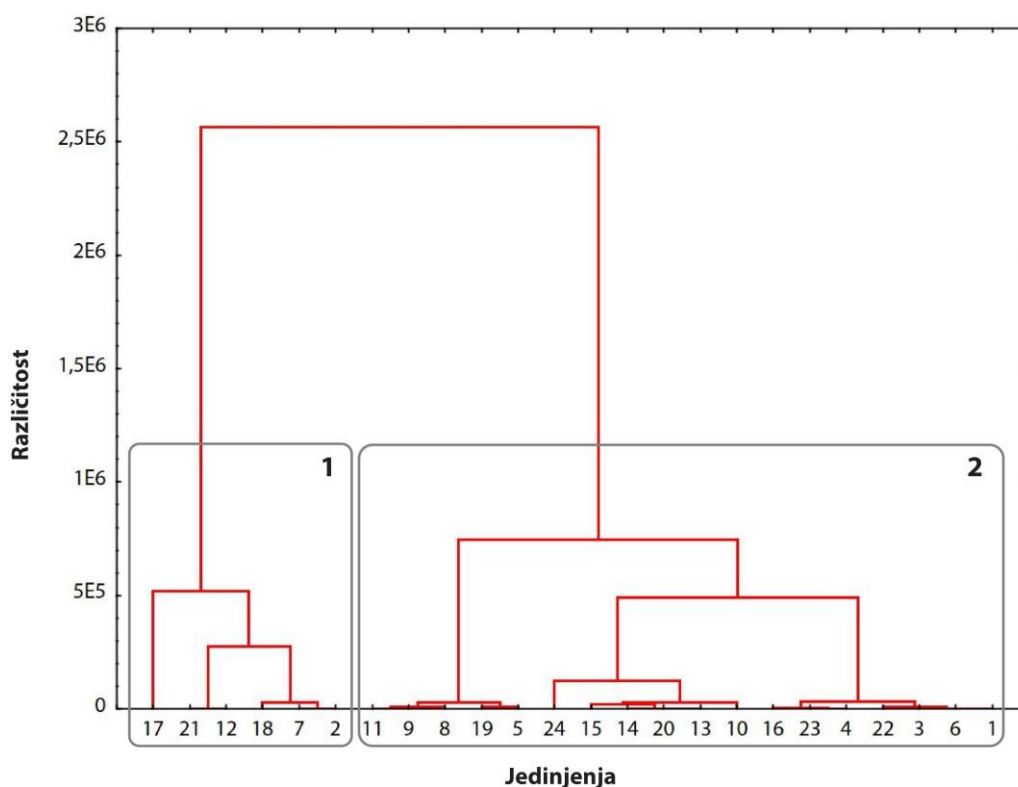
Distribucija jedinjenja na grafiku skorova ukazuje na razdvajanje jedinjenja na dve grupe duž PC1 ose. Grupa 17 α -pikolil derivata pozicionirana je ka negativnim vrednostima skorova PC1, dok se grupa 17(E)-pikoliniliden derivata nalazi u oblasti pozitivnih vrednosti skorova PC1 ose. Izuzetak je ponovo jedinjenje **10**, koje iako strukturno pripada grupi 17 α -pikolil derivata, na grafiku skorova pozicionirano je među 17(E)-pikoliniliden derivatima. Grafik koeficijenata latentnih varijabli ukazuje na to da distribuciji jedinjenja duž PC1 ose doprinosi većina $\log P$ i $\log D$ deskriptora, koji imaju pozitivne koeficijente latentnih varijabli, uključujući i ALogpS deskriptor sa negativnim koeficijentom. Na osnovu prikazanih grafika, može se zaključiti da jedinjenja locirana u oblasti negativnih vrednosti skorova PC1 ose imaju generalno manju lipofilnost i veću rastvorljivost u vodi (17 α -pikolil derivati) u odnosu na jedinjenja koja se nalaze u oblasti pozitivnih vrednosti skorova iste ose (17(E)-pikoliniliden derivati). Ovo razdvajanje jedinjenja je u skladu sa razdvajanjem postignutim klaster analizom. Dakle, PC1 se može smatrati faktorom diskriminacije jedinjenja na osnovu njihove lipofilnosti.



Raspored jedinjenja duž PC2 ose ukazuje na priličnu udaljenost jedinjenja **5**, **16** i **18** u odnosu na ostala. Grafik koeficijenata latentnih varijabli ukazuje na značajan uticaj AClogS, Average logS i XLOGP2 parametara na položaj jedinjenja duž ove ose i navodi na zaključak da ova jedinjenja imaju najmanju rastvorljivost (AClogS) u vodi u odnosu na ostala, ali i relativno visoku vrednost lipofilnosti izraženu preko XLOGP2 parametra. Glavna strukturna karakteristika jedinjenja **5**, **16** i **18** koja ih izdvaja u odnosu na ostale je N-oksidi funkcija na piridinskom prstenu.

Na osnovu T^2 -Hotelling elipse, definisane intervalom poverenja od 99%, može se zaključiti da u skupu podataka nisu detektovani tzv. autlajeri (engl. *outliers*), odnosno jedinjenja koja po posmatranoj osobini (lipofilnosti) drastično odstupaju od ostalih, budući da se sva jedinjenja nalaze unutar T^2 -Hotelling elipse.

U narednom koraku izvedena je klaster analiza androstanskih derivata na osnovu fizičko-hemijskih i topoloških deskriptora. Formirani dendrogram prikazan je na slici 4.5. Na dendrogramu se jasno uočavaju dva glavna klastera, od kojih klaster 2 sadrži 3 veća podklastera. Klasteri 1 i 2 se nalaze na značajnoj udaljenosti. Prisustvo acetoksi grupe u položaju 3 steroidnog jezgra glavna je strukturna karakteristika jedinjenja iz klastera 1, na osnovu koje se razlikuju od jedinjenja koja pripadaju klasteru 2.

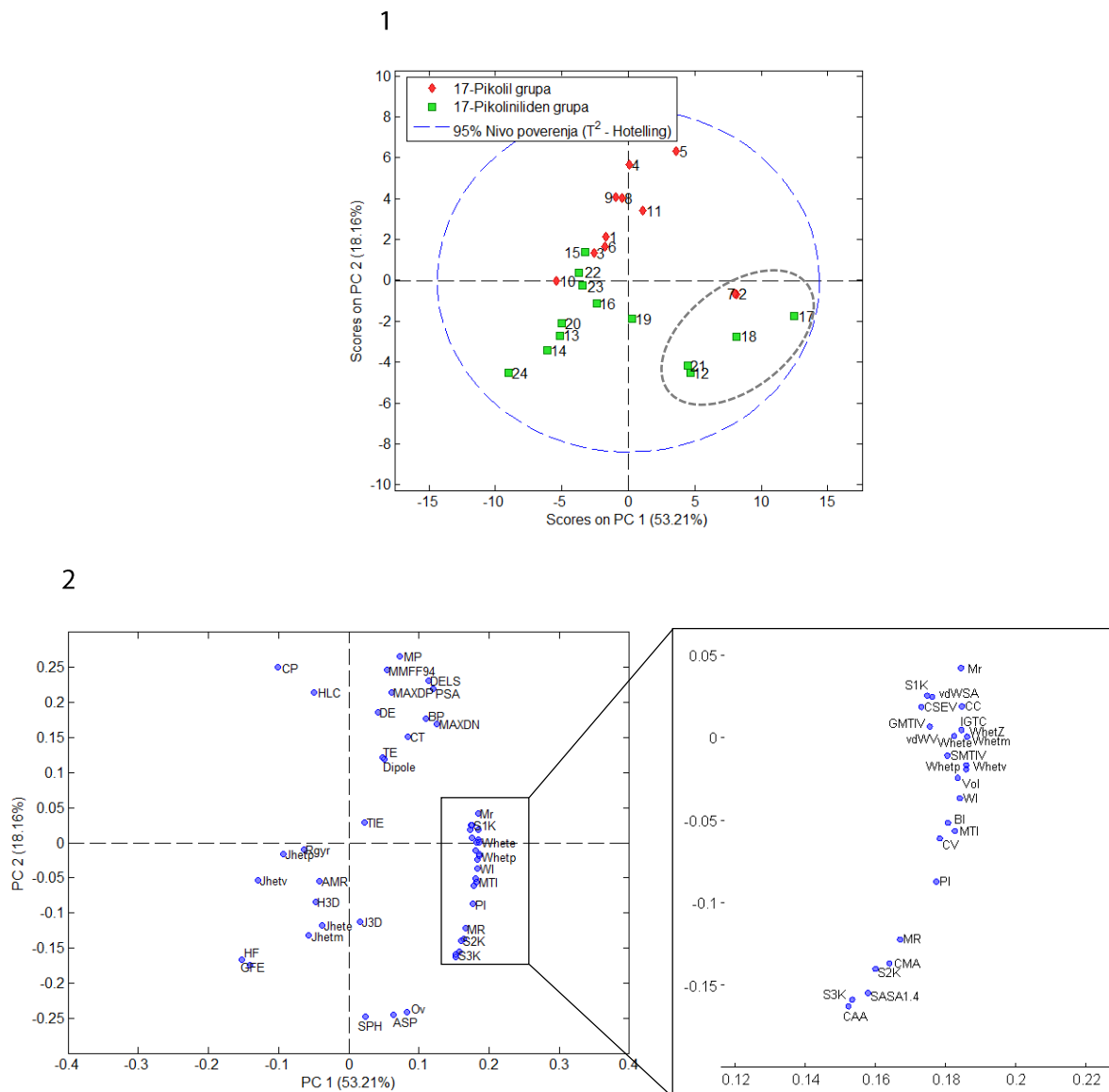


Slika 4.5. Dendrogram kao rezultat klasterovanja jedinjenja na osnovu vrednosti fizičko-hemijskih i topoloških deskriptora

Da bi se stekao detaljniji uvid u grupisanje ispitivanih derivata na osnovu fizičko-hemijskih i topoloških karakteristika, primenjena je PCA metoda. Dobijeni PCA model obuhvata šest glavnih komponenta sa EIV vrednostima većim od 1 ($EiV_{PC1} = 28,70$; $EiV_{PC2} = 9,81$; $EiV_{PC3} = 4,61$, $EiV_{PC4} = 2,93$; EiV_{PC5}



= 2,55; $EiV_{PC6} = 1,33$). Šest izdvojenih glavnih komponenta opisuju 92,53% varijanse, kojoj PC1 doprinosi sa 53,21%, PC2 sa 18,16%, PC3 sa 8,54%, PC4 sa 5,43%, PC5 sa 4,73% i PC6 sa 2,46%. Grafik skorova i grafik koeficijenta latentnih varijabli za PC1 i PC2 predstavljeni su na slici 4.6.



Slika 4.6. Rezultati PCA zasnovani na fizičko-hemijskim i topološkim deskriptorima: grafik skorova (1) i grafik koeficijenta latentnih varijabli (*loadings plot*) (2) za prve dve glavne komponente (PC1-PC2)

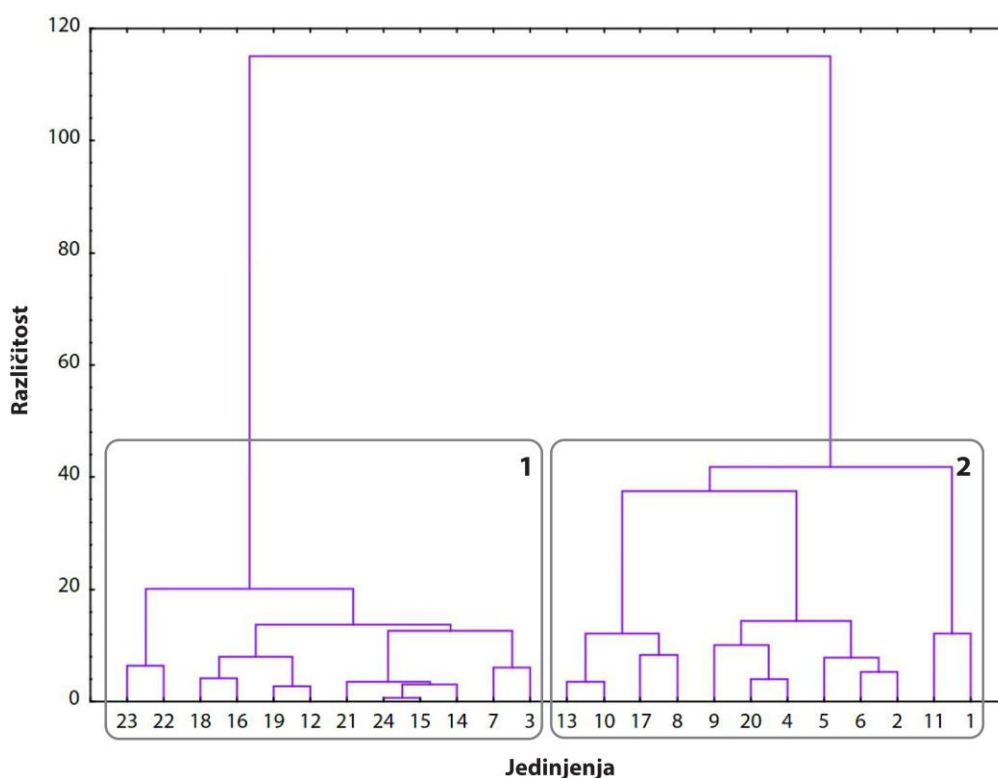
Razdvajanje 17 α -pikolil od 17(E)-pikoliniliden derivata je najizraženije duž PC2 ose grafika skorova. Na položaje jedinjenja duž ove ose dominantan uticaj pokazuje MP (temperatura topljenja) i deskriptori geometrije – SPH (sferičnost), ASP (asferičnost) i Ov (ovalnost). Na raspored jedinjenja duž PC1 ose uticaj ispoljava većina deskriptora, što se može uočiti pomoću grafika koeficijenta latentnih varijabli (uvećanje prikazano na slici 4.6(2)). Tako se, na primer, jedinjenja **2, 7, 12, 17, 18** i **21** razlikuju od ostalih po visokoj polarizabilnosti (PI), molekulskoj zapremini (Vol), relativnoj molekulskoj masi (Mr), van der



Valsovoj površini molekula (vdWSA), kritičnoj zapremini (CV), Vinerovim topološkim indeksima (WI, WhetZ, Whetm, Whetp, Whetv) i drugim deskriptorima navedenim na grafiku koeficijenta latentnih varijabli, ali i po prisustvu acetoksi grupe u položaju 3 steroidnog jezgra, kao što je ranije naglašeno. Ova jedinjenja se na grafiku skorova nalaze na primetnoj distanci u odnosu na ostala, što je u skladu sa rezultatima klaster analize, koja je takođe pokazala njihovo izdvajanje na dendrogramu.

T^2 -Hotelling elipsa na grafiku skorova, definisana za interval poverenja od 95%, obuhvata celokupan skup jedinjenja, ukazujući na odsustvo autlajera.

Klaster analiza androstanskih derivata, zasnovana na kvantno-hemijskim deskriptorima, pokazala je razdvajanje jedinjenja u dve glavne grupe, što se može uočiti na dendrogramu datom na slici 4.7. Jedinjenja 17(E)-pikoliniliden grupe su u najvećoj meri zastupljena u klasteru 1, dok u klasteru 2 većinu čine jedinjenja 17 α -pikolil grupe. U klasteru 1 izuzetke čine jedinjenja **3** i **7**, a u klasteru 2 jedinjenja **13**, **17** i **20**. Ovi rezultati ukazuju na postojanje određene razlike između 17 α -pikolil grupe i 17(E)-pikoliniliden grupe androstanskih derivata sa aspekta njihovih kvantno-hemijskih osobina.

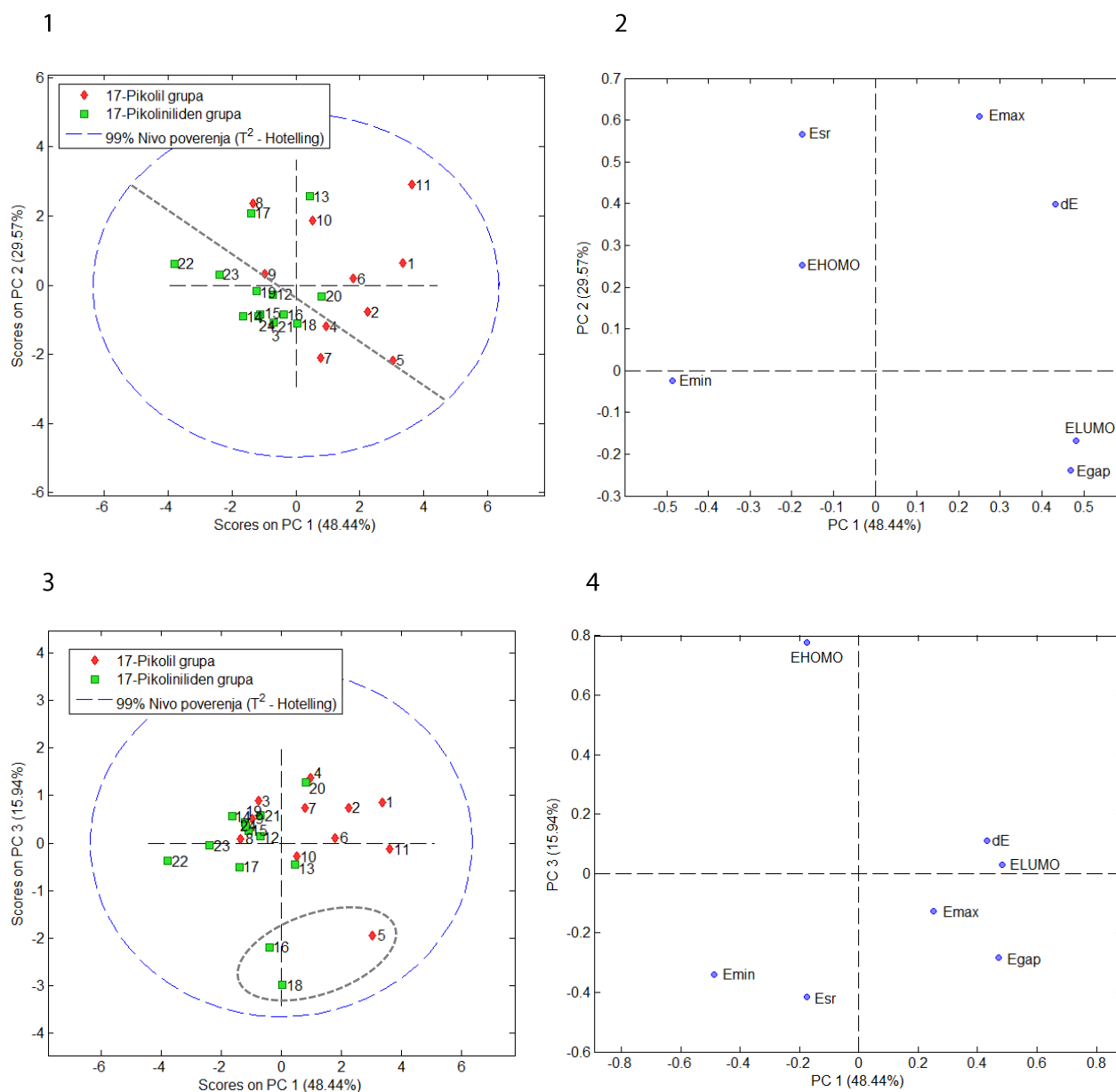


Slika 4.7. Dendrogram kao rezultat klasterovanja jedinjenja na osnovu vrednosti kvantno-hemijskih deskriptora

Nakon klaster analize, izvedena je PCA korišćenjem istog seta kvantno-hemijskih deskriptora. PCA model obuhvata tri glavne komponente koje imaju EiV vrednost veću od 1 ($EiV_{PC1} = 3,39$; $EiV_{PC2} = 2,07$; $EiV_{PC3} = 1,12$). Izdvojene glavne komponente opisuju 93,95% varijanse, kojoj PC1 doprinosi sa 48,44%, PC2



sa 29,57% i PC3 sa 15,94%. Grafik skorova i grafik koeficijenata latentnih varijabli za PC1-PC2 i PC1-PC3 dati su na *slici 4.8*.



Slika 4.8. Rezultati PCA zasnovani na kvantno-hemijskim deskriptorima: grafici skorova (1 i 3) i grafici koeficijenata latentnih varijabli (*loadings plot*) (2 i 4) za PC1-PC2 i PC1-PC3

Poređenjem PC1 i PC2, na grafiku skorova uočava se delimično izdvajanje jedinjenja 17 α -pikolil grupe i to ka smeru pozitivnih vrednosti skorova. Na raspored jedinjenja duž PC1 ose najveći uticaj ispoljavaju E_{LUMO} , ΔE i E_{GAP} deskriptori. Isprekidana prava, koja predstavlja tetivu T^2 -Hotelling elipse, deli skup jedinjenja na dve grupe, identične klasterima prikazanim na dendrogramu na *slici 4.7*. Na istom grafiku može se zapaziti izdvajanje jedinjenja **8**, **10**, **11**, **13** i **17** duž PC2 ose, na osnovu visoke maksimalne (E_{max}) i srednje vrednosti (E_{sr}) elektrostatičkog potencijala.



Grafik skorova formiran na osnovu PC1 i PC3, ukazuje na izdvajanje jedinjenja **5**, **16** i **18** duž PC3 ose u odnosu na ostala jedinjenja. Grafik koeficijenta latentnih varijabli pokazuje da se ova jedinjenja izdvajaju po najnižoj energiji HOMO orbitala (E_{HOMO}). Još jedna karakteristika koja izdvaja jedinjenja **5**, **16** i **18** u odnosu na ostala je prisustvo N-oksid funkcije na piridinskom prstenu.

Celokupan skup jedinjenja se nalazi unutar T^2 -Hotelling elipse definisane za interval poverenja od 99%, što ukazuje na odsustvo autlajera.

Na osnovu predstavljenih rezultata PCA, može se zaključiti da se analizirana 17 α -pikolil i 17(*E*)-pikoliniliden jedinjenja najviše razlikuju po lipofilnosti, dok su razlike između ove dve grupe jedinjenja po fizičko-hemijskim, topološkim i kvantno-hemijskim deskriptorima manje izražene. Neophodno je naglasiti da između 17 α -pikolil i 17(*E*)-pikoliniliden jedinjenja nije utvrđena drastična udaljenost u prostoru razmatranih varijabli. Ovo je važan podatak za QSRR i QSAR modelovanje androstanskih derivata, s obzirom na to da modeli koji se definišu u tom slučaju mogu biti zasnovani na obe grupe jedinjenja. U slučaju jako izraženih razlika između grupa jedinjenja, bilo bi neophodno definisati matematičke modele za svaku grupu posebno.

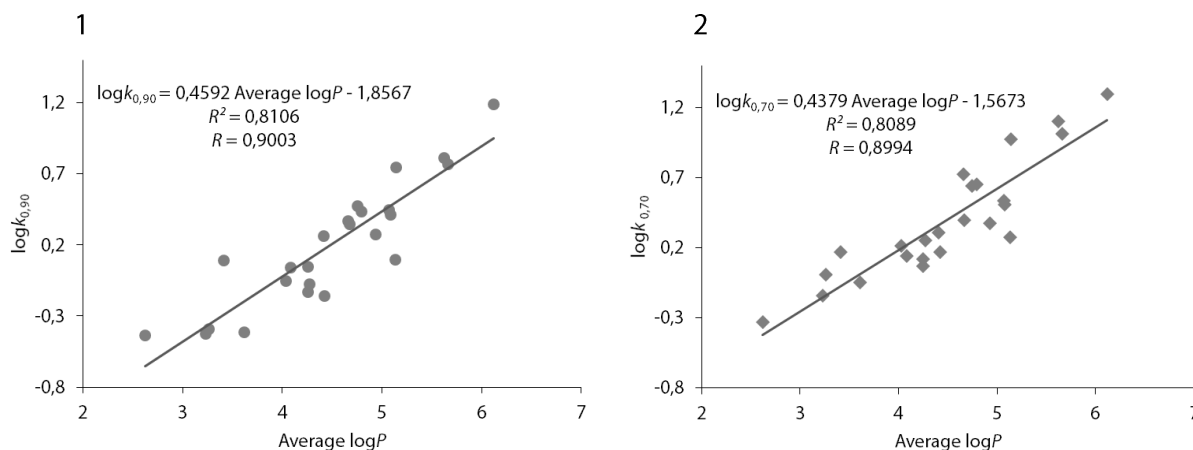


4.4. QSRR analiza retencionog ponašanja androstanskih derivata

4.4.1. Linearno QSRR modelovanje

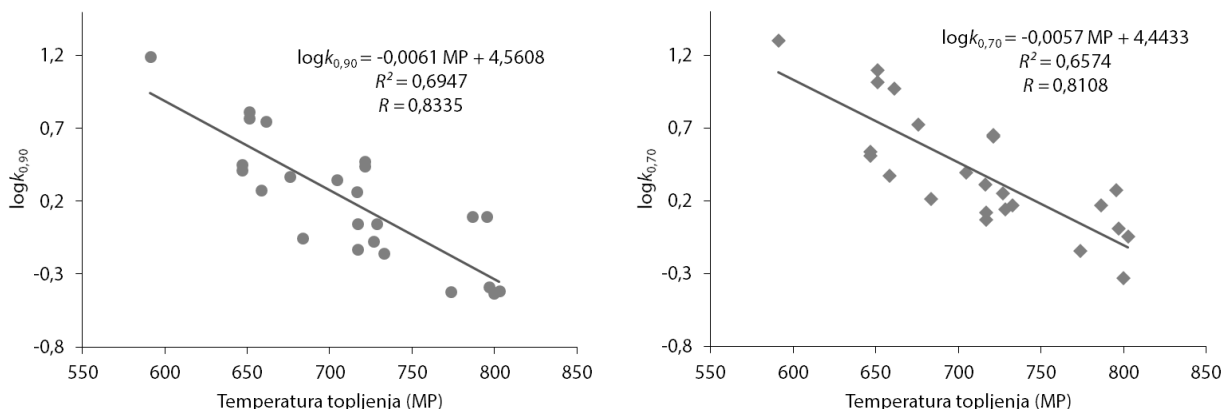
Cilj QSRR analize 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata je da se ustanove linearne i/ili nelinearne matematičke veze između odgovarajućih *in silico* molekulskih deskriptora i eksperimentalno određenih retencionih parametara. Ukoliko se QSRR modelovanjem dobiju statistički validni modeli, njihova uloga u predviđanju retencionog ponašanja 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata u RP-HPLC sistemu, a samim tim i hromatografske lipofilnosti ispitivanih jedinjenja mogla bi biti od velikog značaja. QSRR analiza polazi od primene najjednostavnijih regresionih tehnika, kao što su linearna i polinomska regresija, uključujući po jednu nezavisno promenljivu veličinu. Složenije zavisnosti između retencionih parametara i molekulskih deskriptora utvrđuju se primenom multivarijacionih regresionih tehnika, kao što su MLR, PCR, PLS i ANN.

Prvi korak QSRR modelovanja obuhvata ispitivanje zavisnosti između deskriptora lipofilnosti i $\log k$ vrednosti. Na osnovu korelacione matrice, utvrđena je veoma jaka korelacija između lipofilnosti (Average $\log P$) i retencionih koeficijenata definisanih pomoću mobilne faze sa većim ($\log k_{0,90}$) i manjim udelom metanola ($\log k_{0,70}$) (slika 4.9.), što predstavlja potvrdu hipoteze da se $\log k$ parametar, definisan pri kontrolisanim hromatografskim uslovima, može smatrati alternativnim parametrom lipofilnosti 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata (hromatografska lipofilnost).

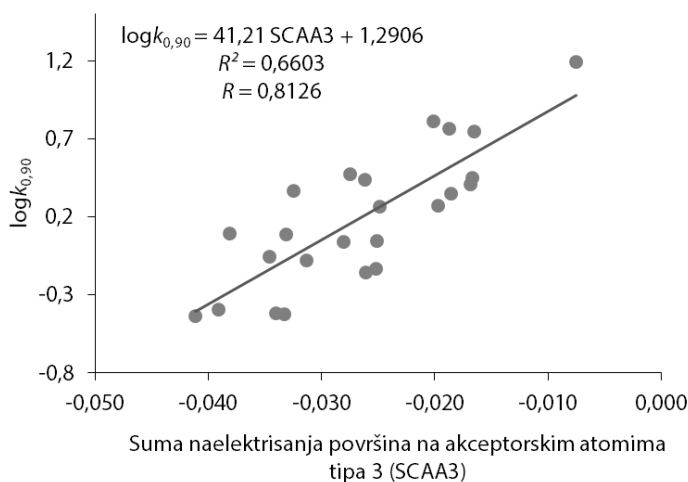


Slika 4.9. Linearna korelacija između *in silico* deskriptora lipofilnosti (Average $\log P$) i retencionog koeficijenta $\log k$ definisanog primenom mobilne faze sa većim (1) i manjim udelom metanola (2)

Jaka linearna veza ustanovljena je između *in silico* temperature topljenja (MP) i $\log k$ definisanog pri većem udelu metanola, i deskriptora ponderisane naelektrisane površine na akceptorima vodonične veze (SCAA3) i $\log k$ definisanog pri većem i pri manjem udelu metanola u mobilnoj fazi. Grafici ovih linearnih modela prikazani su na slici 4.10 i slici 4.11.



Slika 4.10. Linearna korelacija između *in silico* temperature topljenja (MP) i retencionih koeficijenata $\log k_{0,90}$ (1) i $\log k_{0,70}$ (2)



Slika 4.11. Linearna korelacija između *in silico* deskriptora ponderisane naelektrisane površine na akceptorima vodonične veze (SCAA3) i retencionog koeficijenta $\log k_{0,90}$

Predstavljeni modeli služe samo kao pokazatelji određenih zavisnosti i nisu namenjeni predviđanju logk parametra. Dakle, pokazan je značajan uticaj deskriptora lipofilnosti i deskriptora koji opisuje uticaj određenih elektrostatičkih osobina na retenciju. Da bi se dobili statistički validni QSRR modeli, koji se mogu pouzadno koristiti u predviđanju retencionog ponašanja, neophodno je sprovesti internu i eksternu validaciju. Eksterna validacija zahteva formiranje eksternog test seta, stoga celokupni set od 24 jedinjenja je podeljen na dva seta:



1. kalibracioni set od 19 jedinjenja (\approx 80% jedinjenja od ukupnog broja) – **1, 3, 4, 5, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24;**
2. eksterni test set od 5 jedinjenja (\approx 20% jedinjenja od ukupnog broja) – **2, 6, 12, 19, 23.**

Najbolji linearni modeli (LR) zasnovani su na deskriptorima lipofilnosti:

$$\text{LR1: } \log k_{0,90} = 0,35797 (\pm 0,04005) \mathbf{ALOGPs} - 1,33462 (\pm 0,17507) \quad (4.1.)$$

$$\text{LR2: } \log k_{0,90} = 0,44518 (\pm 0,04913) \mathbf{Average logP} - 1,78180 (\pm 0,22085) \quad (4.2.)$$

(brojevi u zagradama predstavljaju standardnu grešku regresionih koeficijenata)

Analizom korelacije dobijen je i polinomski model (PR) drugog stepena, sledećeg oblika:

$$\text{PR: } \log k_{0,90} = 0,000018 (\pm 0,000015) \mathbf{MP}^2 - 0,03202 (\pm 0,02075) \mathbf{MP} + 13,62574 (\pm 7,36103) \quad (4.3.)$$

Navedeni QSRR modeli su dobijeni pomoću kalibracionog seta, a potom je obavljena njihova interna i eksterna validacija. Osnovni i validacioni statistički parametri dobijenih modela prikazani su u tabeli 4.5.

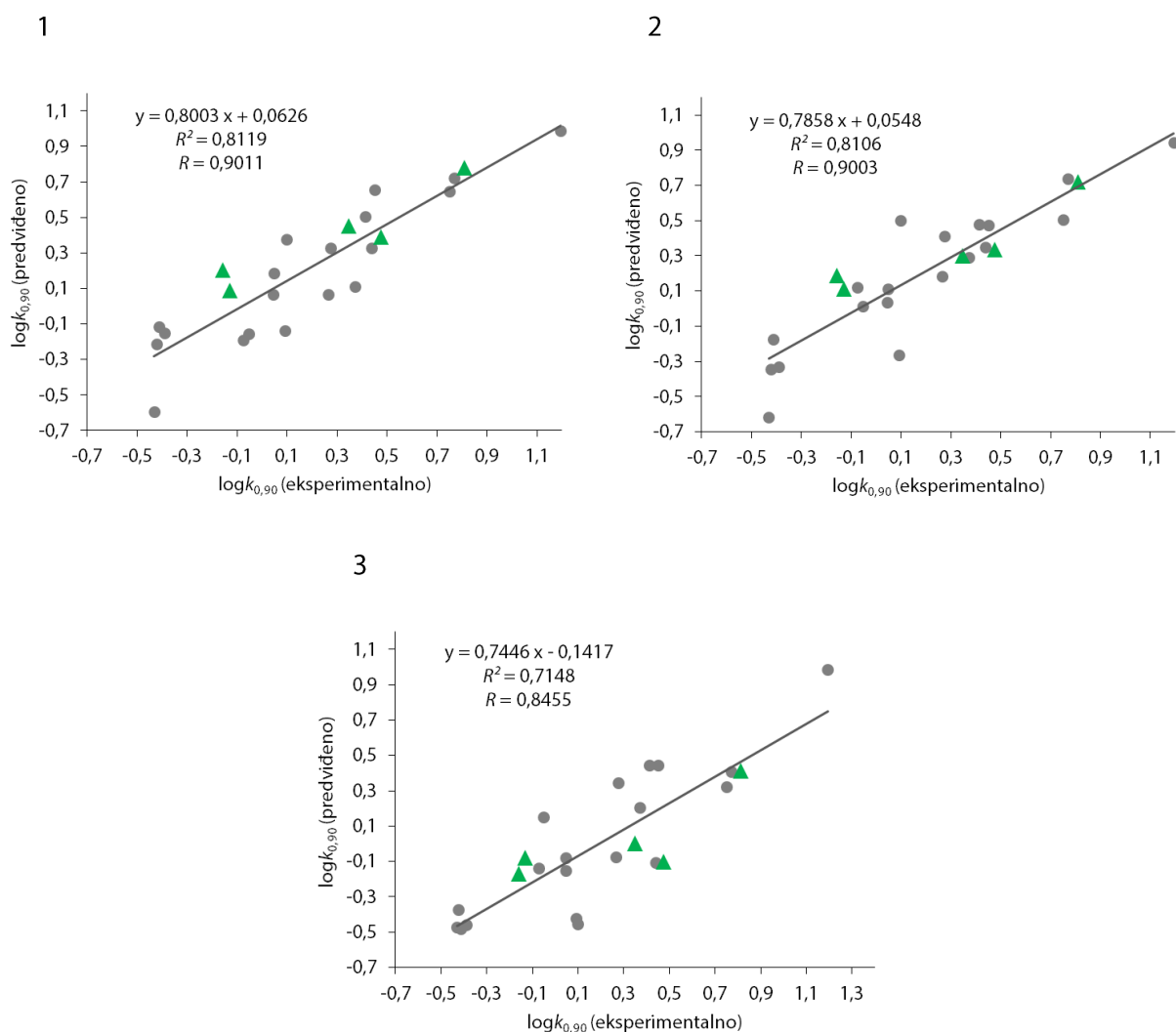
Tabela 4.5. Statistički parametri interne i eksterne validacije LR i PR QSRR modela za predviđanje $\log k_{0,90}$

Parametri	LR1	LR2	PR
R	0,9081	0,9102	0,8658
R^2	0,8246	0,8285	0,7496
R^2_a	0,1754	0,1715	0,2504
R^2_{adj}	0,8142	0,8184	0,7183
F -vrednost	79,90	82,10	23,94
$RMSE$	0,1909	0,1887	0,2351
p -vrednost	0,000000	0,000000	0,000015
R^2_{cv} (LOO)	0,7781	0,7746	0,6476
$ R^2 - R^2_{cv} $	0,0465	0,0539	0,1020
TSS	3,5303	3,5303	3,5303
$PRESS$	0,7832	0,7958	1,2440
$PRESS/TSS$	0,2219	0,2254	0,3524
SD_{PRESS}	0,2030	0,2047	0,2559
R_{test}	0,9489	0,9199	0,8147
$RMSE_{test}$	0,0963	0,1067	0,1601

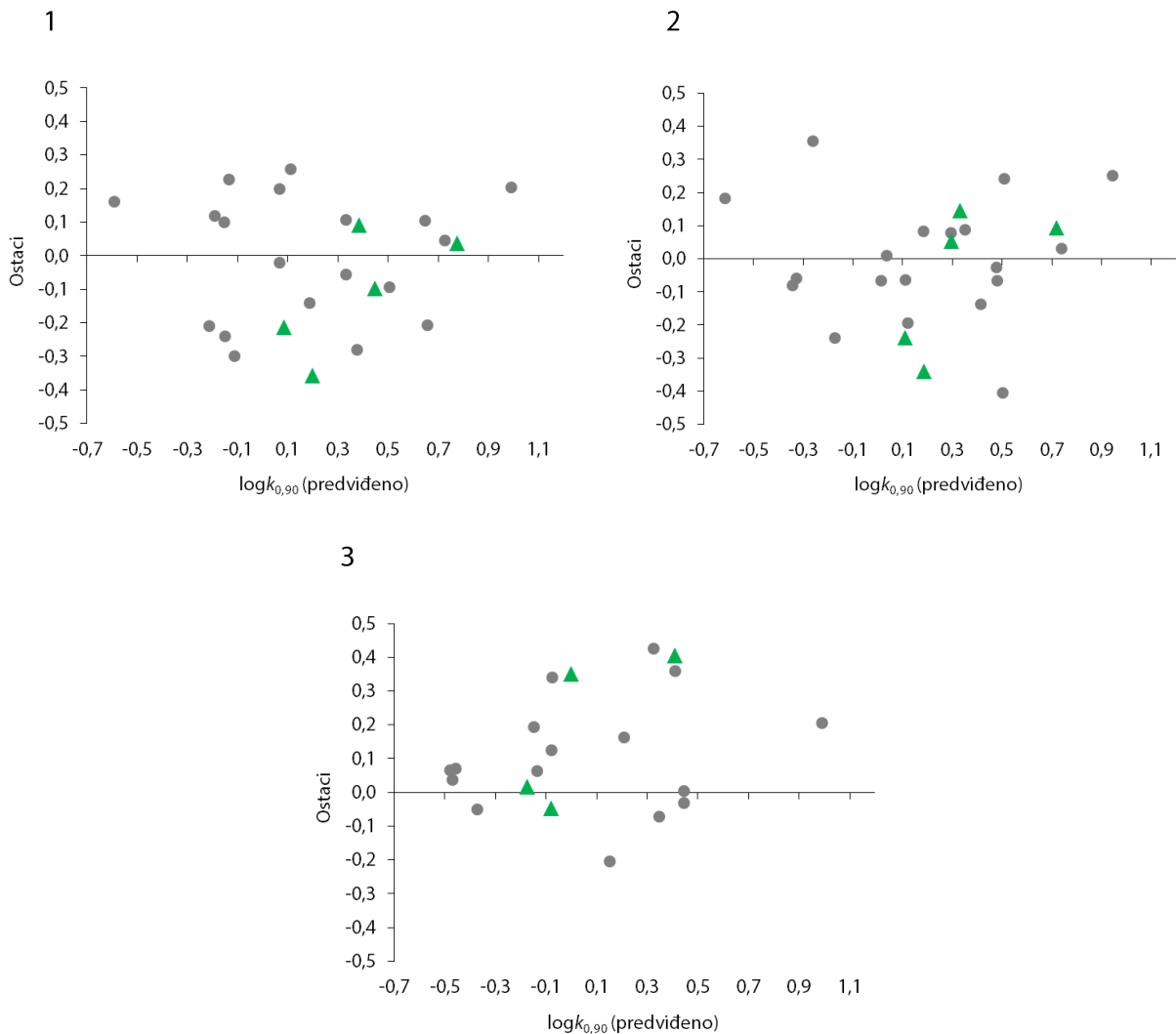


Rezultati statističke analize ukazuju na validnost dobijenih LR i PR modela, ističući poseban kvalitet linearnih jednačina u odnosu na polinomski model. Koeficijenti determinacije značajno veći od 0,60 ukazuju na veoma jaku korelaciju između varijabli, dok parametri unakrsne i eksterne validacije potvrđuju prediktivnu sposobnost formiranih modela. U prilog kvalitetu modela govore i relativno niske *RMSE* i *PRESS* vrednosti. Nedostatak PR modela odnosi se na apsolutnu razliku R^2 i R^2_{CV} vrednosti, koja u ovom slučaju neznatno prelazi granicu prihvatljivosti (0,10), kao i na nešto veću grešku predviđanja (visoke vrednosti *PRESS* i SD_{PRESS} parametara).

Prediktivna moć modela testirana je i poređenjem eksperimentalnih i predviđenih $\log k_{0,90}$ vrednosti (slika 4.12.), uključujući i analizu njihovih razlika, tzv. ostataka (slika 4.13.).



Slika 4.12. Poređenje eksperimentalnih retencionih parametara, dobijenih u sistemu sa zapreminskim udelom metanola od 0,90 ($\log k_{0,90}$), sa $\log k_{0,90}$ vrednostima predviđenim primenom LR1 (1), LR2 (2) i PR (3) QSRR modela (● - kalibracioni set, ▲ - eksterni test set)



Slika 4.13. Poređenje $\log k_{0,90}$ vrednosti predviđenih primenom LR1 (1), LR2 (2) i PR (3) QSRR modela sa vrednostima ostataka (● - kalibracioni set, ▲ - eksterni test set)

Poređenje eksperimentalnih i predviđenih vrednosti ukazuje na dobro slaganje podataka kod oba LR modela, s obzirom na visoke vrednosti koeficijenta korelacije, relativno malo rasipanje tačaka oko prave, vrednost nagiba prave blisku jedinici i odsečak blizak nuli. Sa ovog aspekta, polinomski model je okarakterisan nešto lošijim vrednostima pomenutih parametara.

Vrednosti ostataka se nalaze u prihvatljivim granicama kod sva tri modela. Nasumičnost njihove raspodele oko $y = 0$ ose je očigledna kod linearnih modela. Kod polinomskog modela su zastupljene znatno veće vrednosti ostataka sa pozitivnim predznakom u odnosu na negativne ostatke. Ova pojava može ukazati na potrebu za polinomom većeg stepena, međutim u ovom slučaju povećanje složenosti polinoma dovelo bi do drastičnog pogoršanja prediktivne moći. Stoga, polinomski model se može smatrati manje preciznim u predviđanju $\log k$ vrednosti u odnosu na linearne. Nasumičnost raspodele ostataka oko $y = 0$ ose potvrđena je Vald-Volfovici testom ($n_1 = 11, n_2 = 5, r_{kr} = 5, r = 11; r > r_{kr}$ hipoteza H_0 o nasumičnoj raspodeli ostataka je prihvaćena).



Na osnovu prezentovanih rezultata statističke validacije formiranih LR i PR modela, njihovo rangiranje prema kvalitetu predviđanja $\log k_{0,90}$ vrednosti može se predstaviti na sledeći način:

$$LR1 \approx LR2 > PR$$

Linearni modeli se međusobno neznatno razlikuju u pogledu statističkih parametara i imaju sličnu moć predviđanja.

QSRR modelovanjem $\log k$ parametra, definisanog primenom mobilne faze sa manjim zapreminskim udelom metanola ($\varphi = 0,70$), dobijena su dva linearna modela zadovoljavajućih statističkih karakteristika:

$$LR3: \log k_{0,70} = 0,41726 (\pm 0,04797) \text{ **Average logP** } - 1,46768 (\pm 0,21561) \quad (4.4.)$$

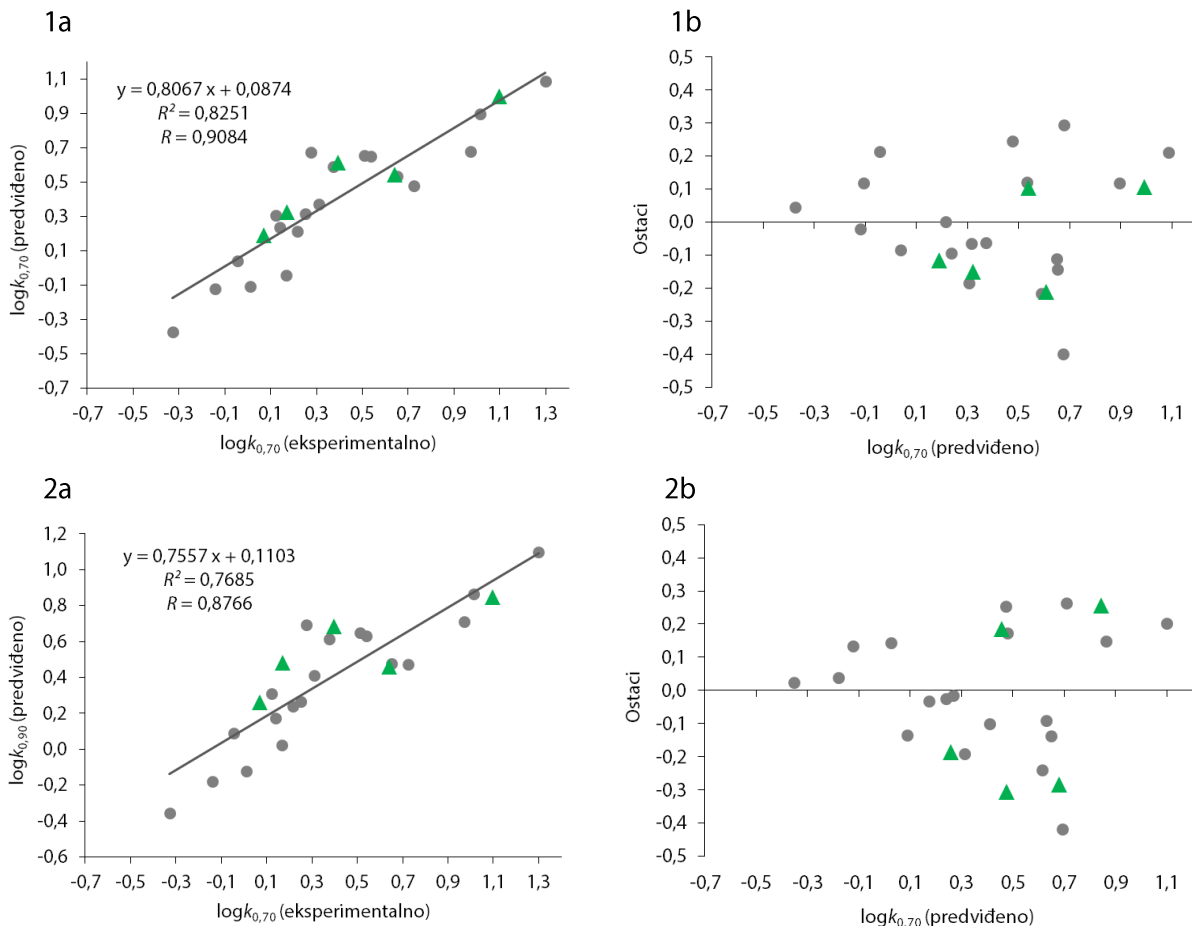
$$LR4: \log k_{0,70} = 0,37777 (\pm 0,04409) \text{ **XLOGP3** } - 1,30892 (\pm 0,20075) \quad (4.5.)$$

U tabeli 4.6. dati su osnovni i validacioni statistički parametri navedenih matematičkih modela.

Tabela 4.6. Statistički parametri interne i eksterne validacije linearnih QSRR modela za predviđanje $\log k_{0,70}$ parametra

Parametri	LR3	LR4
R	0,9037	0,9011
R^2	0,8166	0,8120
R^2_a	0,1834	0,1880
R^2_{adj}	0,8058	0,8009
F -vrednost	75,68	73,41
$RMSE$	0,1843	0,1866
p -vrednost	0,000000	0,000000
R^2_{cv} (LOO)	0,7714	0,7664
$ R^2 - R^2_{cv} $	0,0452	0,0345
TSS	3,1465	3,1465
$PRESS$	0,7191	0,7351
$PRESS/TSS$	0,2285	0,2336
SD_{PRESS}	0,1945	0,1967
R_{test}	0,9546	0,8036
$RMSE_{test}$	0,1062	0,1545

Kvalitet modela testiran je poređenjem eksperimentalnih i predviđenih $\log k_{0,70}$ vrednosti, kao i analizom distribucije ostataka (slika 4.14).



Slika 4.14. Poređenje eksperimentalnih retencionih parametara, dobijenih u sistemu sa zapreminskim udelom metanola od 0,70 ($\log k_{0,70}$), sa $\log k_{0,70}$ vrednostima predviđenim primenom LR3 (1a) i LR4 (2a) QSRR modela, uključujući i prikaz distribucije ostataka za LR3 (1b) i LR4 (2b) model (● - kalibracioni set, ▲ - eksterni test set)

Modeli LR3 i LR4 okarakterisani su zadovoljavajućim vrednostima svih statističkih parametara koji govore o njihovoj prediktivnoj sposobnosti, koja je potvrđena i primenom eksternog test seta. Prednost LR3 modela nad LR4 modelom ogleda se u većoj vrednosti nagiba zavisnosti eksperimentalno–predviđeno i odsečku koji je bliži nuli. Nasumičnost raspodele ostataka oko $y = 0$ ose je očigledna kod oba modela.

Nedostaci i LR i PR modela ogledaju se u relativno visokim vrednostima ostataka, kako za eksterni tako i kalibracioni set jedinjenja. Ostaci se za LR1 model nalaze u opsegu od -0,357 do 0,260, za LR2 od -0,405 do 0,357, za LR3 od -0,398 do 0,295, za LR4 od -0,419 do 0,263 i za PR od -0,202 do 0,580. Prosečne vrednosti pozitivnih i negativnih ostataka za LR i PR modele, prikazane u tabeli 4.7, potvrđuju prednost LR modela nad PR modelom sa aspekta srednje razlike između eksperimentalnih i predviđenih vrednosti. Na osnovu maksimalnog opsega vrednosti ostataka može se steći uvid u maksimalnu i minimalnu razliku između eksperimentalne i predviđene vrednosti koja se dobija primenom određenog modela, dok prosečne vrednosti ostataka pokazuju koja je srednja razlika između eksperimentalnih i predviđenih vrednosti. Na ovaj način se dobija realniji uvid u grešku predviđanja određenog modela sa aspekta analize ostataka.

**Tabela 4.7.** Prosečne (srednje) vrednosti pozitivnih i negativnih ostataka za LR i PR modele

Modeli	Prosečna vrednost pozitivnih ostataka	Prosečna vrednost negativnih ostataka
LR1	0,138	-0,184
LR2	0,135	-0,159
LR3	0,144	-0,143
LR4	0,165	-0,166
PR	0,264	-0,080

S obzirom na to da predstavljeni LR i PR modeli opisuju uticaj samo jednog faktora na retenciju ispitivanih jedinjenja, uzimajući u obzir složenost hromatografskih procesa bilo je neophodno primeniti multivarijantne regresione metode, koje ukazuju na zajednički uticaj više faktora (molekulskih deskriptora) na $\log k$ parametar.

U prvom koraku multivarijantne regresione analize primenjena je multipla linearna regresija pomoću NCSS 2007 programa. Odabir deskriptora koji figurišu u MLR modelima postignut je pomoću stepenaste selekcije. Maksimalan broj nezavisno promenljivih u MLR modelima je tri, što je u skladu sa Toplis-Kostelo pravilom. Prilikom odabira molekulskih deskriptora, koji se nalaze u MLR modelima, javio se problem jake multikolinearnosti, te su molekularni deskriptori koji doprinose ovom problemu izuzeti iz modela, a potom zamenjeni adekvatnim deskriptrom. Zamena pomenutih deskriptora je obavljena metodom probe i greške, tako da su zadržani oni deskriptori koji doprinose povećanju koeficijenta korelacije, odnosno smanjenju ukupne greške predviđanja bez povećanja multikolinearnosti (u ovom radu tolerisane su vrednosti *VIF* faktora ispod 10, za sve nezavisno promenljive). Ovim pristupom dobijeni su sledeći MLR modeli namenjeni predviđanju $\log k_{0,90}$ i $\log k_{0,70}$ parametara:

$$\text{MLR1:} \quad (4.6.)$$

$$\log k_{0,90} = 0,39009 (\pm 0,03960) \text{ Average logP} - 0,00644 (\pm 0,00128) \text{ CT} + 0,00205 (\pm 0,00128) \text{ DE} + 4,28539 (\pm 1,33568)$$

$$\text{MLR2:} \quad (4.7.)$$

$$\log k_{0,90} = 0,31379 (\pm 0,03399) \text{ XLOGP3} - 0,00685 (\pm 0,00135) \text{ CT} - 1,02150 (\pm 0,36183) \text{ Jhetv} + 6,68200 (\pm 1,46962)$$

$$\text{MLR3:} \quad (4.8.)$$

$$\log k_{0,90} = 40,04356 (\pm 3,93576) \text{ SCAA3} - 0,11597 (\pm 0,02074) \text{ CP} + 0,00212 (\pm 0,00077) \text{ DE} + 2,89430 (\pm 0,31819)$$

$$\text{MLR4:} \quad (4.9.)$$

$$\log k_{0,90} = 0,36138 (\pm 0,05960) \text{ Average logP} + 0,00277 (\pm 0,00076) \text{ DE} - 0,00276 (\pm 0,00083) \text{ MP} + 0,19114 (\pm 0,81162)$$



MLR5: (4.10.)

$$\log k_{0,90} = 0,45822 (\pm 0,04500) \text{ Average logP} - 0,02847 (\pm 0,01322) \text{ E}_{\text{sr}} - 1,96110 (\pm 0,21706)$$

MLR6: (4.11.)

$$\log k_{0,90} = 0,36361 (\pm 0,03008) \text{ ALOGPs} - 0,01969 (\pm 0,00554) \text{ E}_{\text{max}} + 0,02356 (\pm 0,01097) \text{ E}_{\text{GAP}} - 1,21035 (\pm 0,17564)$$

MLR7: (4.12.)

$$\log k_{0,90} = 0,39182 (\pm 0,03467) \text{ ALOGPs} - 0,01714 (\pm 0,00556) \text{ E}_{\text{max}} - 0,06415 (\pm 0,02959) \text{ E}_{\text{HOMO}} - 1,78929 (\pm 0,38008)$$

MLR8: (4.13.)

$$\log k_{0,90} = 0,32753 (\pm 0,03418) \text{ ALOGPs} - 17,49395 (\pm 5,68512) \text{ FPSA3} - 0,60873 (\pm 0,27587)$$

MLR9: (4.14.)

$$\log k_{0,90} = 0,35765 (\pm 0,02826) \text{ ALOGPs} - 0,01475 (\pm 0,00363) \text{ \Delta E} + 0,03507 (\pm 0,01086) \text{ E}_{\text{GAP}} - 0,98360 (\pm 0,19327)$$

MLR10: (4.15.)

$$\log k_{0,90} = 0,20208 (\pm 0,04283) \text{ ALOGPs} - 0,01641 (\pm 0,00345) \text{ PSA} + 0,00436 (\pm 0,00120) \text{ vdWSA} - 2,50896 (\pm 0,56880)$$

MLR11: (4.16.)

$$\log k_{0,90} = 0,33171 (\pm 0,04212) \text{ ALOGPs} + 0,00424 (\pm 0,00127) \text{ TE} - 0,00436 (\pm 0,00141) \text{ BP} + 1,79538 (\pm 1,22230)$$

MLR12: (4.17.)

$$\log k_{0,70} = 0,36060 (\pm 0,03857) \text{ ALOGPs} - 0,01049 (\pm 0,00405) \text{ \Delta E} - 0,07152 (\pm 0,03255) \text{ E}_{\text{HOMO}} - 1,38469 (\pm 0,45179)$$

MLR13: (4.18.)

$$\log k_{0,70} = -0,00507 (\pm 0,00049) \text{ MP} - 0,95194 (\pm 0,22326) \text{ S3K} + 0,13289 (\pm 0,02329) \text{ AMR} - 8,78772 (\pm 2,19835)$$

MLR14: (4.19.)

$$\log k_{0,70} = 0,00205 (\pm 0,00089) \text{ DE} - 0,02148 (\pm 0,00219) \text{ PSA} + 0,07228 (\pm 0,01017) \text{ AMR} - 7,18147 (\pm 1,19329)$$



Tabela 4.8. Statistički parametri interne i eksterne validacije MLR–QSRR modela za predviđanje $\log k_{0,90}$ i $\log k_{0,70}$ parametara

Parametri	MLR1	MLR2	MLR3	MLR4	MLR5	MLR6	MLR7
R	0,9766	0,9739	0,9597	0,9634	0,9311	0,9568	0,9570
R^2	0,9538	0,9485	0,9210	0,9282	0,8670	0,9155	0,9159
R^2_a	0,0462	0,0515	0,0790	0,0718	0,1330	0,0845	0,0841
R^2_{adj}	0,9445	0,9382	0,9052	0,9138	0,8504	0,8986	0,8990
F -vrednost	103,20	92,10	58,26	64,65	52,15	54,16	54,43
$RMSE$	0,1043	0,1101	0,1364	0,1300	0,1713	0,1410	0,1407
p -vrednost	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000
VIF	2,13 (Av. logP)	1,71 (XLOGP3)	1,28 (SCAA3)	3,10 (Av. logP)	1,02 (Av. logP)	1,03 (ALOGPs)	1,38 (ALOGPs)
	1,67 (CT)	1,67 (CT)	1,16 (CP)	1,41 (DE)	1,02 (Est)	1,01 (E _{max})	1,02 (E _{max})
	1,38 (DE)	1,05 (Jhetv)	1,32 (DE)	3,05 (MP)		1,04 (E _{gap})	1,39 (E _{homo})
R^2_{CV} (LOO)	0,9191	0,9067	0,8700	0,8439	0,8055	0,8707	0,8752
$ R^2 - R^2_{CV} $	0,0347	0,0418	0,0510	0,0843	0,0615	0,0448	0,0407
TSS	3,5303	3,5303	3,5303	3,5303	3,5303	3,5303	3,5303
$PRESS$	0,2855	0,3292	0,4590	0,5511	0,6868	0,4565	0,4407
$PRESS/TSS$	0,0809	0,0932	0,1300	0,1561	0,1945	0,1293	0,1248
SD_{PRESS}	0,1226	0,1316	0,1554	0,1703	0,1901	0,1550	0,1523
R_{test}	0,9160	0,9494	0,9096	0,9239	0,9714	0,9829	0,9398
$RMSE_{test}$	0,1258	0,1011	0,1684	0,1261	0,0688	0,0612	0,1213

Većina formiranih MLR modela sadrži deskriptore lipofilnosti (Average logP, XLOGP3, ALOGPs). Pored toga, modeli uključuju i kvantno-hemijske deskriptore (ΔE , E_{HOMO} , E_{GAP} , E_{max} , E_{st}), fizičko-hemijske deskriptore (CT, CP, DE, MP, TE, BP, PSA, vdWSA, AMR), topološke deskriptore (S3K, Jhetv) i elektrodeskriptore (FPSA3, SCAA3). Od 14 formiranih modela, 11 se odnosi na predviđanje $\log k_{0,90}$ parametra, dok preostala tri služe za predviđanje $\log k_{0,70}$ parametra.

Formirane MLR jednačine (MLR1, MLR2, MLR4–MLR12) su dobar pokazatelj zavisnosti retencionog ponašanja ispitivanih androstanskih derivata u RP-HPLC sistemu od njihove lipofilnosti, od koje u najvećoj meri zavisi raspodela molekula između stacionarne i mobilne faze. U MLR jednačinama, u kojima se javlja deskriptor lipofilnosti, regresioni koeficijent koji stoji uz ovaj deskriptor ukazuje na njegov najveći uticaj na retenciju u odnosu na ostale deskriptore koji figurišu u modelu. Izuzetak je model MLR8, koji ukazuje na dominantniji uticaj deskriptora delimično naelektrisane pozitivne površine molekula (FPSA3), sa značajno većim regresionim koeficijentom negativnog predznaka. Većina kvantno-hemijskih deskriptora ima negativan regresioni koeficijent, izuzev deskriptora energetske praznine (E_{GAP}). CT, CP, MP, BP, PSA, S3K i Jhetv deskriptori imaju negativan regresioni koeficijent u svim MLR jednačinama, za razliku od TE, DE, AMR, SCAA3 i vdWSA deskriptora.



Tabela 4.8. Statistički parametri interne i eksterne validacije MLR QSRR modela za predviđanje $\log k_{0,90}$ i $\log k_{0,70}$ parametara (*nastavak*)

Parametri	MLR8	MLR9	MLR10	MLR11	MLR12	MLR13	MLR14
R	0,9433	0,9622	0,9644	0,9675	0,9413	0,9576	0,9473
R^2	0,8898	0,9259	0,9301	0,9360	0,8861	0,9170	0,8973
R^2_a	0,1102	0,0741	0,0699	0,0640	0,1139	0,0830	0,1027
R^2_{adj}	0,8760	0,9110	0,9161	0,9232	0,8633	0,9003	0,8767
F -vrednost	64,59	62,43	66,53	73,16	38,90	55,21	43,67
$RMSE$	0,1559	0,1321	0,1283	0,1227	0,1546	0,1320	0,1468
p -vrednost	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000
VIF	1,09 _(ALOGPs)	1,04 _(ALOGPs)	2,53 _(ALOGPs)	2,68 _(ALOGPs)	1,41 _(ALOGPs)	1,01 _(MP)	1,52 _(DE)
	1,09 _(FPSA3)	1,14 _(ΔE)	4,65 _(PSA)	1,44 _(TE)	1,04 _(ΔE)	8,33 _(S3K)	1,43 _(PSA)
		1,16 _(Egap)	2,55 _(vdWSA)	2,06 _(BP)	1,39 _(Ehomo)	8,35 _(AMR)	1,29 _(AMR)
R^2_{CV} (LOO)	0,8501	0,8923	0,8883	0,8957	0,8055	0,8712	0,8486
$ R^2 - R^2_{CV} $	0,0397	0,0336	0,0418	0,0403	0,0806	0,0458	0,0487
TSS	3,5303	3,5303	3,5303	3,5303	3,1465	3,1465	3,1465
$PRESS$	0,5291	0,3803	0,3945	0,3683	0,6121	0,4054	0,4763
$PRESS/TSS$	0,1499	0,1077	0,1117	0,1043	0,1945	0,1288	0,1514
SD_{PRESS}	0,1669	0,1415	0,1441	0,1392	0,1795	0,1461	0,1583
R_{test}	0,9172	0,9548	0,9720	0,9598	0,9074	0,9652	0,9482
$RMSE_{test}$	0,1444	0,0971	0,0989	0,1039	0,1403	0,1005	0,1373

MLR modeli su validovani računanjem statističkih parametara interne i eksterne validacije, koji su prikazani u *tabeli 4.8*. Na osnovu koeficijenata korelacije i determinacije može se zaključiti da svi predstavljeni MLR modeli opisuju veoma jaku zavisnost između varijabli, kao i da se predviđanje $\log k$ parametra postiže sa greškom predviđanja u prihvatljivim granicama. Visoke vrednosti F -parametra ukazuju na dobro slaganje podataka. VIF vrednosti pokazuju da je multikolinearnost prisutna u MLR modelima u granicama tolerancije ($VIF < 10$).

Najjasniji uvid u moć predviđanja modela dobija se primenom eksternog test seta. Parametri eksterne validacije (R_{test} i $RMSE_{test}$) ukazuju na veoma jaku korelaciju između promenljivih, kao i na relativno malu grešku predviđanja.

Poređenje eksperimentalnih i predviđenih $\log k$ vrednosti za kalibracioni i eksterni test set, kao i raspodela ostatka u odnosu na predviđene $\log k$ vrednosti prikazano je na *slikama 4.15. – 4.18*. Dati grafici ukazuju na dobro slaganje eksperimentalnih i predviđenih $\log k$ vrednosti, kao i na dobro uklapanje eksternog test seta u kalibracioni set podataka. Jednačine koje opisuju zavisnosti eksperimentalno-predviđeno prikazane su na datim graficima i odnose se na kalibracioni i eksterni test set. Odsecci ovih zavisnosti su za sve modele veoma bliski nuli, dok su nagibi bliski jedinici, što govori u prilog dobrom slaganju eksperimentalnih i predviđenih vrednosti. Preciznost predviđanja $\log k$ parametra može se proceniti i na osnovu amplitude ostataka, nasumično raspoređenih oko $y = 0$ ose. Model MLR4 je

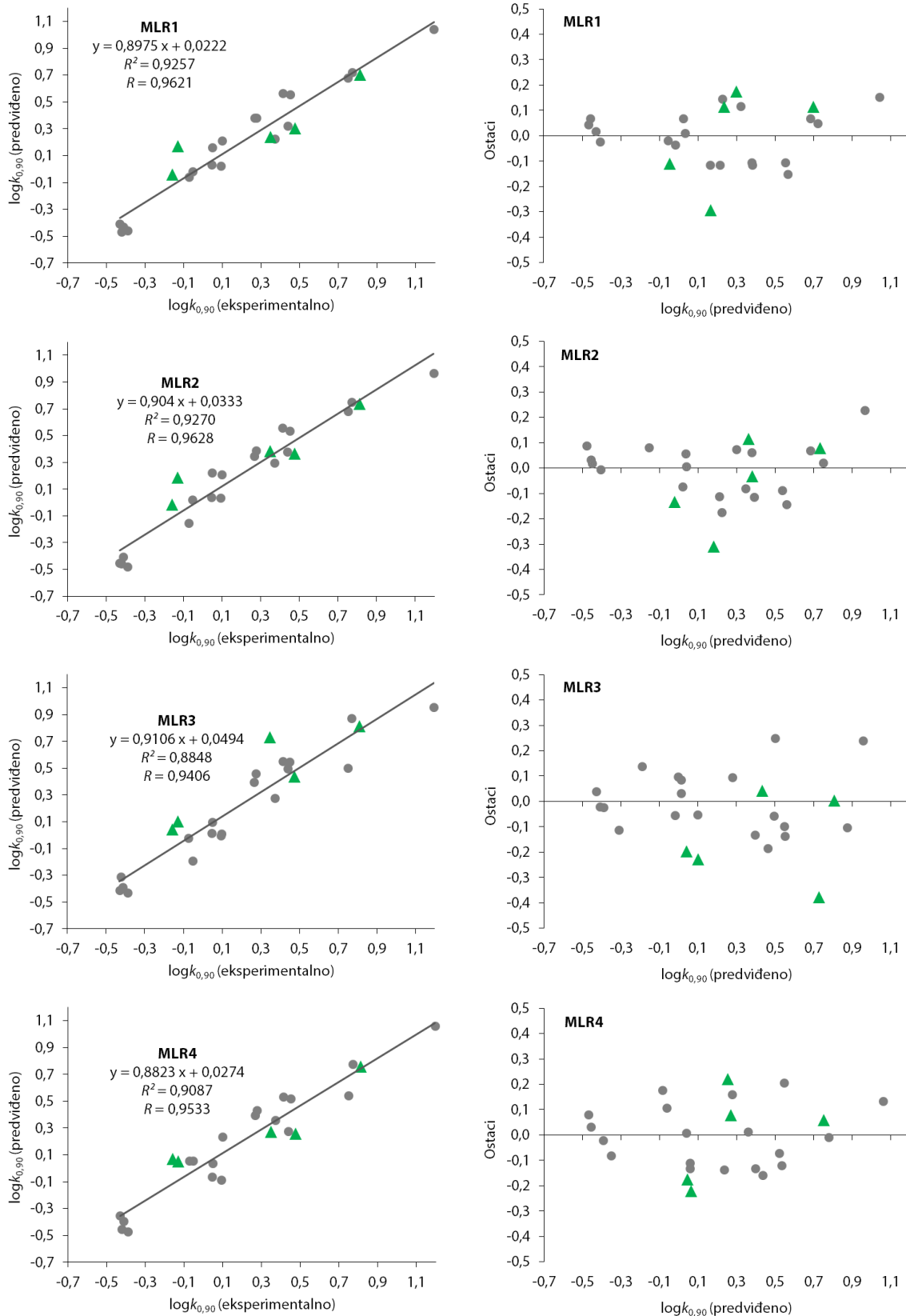


okarakterisan vrednostima ostataka koji se kreću u najnižem opsegu od -0,222 do 0,219. Model koji u ovom pogledu pravi najveću grešku predviđanja je model MLR3 čiji se ostaci nalaze u opsegu od -0,379 do 0,249. Prosečne vrednosti pozitivnih i negativnih ostataka mogu da ukažu na prosečnu grešku predviđanja određenog modela (*tabela 4.9.*). Prema ovom kriterijumu, najmanja prosečna razlika između eksperimentalnih i predviđenih vrednosti javlja se kod modela MLR11, za razliku od modela MLR12 koji je okarakterisan najvećom prosečnom razlikom eksperimentalnih i predviđenih vrednosti.

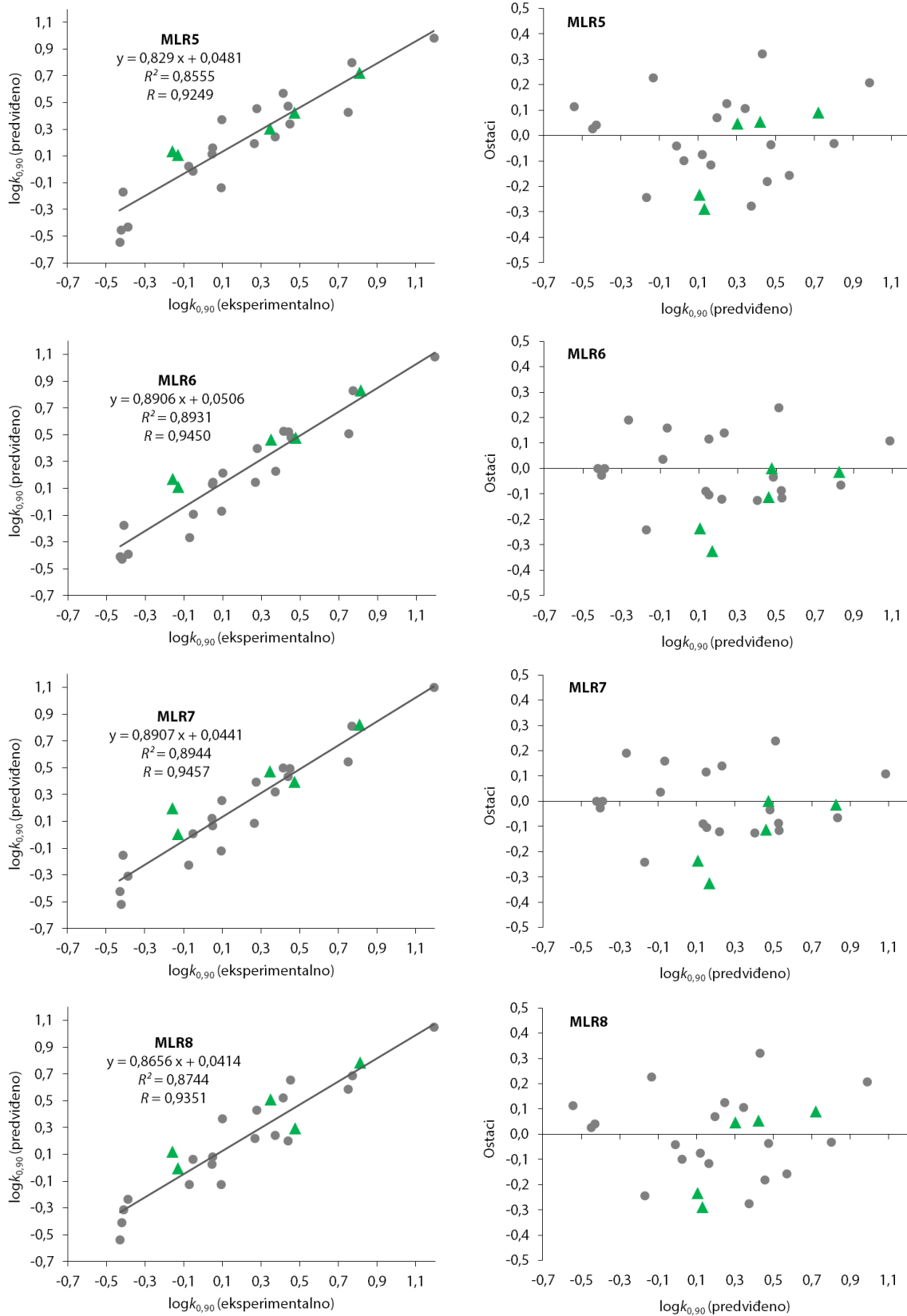
Tabela 4.9. Prosečne (srednje) vrednosti pozitivnih i negativnih ostataka za MLR modele

Modeli	Prosečna vrednost pozitivnih ostataka	Prosečna vrednost negativnih ostataka
MLR1	0,088	-0,108
MLR2	0,072	-0,116
MLR3	0,102	-0,127
MLR4	0,106	-0,144
MLR5	0,119	-0,147
MLR6	0,111	-0,113
MLR7	0,132	-0,099
MLR8	0,115	-0,144
MLR9	0,101	-0,108
MLR10	0,103	-0,101
MLR11	0,094	-0,088
MLR12	0,120	-0,150
MLR13	0,111	-0,102
MLR14	0,136	-0,081

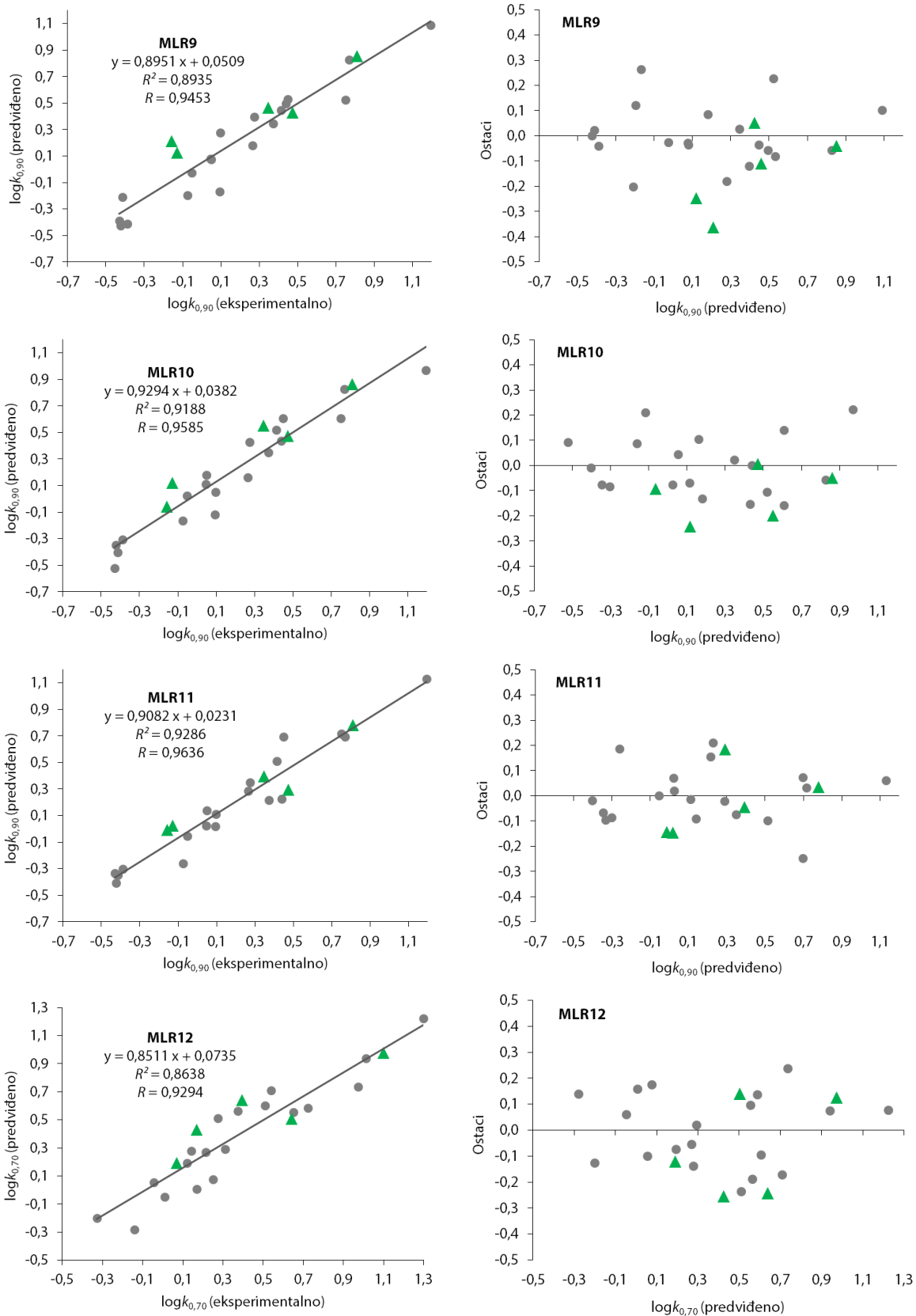
U poređenju sa LR i PR modelima, MLR modeli imaju bolje statističke parametre, što im daje prednost u predviđanju retencionih parametara ispitivanih androstanskih derivata.



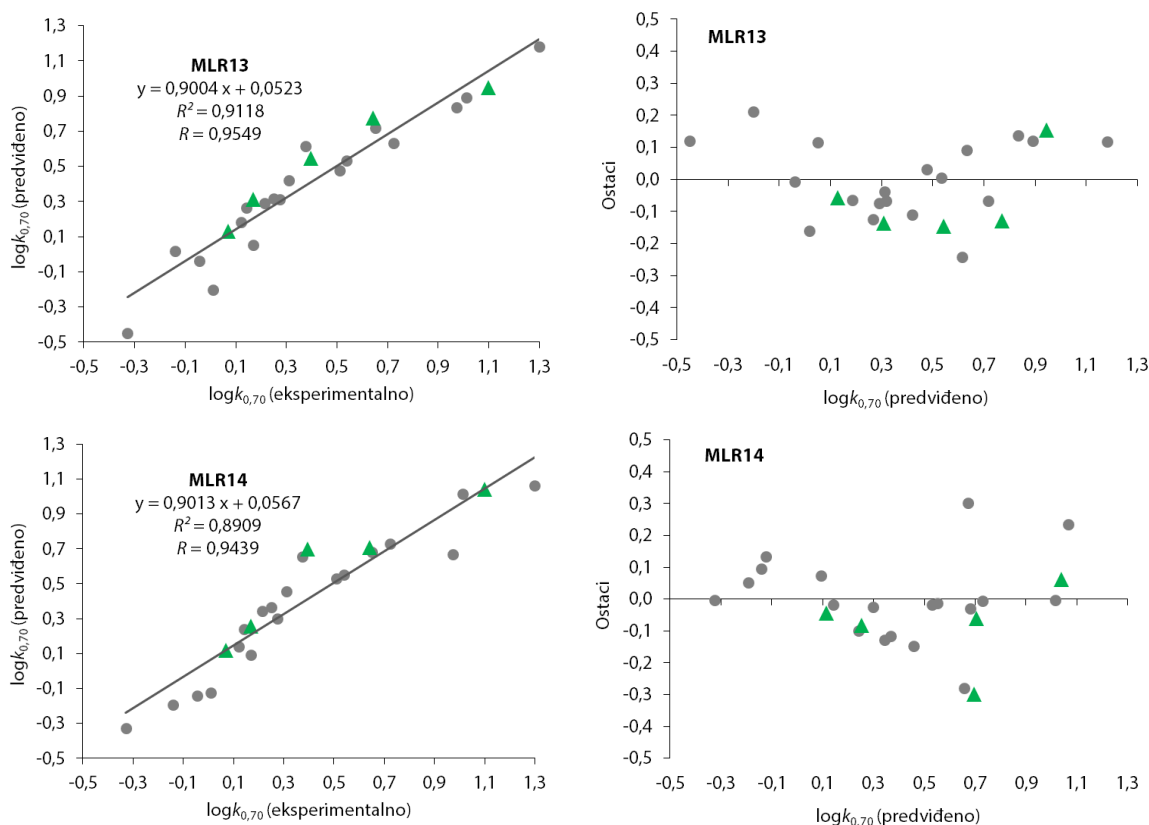
Slika 4.15. Poređenje eksperimentalnih i predviđenih $\log k$ parametara i distribucija ostataka oko $y = 0$ ose za MLR1, MLR2, MLR3 i MLR4 modele (● - kalibracioni set, ▲ - eksterni test set)



Slika 4.16. Poređenje eksperimentalnih i predviđenih logk parametara i distribucija ostataka oko $y = 0$ ose za MLR5, MLR6, MLR7 i MLR8 modele (● - kalibracioni set, ▲ - eksterni test set)



Slika 4.17. Poređenje eksperimentalnih i predviđenih $\log k$ parametara i distribucija ostataka oko $y = 0$ ose za MLR9, MLR10, MLR11 i MLR12 modele (● - kalibracioni set, ▲ - eksterni test set)



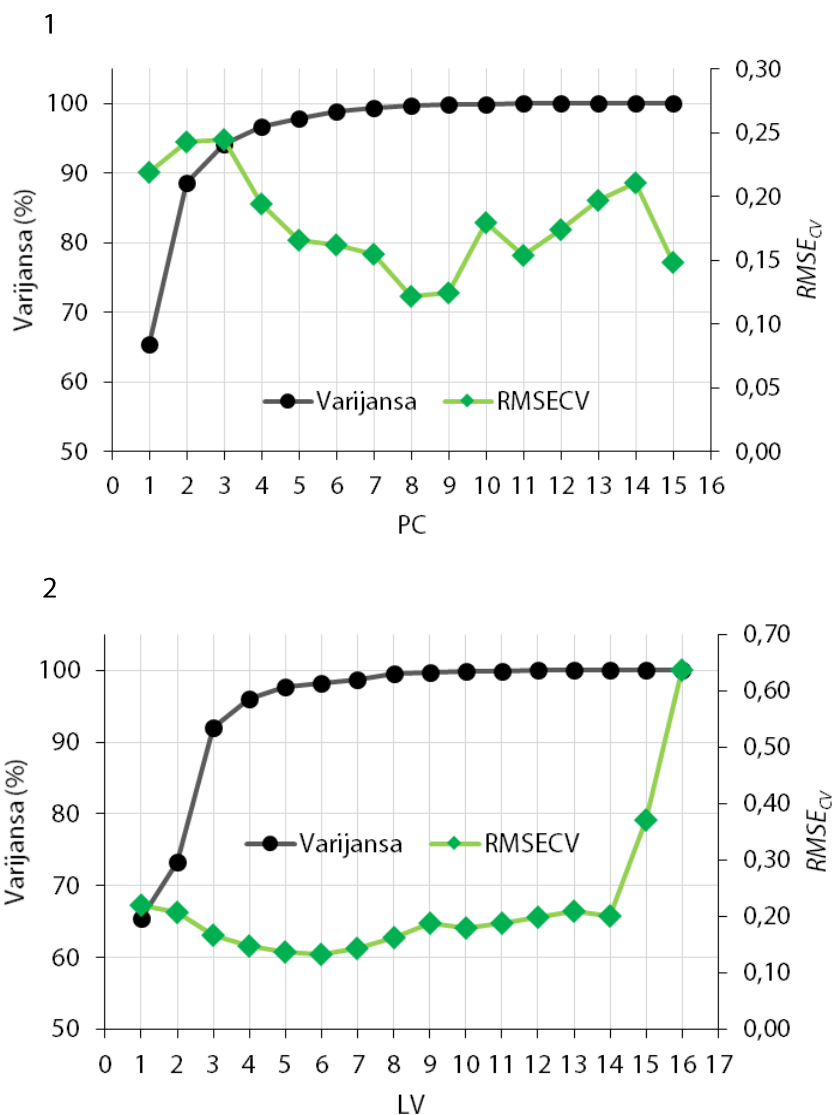
Slika 4.18. Poređenje eksperimentalnih i predviđenih $\log k$ parametara i distribucija ostataka oko $y = 0$ ose za MLR13 i MLR14 modele (● - kalibracioni set, ▲ - eksterni test set)

S obzirom na potencijalni problem pojave značajne multikolinearnosti, koja bi destabilizovala MLR modele u slučaju uključivanja međusobno jako koreliranih deskriptora, u narednom koraku regresione analize primenjene su regresija glavnih komponentata i regresija metodom parcijalnih najmanjih kvadrata, čiji je uslov primene upravo jaka korelacija između varijabli, s tim što kod PLS modela varijable koje su jako korelirane sa zavisno promenljivom dodatno dobijaju na značaju. PCR i PLS modelovanje izvedeno je pomoću programa MATLAB R2013a sa PLS_Toolbox dodatkom. Na osnovu korelacione matrice birani su setovi deskriptora koji su jako korelirani, a potom su primenjeni kao nezavisno promenljive. Najbolji PCR i PLS modeli dobijeni su primenom seta koji sadrži trinaest deskriptora lipofilnosti (ALOGPs, AClogP, ALOGP, MLOGP, KOWWIN, XLOGP2, XLOGP3, Average logP, miLogP, $\log P_{vgr}$, $\log P_{klopr}$, $\log P_{physr}$, $\log P_{wgt}$) i pet fizičko-hemijskih deskriptora (vdWV, PSA, vdWSA, SASA1.4, MR).

Odabir broja glavnih komponentata (PC) kod PCR modela, odnosno latentnih varijabli (LV) kod PLS modela, izveden je na osnovu najmanje $RMSE$ vrednosti LOO unakrsne validacije ($RMSE_{CV}$). U slučaju PCR modela najmanja $RMSE_{CV}$ vrednost odgovara broju od osam glavnih komponentata, koje obuhvataju 99,66% varijanse u nezavisno promenljivim varijablama. PLS model uključuje šest latentnih varijabli koje obuhvataju 98,20% varijanse u nezavisno promenljivim varijablama. Odabir broja PC, odnosno LV, na osnovu najmanje $RMSE_{CV}$ vrednosti prikazan je grafički na slici 4.19. Regresioni koeficijenti varijabli koje ulaze u PCR i PLS modele dati su u tabeli 4.10.



Najbolji PCR i PLS modeli dobijeni su u slučaju predviđanja $\log k_{0,90}$ parametra, dok su za predviđanje $\log k_{0,70}$ dobijeni modeli sa veoma velikom greškom predviđanja, te nisu uzeti u dalje razmatranje.



Slika 4.19. Zavisnost varijanse nezavisno promenljivih varijabli i $RMSE_{CV}$ od broja PC kod PCR modela (1) i broja LV kod PLS modela (2)

Na osnovu prikazanih regresionih koeficijenata, može se uočiti da deskriptori lipofilnosti $ALOGPs$ i $\log P_{klop}$ imaju najveći pozitivan uticaj na vrednosti zavisno promenljive, dok najveći negativan uticaj ispoljava $ALOGP$ deskriptor. Ukoliko se uporede vrednosti regresionih koeficijenata koje molekularni deskriptori imaju u PCR modelu sa njihovim regresionim koeficijentima dodeljenim u PLS modelu, može se uočiti da su ove razlike veoma male, što u ovom koraku analize može da nagovesti sličnost ova dva modela u predviđanju zavisno promenljive.

**Tabela 4.10.** Regresioni koeficijenti nezavisno promenljivih varijabli koje figurišu u PCR i PLS modelima

Nezavisno promenljiva	Regresioni koeficijenti (PCR)	Regresioni koeficijenti (PLS)
ALOGPs	0,20582	0,18288
AClogP	-0,03762	-0,04956
ALOGP	-0,12188	-0,12961
MLOGP	-0,02927	-0,07085
KOWWIN	0,04317	0,07440
XLOGP2	0,13042	0,08461
XLOGP3	0,05198	0,04586
Average logP	0,03302	0,03667
miLogP	0,00439	0,04867
$\log P_{vg}$	-0,05226	-0,09885
$\log P_{klop}$	0,17335	0,18071
$\log P_{phys}$	-0,00812	0,00024
$\log P_{wgt}$	0,01060	0,04618
vdWV	0,00227	0,00197
PSA	-0,01006	-0,01136
vdWSA	0,00225	0,00198
SASA1.4	-0,00016	-0,00032
MR	-0,01133	-0,00366
Intercept	-1,90881	-2,16348

Sličnost PCR i PLS modela uočljiva je i ukoliko se uporede njihove statističke karakteristike, koje su prikazane u tabeli 4.11. Iako se prema statističkim parametrima kalibracije prednost može dati PLS modelu, PCR model ipak ima prednost posmatrajući validacione parametre. Ove statističke razlike između PCR i PLS modela su male, te se oba modela mogu prihvatiti kao statistički validna sa relativno malom greškom predviđanja $\log k_{0,90}$ vrednosti. Koeficijenti korelacije i determinacije za kalibracioni i eksterni test set ukazuju na veoma jaku korelaciju između eksperimentalnih $\log k_{0,90}$ vrednosti i $\log k_{0,90}$ vrednosti predviđenih PCR i PLS modelima.


Tabela 4.11. Statistički parametri interne i eksterne validacije PCR i PLS modela za predviđanje $\log k_{0,90}$

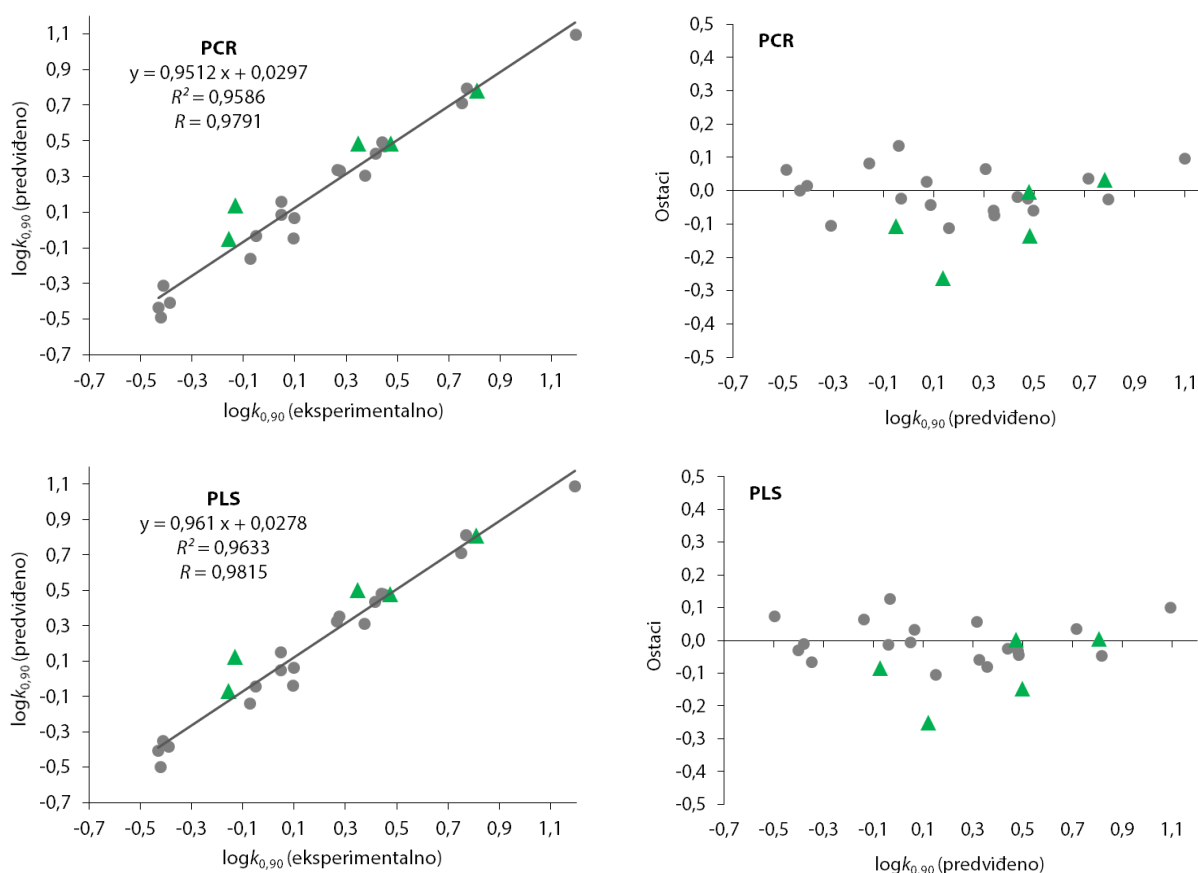
Parametri	PCR	PLS
$R^2 Y_{kumul}$	0,9754	0,9789
$Q^2 Y_{kumul}$	0,9189	0,9047
R	0,9879	0,9896
R^2	0,9760	0,9793
R_{test}	0,9777	0,9759
R^2_{test}	0,9558	0,9524
R^2_{CV}	0,9210	0,9075
$RMSE$	0,0667	0,0619
$RMSE_{CV}$	0,1213	0,1317
$RMSE_{test}$	0,1418	0,1357
$PRESS/TSS$	0,0790	0,0925
F -vrednost (eksperimentalno-predviđeno)	512,00	577,24
p -vrednost (eksperimentalno-predviđeno)	0,000000	0,000000

Grafičko poređenje eksperimentalnih i predviđenih $\log k_{0,90}$ vrednosti prikazano je na slici 4.20. Oba modela su okarakterisana malim rasipanjem tačaka oko prave, visokim koeficijentom korelacije i nagibom ove zavisnosti koji je veoma blizak jedinici, i odsečkom bliskim nuli, što govori u prilog dobrom slaganju eksperimentalnih i predviđenih vrednosti. Sa ovog aspekta poređenja modela, neznatna prednost se može dati PLS modelu. Takođe, može se uočiti i dobro uklapanje eksternog test seta u kalibracioni set podataka kod oba modela. Slika 4.20. pokazuje i nasumičnu raspodelu ostataka oko $y = 0$ ose, kao i njihove relativno niske vrednosti. Ostaci se u slučaju PCR modela nalaze u opsegu od -0,265 do 0,135, a u slučaju PLS modela od -0,251 do 0,128. Prosečne vrednosti pozitivnih i negativnih ostataka za PCR i PLS modele date su u tabeli 4.12. One ukazuju na to da je prosečna razlika između eksperimentalnih i predviđenih vrednosti neznatno manja u slučaju PLS modela.

Tabela 4.12. Prosečne (srednje) vrednosti pozitivnih i negativnih ostataka za PCR i PLS modele

Model	Prosečna vrednost pozitivnih ostataka	Prosečna vrednost negativnih ostataka
PCR	0,056	-0,074
PLS	0,056	-0,066

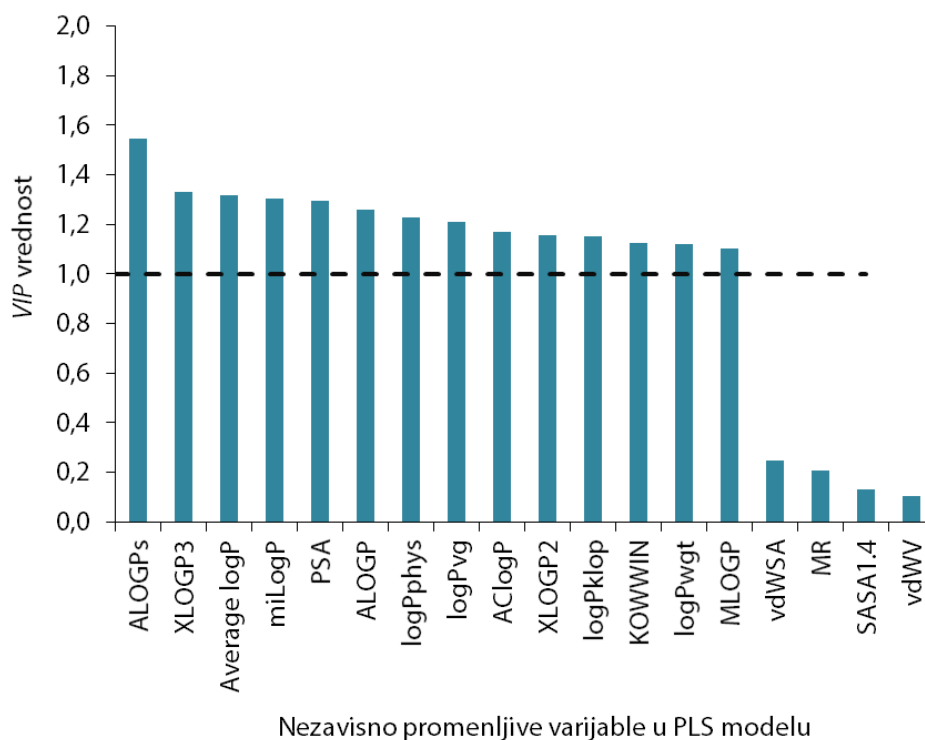
Vrednosti fizičko-hemijskih, topoloških i elektrostatičkih deskriptora prisutnih u linearnim QSRR modelima dati su u prilogu u tabeli P-7.4.



Slika 4.20. Poređenje eksperimentalnih i predviđenih $\log k_{0,90}$ parametara i distribucija ostataka oko $y = 0$ ose za PCR i PLS modele (● - kalibracioni set, ▲ - eksterni test set)

Da bi se utvrdio značaj nezavisno promenljivih varijabli u pogledu njihovog udela u objašnjenju zavisno promenljive, računane su njihove *VIP* vrednosti na osnovu *formule P-7.5.* date u prilogu. *VIP* vrednosti varijabli primenjenih u PLS modelu prikazane su na *slici 4.21.* Varijable čija je *VIP* vrednost veća od 1, smatraju se najznačajnijim. U slučaju dobijenog PLS modela za predviđanje $\log k_{0,90}$ parametra ispitivanih androstanskih derivata, parametri lipofilnosti najviše doprinose objašnjenju retencionog parametra, dok se od ostalih pet fizičko-hemijskih deskriptora jedino izdvaja polarna površina (PSA) kao najznačajnija varijabla. Analizom *VIP* vrednosti dobijena je još jedna potvrda dominantnog uticaja lipofilnosti na retenciono ponašanje 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata, ali ukazuje i na značajan doprinos PSA deskriptora u objašnjenju retencionog faktora.

Na osnovu iznetih rezultata linearnog modelovanja, može se zaključiti da je najbolje slaganje eksperimentalnih i predviđenih $\log k_{0,90}$ vrednosti postignuto pomoću PCR i PLS modela, dok se predviđanje $\log k_{0,70}$ parametra može uspešno izvesti primenom MLR modela.



Slika 4.21. VIP vrednosti nezavisno promenljivih varijabli koje figurišu u PLS modelu

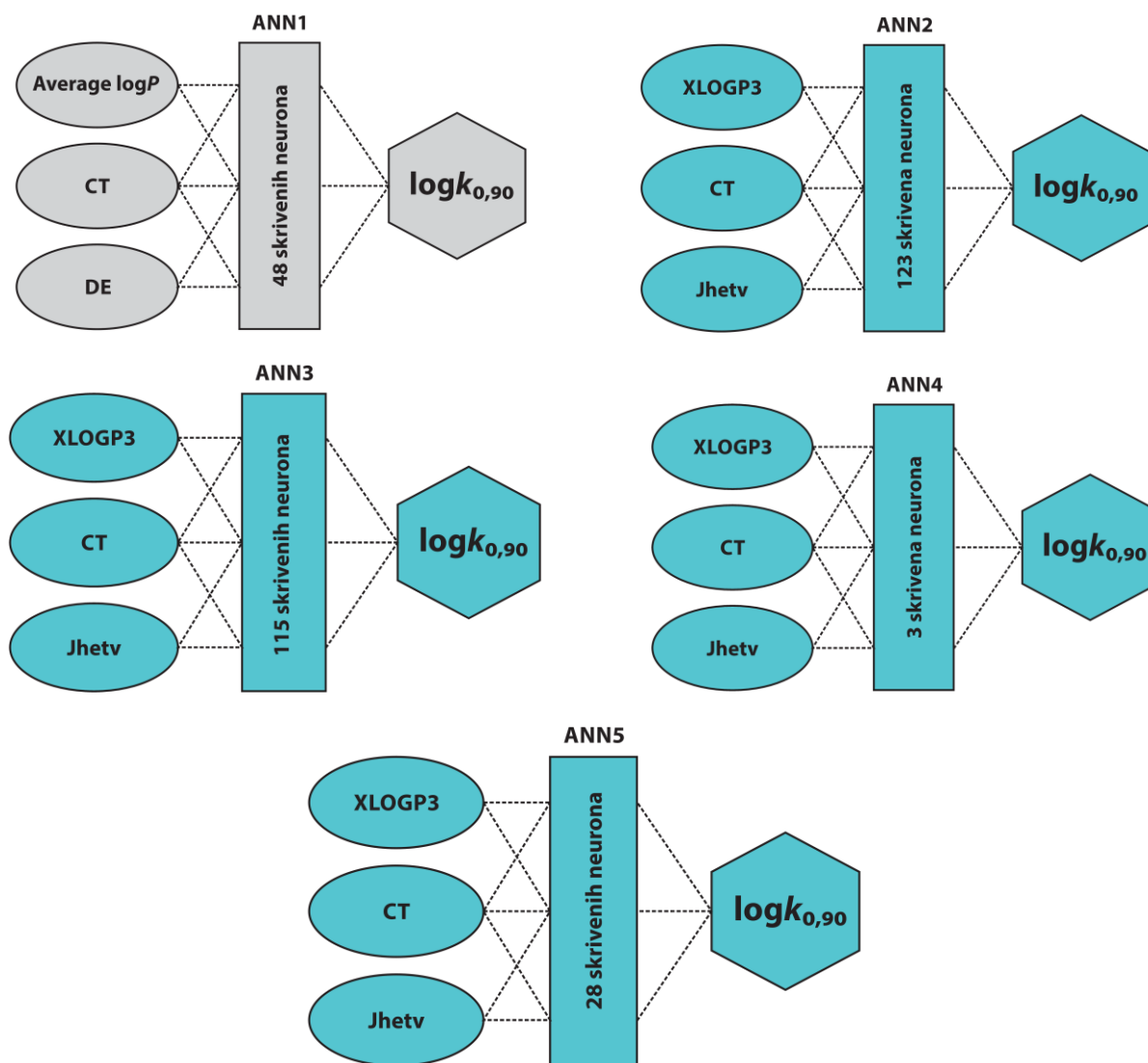
4.4.2. Nelinearno QSRR modelovanje

Nakon formiranja linearnih QSRR modela, izvedeno je nelinearno QSRR modelovanje primenom regresione metode zasnovane na veštačkim neuronskim mrežama. ANN modelovanje je izvedeno pomoću programa Statistica 10.0, primenom *Automated Network Search (ANS)* opcije. Celokupni set podataka je podeljen na tri seta – kalibracioni (jedinjenja **1-6, 8, 9, 11, 13, 15, 17-23**), test (jedinjenja **7, 10, 16**) i validacioni set (jedinjenja **12, 14, 24**). Formiranje setova izvedeno je nasumičnim odabirom jedinjenja. Kako je ranije napomenuto, test set je neophodan za definisanje opšte greške modela, a validacioni set za pronalaženje najbolje konfiguracije mreže poređenjem greške validacije i greške treniranja mreže. Stoga, treniranje mreža je izvedeno na osnovu osamnaest jedinjenja, dok su preostalih šest jedinjenja upotrebljeni za testiranje (tri jedinjenja) i validaciju (tri jedinjenja). S obzirom na to da je pri ANN modelovanju neophodno koristiti tri seta jedinjenja, za razliku od linearnog modelovanja kod koga su primenjena dva seta jedinjenja (kalibracioni i eksterni test set), razlika u načinima validacije QSRR modela kod ova dva pristupa je očigledna. ANN modeli su formirani na osnovu kombinacija varijabli koje se nalaze u MLR modelima. Ulazne varijable su pre treniranja mreža normalizovane pomoću *min-max* metode normalizacije. Za svaku kombinaciju varijabli trenirano je 2000 mreža, tako da je ukupan broj



treniranih mreža iznosio 28 000. Najbolji modeli su odabrani na osnovu maksimalnih R_{kalibr} , R_{test} , R_{valid} i minimalnih $RMSE_{kalibr}$, $RMSE_{test}$, $RMSE_{valid}$ vrednosti kalibracionog, test i validacionog seta. Pri treniranju broj skrivenih neurona variran je od 2 do 200. Treniranje višeslojnih (MLP) nerekurentnih (*feedforward*) mreža obavljeno je primenom BFGS algoritma i kombinacijom sinusne (*Sine*), logističke (*Lgt*), hiperbolične tangens (*Tanh*), eksponencijalne (*Exp*) funkcije i funkcije identiteta (*Idt*), kao aktivacionih funkcija skrivenog i izlaznog sloja. Neuronske težine su određene pomoću primenjenog softvera.

ANN modelovanjem dobijeno je pet najboljih modela za predviđanje $\log k_{0,90}$ parametra androstanskih derivata, čija je arhitektura prikazana na slici 4.22. Model ANN1 sadrži ulazne varijable koje su prisutne u modelu MLR1 (Average $\log P$, CT i DE), dok modeli ANN2, ANN3, ANN4 i ANN5 imaju ulazne varijable prisutne u modelu MLR2 (XLOGP3, CT i Jhetv). Arhitektura formiranih ANN modela se značajno razlikuje u broju neurona u skrivenom sloju. Neophodno je naglasiti da ANN modelovanje $\log k_{0,70}$ faktora nije rezultovalo prihvatljivim modelom.



Slika 4.22. Arhitektura formiranih ANN modela



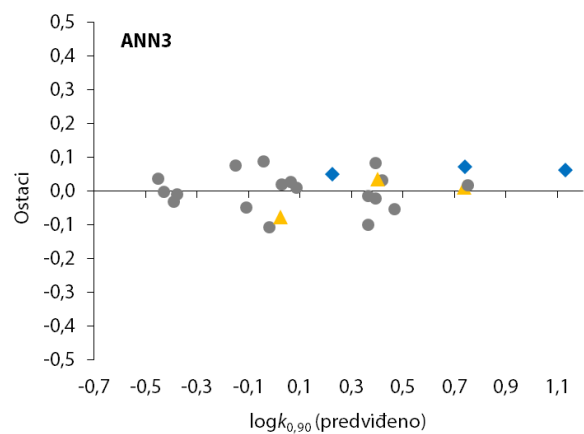
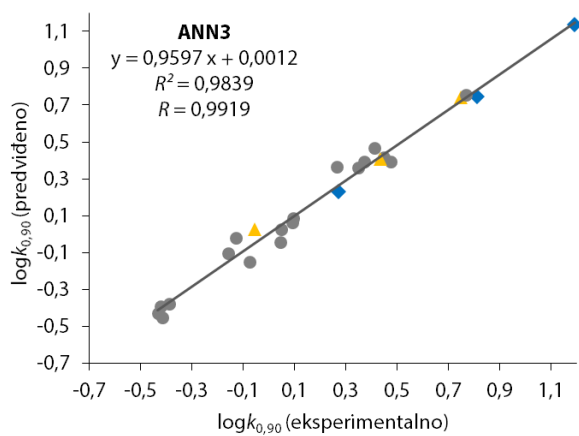
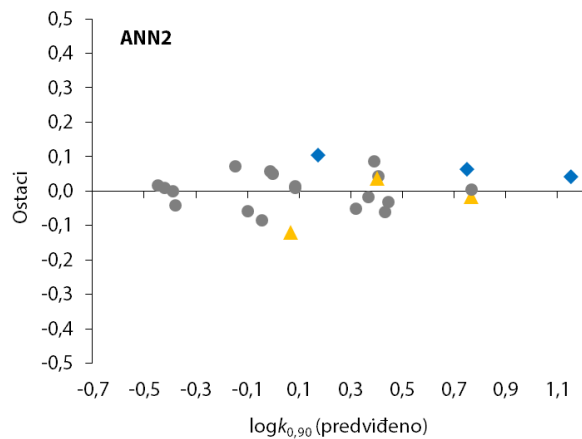
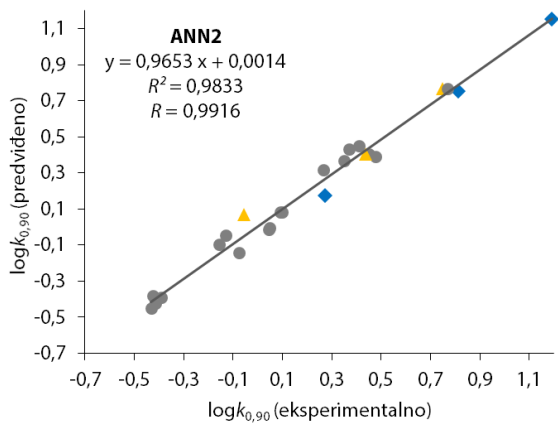
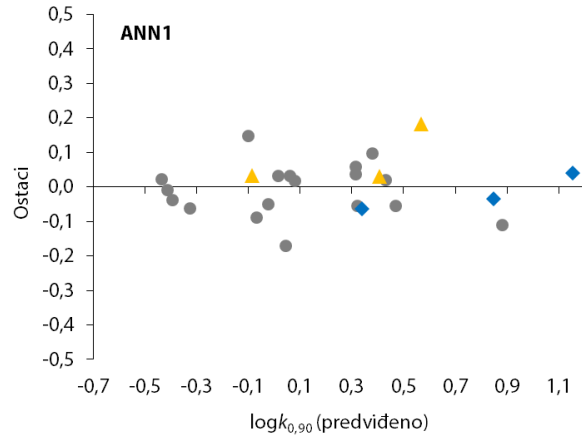
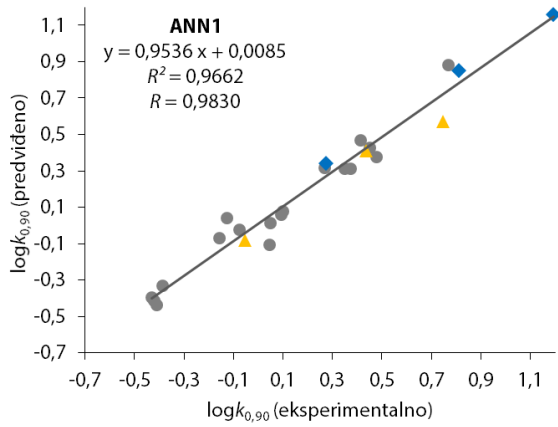
Procena prediktivne sposobnosti formiranih mreža izvedena je na osnovu statističkih parametara datih u *tabeli 4.13*. Na osnovu predstavljenih rezultata može se uočiti da formirani ANN modeli imaju odlične statističke karakteristike, uključujući veoma jaku korelaciju između eksperimentalnih i predviđenih $\log k_{0,90}$ vrednosti, kako za kalibracioni set, tako i za test i validacioni set podataka. Takođe, veoma visoke F -vrednosti ukazuju na odlično slaganje eksperimentalnih i predviđenih vrednosti. Dobijene statističke karakteristike ANN modela ukazuju na to da se najmanja greška predviđanja upravo postiže primenom nelinearnih ANN modela u odnosu na predstavljene linearne modele.

Tabela 4.13. Statističke karakteristike formiranih ANN modela

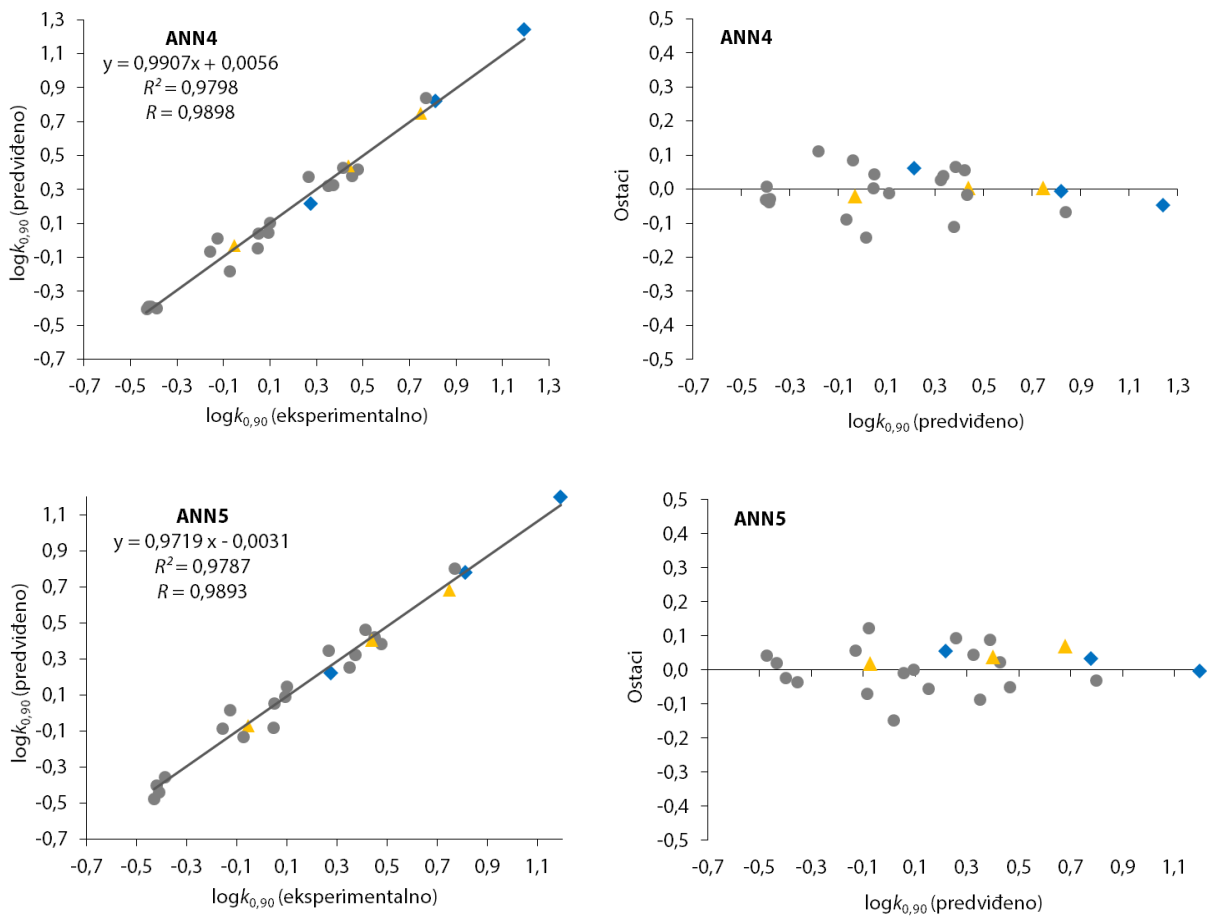
Parametri	ANN1	ANN2	ANN3	ANN4	ANN5
Arhitektura	3-48-1	3-123-1	3-115-1	3-3-1	3-28-1
R_{kalib}	0,9766	0,9906	0,9877	0,9815	0,9810
R_{test}	0,9875	0,9881	0,9956	0,9999	0,9998
R_{valid}	0,9990	1,0000	0,9998	1,0000	0,9999
$RMSE_{kalib}$	0,003	0,001	0,002	0,002	0,002
$RMSE_{test}$	0,006	0,003	0,001	0,000	0,001
$RMSE_{valid}$	0,001	0,003	0,002	0,001	0,001
F -vrednost (eksperimentalno-predviđeno)	630,43	1288,03	1353,00	1061,87	1015,55
p -vrednost (eksperimentalno-predviđeno)	0,000000	0,000000	0,000000	0,000000	0,000000
Algoritam	BFGS (36)*	BFGS (56)	BFGS (35)	BFGS (29)	BFGS (39)
Skrivena aktivaciona funkcija	<i>Exp</i>	<i>Tanh</i>	<i>Tanh</i>	<i>Lgt</i>	<i>Lgt</i>
Izlazna aktivaciona funkcija	<i>Sine</i>	<i>Idt</i>	<i>Idt</i>	<i>Sine</i>	<i>Sine</i>

* broj u zagradi predstavlja broj ciklusa treniranja nakon kojih je dobijena mreža optimalne arhitekture

Kvalitet dobijenih ANN modela potvrđen je grafičkim poređenjem eksperimentalnih i predviđenih $\log k_{0,90}$ vrednosti, uključujući i analizu distribucije ostataka, prikazanim na *slici 4.23*. i *slici 4.24*. Nagib linearne zavisnosti eksperimentalno-predviđeno je veoma blizak jedinici, a odsečak veoma blizak nuli, što govori o odličnom slaganju poređenih podataka. Takođe, može se uočiti i veoma malo rasipanje tačaka oko prave linearne zavisnosti. Nasumična raspodela ostataka oko $y = 0$ ose je očigledna, a amplituda ostataka je generalno manja u odnosu na ostatke linearnih modela. Za ANN1 model ostaci se kreću u opsegu od -0,171 do 0,181, za ANN2 od -0,121 do 0,102, za ANN3 od -0,107 do 0,089, za ANN4 od -0,142 do 0,110, za ANN5 od -0,146 do 0,124. Sa aspekta amplitude ostataka, model ANN3 pravi najmanju razliku između eksperimentalnih i predviđenih vrednosti. Ukoliko se uporede prosečne vrednosti pozitivnih i negativnih ostataka (*tabela 4.14.*), i u tom pogledu prednost se može dati modelu ANN3 (najmanja razlika između prosečnih pozitivnih i prosečnih negativnih ostataka). Neophodno je naglasiti da se modeli ANN2, ANN3, ANN4 i ANN5 vrlo malo razlikuju u pogledu statističkih karakteristika i greške predviđanja. Međutim, za model ANN1 može se reći da pravi nešto veću, ali ipak prihvatljivu grešku predviđanja u odnosu na ostale ANN modele. U poređenju sa linearnim QSRR modelima, neuronske mreže su se pokazale kao preciznije rešenje u predviđanju $\log k_{0,90}$ parametra ispitivanih androstanskih derivata. Njihova prednost se ogleda u generalno manjoj amplitudi i prosečnim vrednostima ostataka, većim koeficijentima korelacije i manjim $RMSE$ vrednostima za kalibracioni, test i validacioni set.



Slika 4.23. Poređenje eksperimentalnih i predviđenih $\log k_{0,90}$ parametara i distribucija ostataka oko $y = 0$ ose za ANN1, ANN2 i ANN3 modele (● - kalibracioni set, ▲ - test set, ◆ - validacioni set)



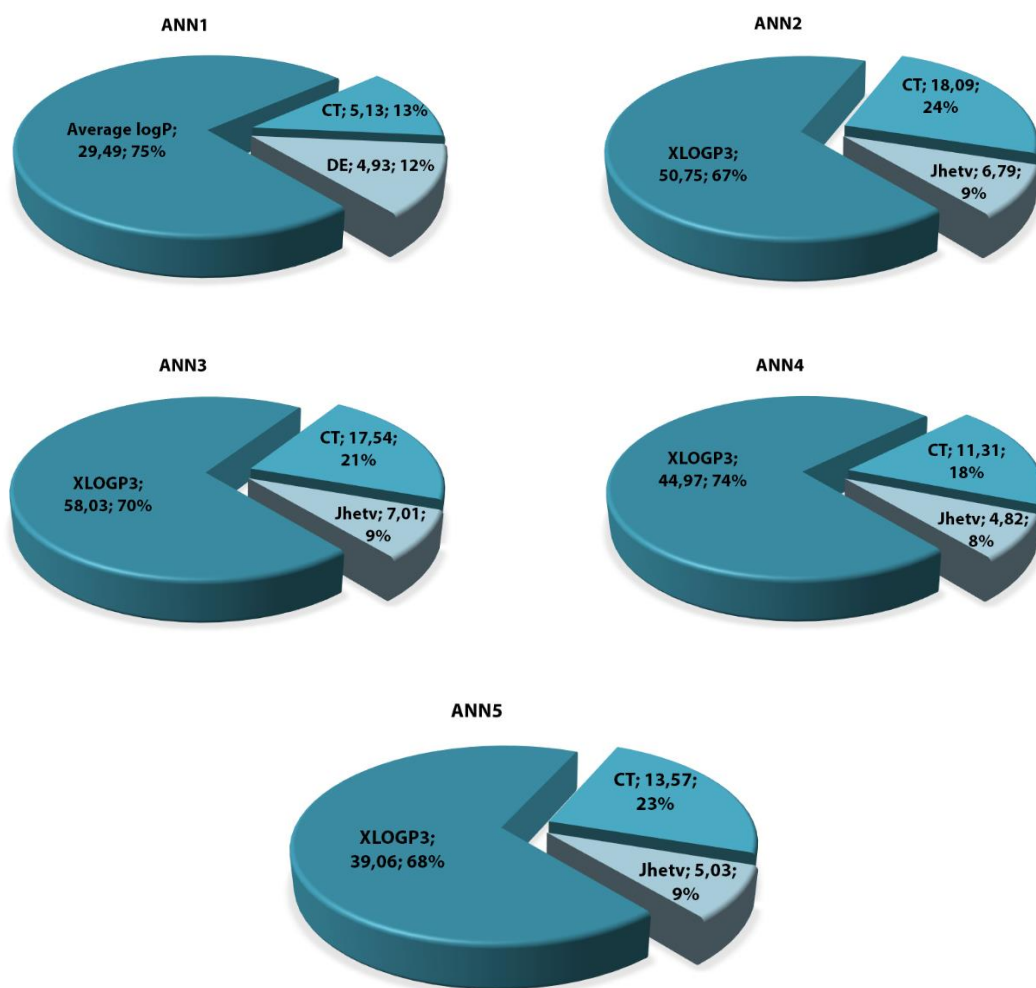
Slika 4.24. Poređenje eksperimentalnih i predviđenih $\log k_{0,90}$ parametara i distribucija ostataka oko $y = 0$ ose za ANN4 i ANN5 modele (● - kalibracioni set, ▲ - test set, ◆ - validacioni set)

Tabela 4.14. Prosečne (srednje) vrednosti pozitivnih i negativnih ostataka za ANN modele

Model	Prosečna vrednost pozitivnih ostataka	Prosečna vrednost negativnih ostataka
ANN1	0,058	-0,067
ANN2	0,041	-0,053
ANN3	0,045	-0,046
ANN4	0,043	-0,050
ANN5	0,051	-0,050



Uticaj svake ulazne varijable na varijacije vrednosti parametara dobijenih ANN modela procenjen je na osnovu analize globalne osetljivosti. S obzirom na to da su GSA koeficijenti ulaznih varijabli svih formiranih ANN modela veći od 1, svaka od varijabli ima značajan uticaj na zavisno promenljivu (slika 4.25.). Na osnovu prikazanih dijagrama može se zaključiti da najveći uticaj na $\log k_{0,90}$ vrednosti ispoljavaju deskriptori lipofilnosti (Average $\log P$ i XLOGP3), a potom CT, Jhetv i DE deskriptori. Ovo je još jedna potvrda značajnog uticaja lipofilnosti na retenciju analiziranih androstanskih derivata u primenjenom hromatografskom sistemu, ali i dodatna potvrda teze da se hromatografska retencija može smatrati alternativnim parametrom lipofilnosti ispitivane serije jedinjenja.

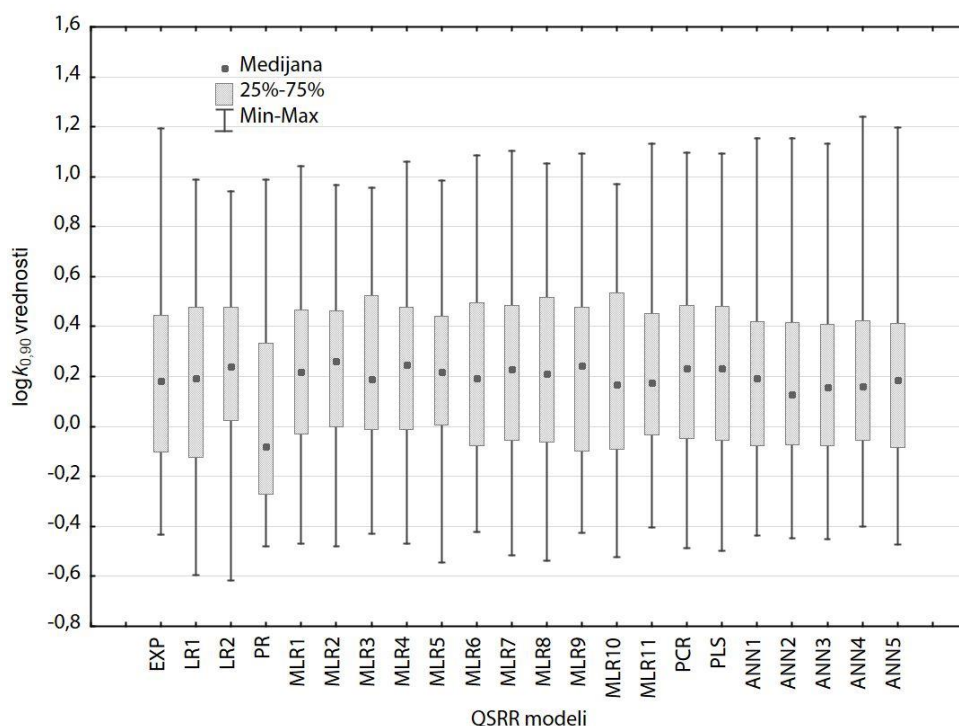


Slika 4.25. GSA koeficijenti ulaznih varijabli ANN modela



4.4.3. Rangiranje QSRR modela

Poređenje formiranih QSRR modela na osnovu izračunatih statističkih parametara može da bude problematičan zadatak. Generalno, može se uočiti da LR i PR modeli imaju najlošije statističke karakteristike u odnosu na ostale (ali u prihvatljivim granicama), potom slede MLR modeli, zatim PCR i PLS, a potom ANN modeli kao najbolji. Prilikom poređenja QSRR modela često se postavlja pitanje u odnosu na koji statistički kriterijum je poređenje najrelevantnije. Neretko se dešava da jedan model ima prednost u odnosu na koeficijent korelacije, dok drugi model daje manju grešku predviđanja, uprkos manjem koeficijentu korelacije. Zato je neophodno uporediti, odnosno rangirati modele i sa drugih aspekata. *Box-Whisker* grafik, prikazan na slici 4.26, ukazuje na izdvajanje PR modela od ostalih prema opsegu predviđenih $\log k_{0,90}$ vrednosti, međutim ukazuje i na teškoće u međusobnom razlikovanju ostalih QSRR modela. Osim klasičnog načina poređenja modela na osnovu vrednosti izračunatih statističkih parametara, razvijen je i relativno nov način rangiranja modela – rangiranje na osnovu razlike sume rangova, tzv. SRD metod, koji je prethodno opisan u teorijskom delu (podpoglavlje 2.7.1.).

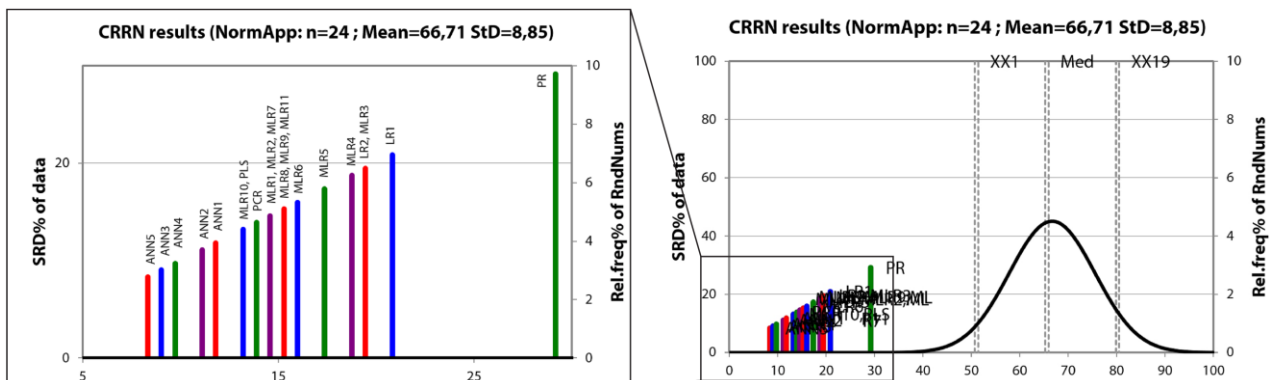


Slika 4.26. *Box-Whisker* grafik eksperimentalnih $\log k_{0,90}$ vrednosti (EXP) i $\log k_{0,90}$ vrednosti predviđenih pomoću formiranih QSRR modela

Rangiranje QSRR modela SRD metodom izvedeno je u odnosu na dva različita idealna ranga – rang definisan na osnovu eksperimentalnih $\log k_{0,90}$ vrednosti, i rang definisan na osnovu prosečnih $\log k_{0,90}$ vrednosti (*row average*). Neophodno je naglasiti da su rangirani modeli koji predviđaju $\log k_{0,90}$ vrednosti, dok rangiranje QSRR modela koji predviđaju $\log k_{0,70}$ parametar nije izvedeno, s obzirom na relativno mali broj ovih modela (pet).



Rangiranje QSRR modela u odnosu na eksperimentalne vrednosti, uzete za referentni rang, prikazano je na slici 4.27. Idealni rang je definisan redosledom analiziranih jedinjenja prema rastućim eksperimentalnim $\log k_{0,90}$ vrednostima. Modeli se porede tako što se jedinjenja poređaju po rastućim predviđenim $\log k_{0,90}$ vrednostima, uključujući kalibracioni i eksterni test set kod LR, PR, MLR, PCR i PLS modela, i kalibracioni, validacioni i test set kod ANN modela, tako da ukupan broj jedinjenja na osnovu kojih se izvodi rangiranje iznosi 24 za sve rangirane QSRR modele. Svako odstupanje od referentnog „idealnog“ redosleda, dovodi do udaljavanja modela od referentnog ranga za određeni stepen. ANN modeli se nalaze najbliže referentnom rangu u odnosu na ostale, dok se PR model nalazi na najvećoj udaljenosti od referentnog ranga. Pomoću ANN modela je i postignuto najbolje fitovanje podataka, stoga je njihov položaj u blizini referentnog ranga na SRD grafiku u potpunosti opravdan. PLS i PCR modeli su najbliži ANN modelima u odnosu na većinu MLR modela, što ukazuje na njihovu prednost u pogledu fitovanja podataka. Model MLR10 ima isti rang kao PLS model, te se ova dva modela mogu smatrati podjednako kvalitetnim. LR1, LR2, MLR3 i MLR4 modeli su predstavljeni kao modeli čiji je rang značajno udaljen od idealnog ranga. Neophodno je istaći da rangiranje modela u odnosu na eksperimentalne vrednosti nije izvedeno sa ciljem opisivanja greške predviđanja, nego fitovanja podataka. Validacija SRD metode je izvedena na osnovu teorijske raspodele slučajnih brojeva (CRRN), koja je približna normalnoj raspodeli s obzirom na to da je $n > 13$ (slika 4.27.), i na osnovu unakrsne validacije.



Slika 4.27. Rangiranje modela na osnovu eksperimentalnih $\log k_{0,90}$ vrednosti uzetih za referentni rang (normalizovane SRD vrednosti od 0 do 100 nalaze se na apscisi)

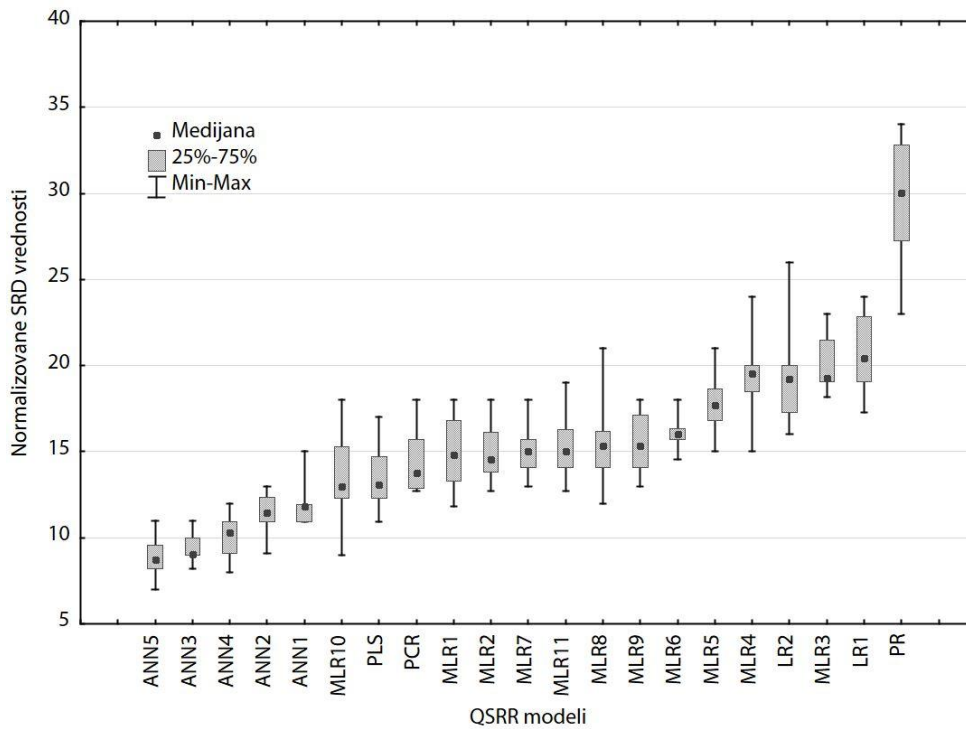
Apsolutne SRD vrednosti i verovatnoća koja opisuje nasumičnost modela dati su u tabeli 4.15. Svi QSRR modeli su opisani veoma malom verovatnoćom, što ukazuje na to da nisu slučajnog karaktera. ANN modeli se ističu sa najmanjom verovatnoćom, te sa ovog aspekta mogu biti preporučeni kao najpouzdaniji za predviđanje $\log k_{0,90}$ parametra analiziranih androstanskih derivata.



Tabela 4.15. Rezultati SRD analize QSRR modela prema eksperimentalnim $\log k_{0,90}$ vrednostima kao referentnom rangu i verovatnoća da su modeli slučajnog (nasumičnog) karaktera ($p\%$)

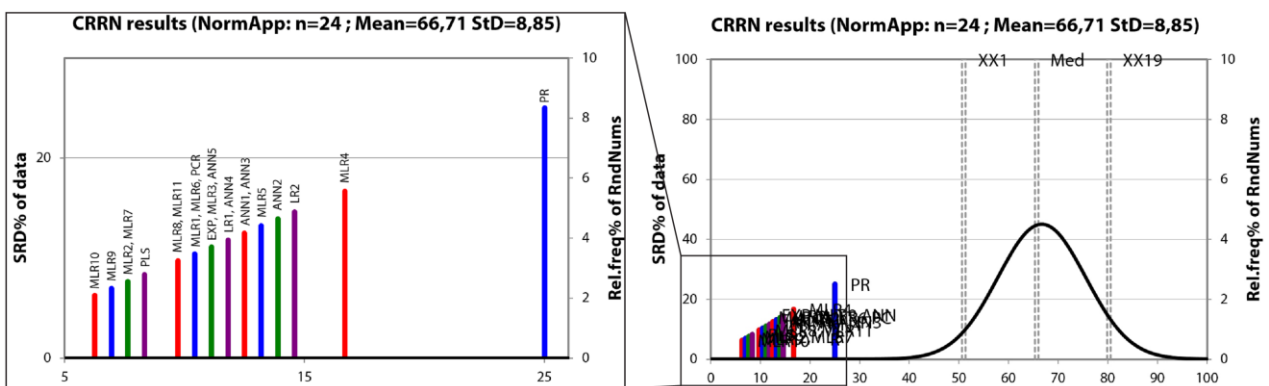
QSRR model	Apsolutna SRD vrednost	Verovatnoća ($p\%$) $x_1 < \text{SRD} \leq x_2$	
ANN5	24	$1,91 \cdot 10^{-9}$	$3,26 \cdot 10^{-9}$
ANN3	26	$3,26 \cdot 10^{-9}$	$5,20 \cdot 10^{-9}$
ANN4	28	$5,20 \cdot 10^{-9}$	$9,05 \cdot 10^{-9}$
ANN2	32	$1,53 \cdot 10^{-8}$	$2,45 \cdot 10^{-8}$
ANN1	34	$2,45 \cdot 10^{-8}$	$4,04 \cdot 10^{-8}$
MLR10	38	$6,78 \cdot 10^{-8}$	$1,08 \cdot 10^{-7}$
PLS	38	$6,78 \cdot 10^{-8}$	$1,08 \cdot 10^{-7}$
PCR	40	$1,08 \cdot 10^{-7}$	$1,71 \cdot 10^{-7}$
MLR1	42	$1,71 \cdot 10^{-7}$	$2,84 \cdot 10^{-7}$
MLR2	42	$1,71 \cdot 10^{-7}$	$2,84 \cdot 10^{-7}$
MLR7	42	$1,71 \cdot 10^{-7}$	$2,84 \cdot 10^{-7}$
MLR8	44	$2,84 \cdot 10^{-7}$	$4,51 \cdot 10^{-7}$
MLR9	44	$2,84 \cdot 10^{-7}$	$4,51 \cdot 10^{-7}$
MLR11	44	$2,84 \cdot 10^{-7}$	$4,51 \cdot 10^{-7}$
MLR6	46	$4,51 \cdot 10^{-7}$	$6,86 \cdot 10^{-7}$
MLR5	50	$1,12 \cdot 10^{-6}$	$1,77 \cdot 10^{-6}$
MLR4	54	$2,66 \cdot 10^{-6}$	$4,21 \cdot 10^{-6}$
LR2	56	$4,21 \cdot 10^{-6}$	$6,57 \cdot 10^{-6}$
MLR3	56	$4,21 \cdot 10^{-6}$	$6,57 \cdot 10^{-6}$
LR1	60	$9,84 \cdot 10^{-6}$	$1,49 \cdot 10^{-5}$
PR	84	$1,03 \cdot 10^{-3}$	$1,44 \cdot 10^{-3}$
XX1	148	4,72	5,50
Q1	174	23,11	25,65
Medijana (Med)	190	48,91	52,04
Q3	208	72,68	75,12
XX19	232	94,72	95,57

Rezultati 7-fold unakrsne validacije prikazani su na slici 4.28. Za svaki model dobijeno je osam SRD vrednosti, uključujući sedam SRD vrednosti dobijenih unakrsnom validacijom i jednu SRD vrednost dobijenu na osnovu celokupnog skupa podataka. Opseg ovih SRD vrednosti prikazan je u formi *Box-Whisker* grafika. Unakrsna validacija SRD analize potvrđuje rangiranje modela koje je prethodno predstavljeno na slici 4.27. i u tabeli 4.15. Unakrsnom validacijom je potvrđeno značajno odstupanje PR modela od ostalih, a samim tim i njegova udaljenost od idealnog ranga.



Slika 4.28. Box-Whisker grafik SRD vrednosti QSRR modela dobijenih pomoću 7-fold unakrsne validacije SRD analize zasnovane na eksperimentalnim $\log_{k_{0,90}}$ vrednostima kao referentnom rang

Ukoliko se SRD analiza izvede na osnovu prosečnih $\log_{k_{0,90}}$ vrednosti (za svako jedinjenje računa se srednja $\log_{k_{0,90}}$ vrednost predviđena na osnovu modela koji se rangiraju), rangiranje modela prikazano na slici 4.29. razlikuje se od rangiranja izvedenog prema eksperimentalnim $\log_{k_{0,90}}$ vrednostima. U ovom slučaju, model MLR10 je najbliži referentnom rang, dok se PR model i u ovom slučaju izdvaja kao najdalji, a samim tim i najlošiji. Ovo je sigurna indikacija da primenu PR modela treba izbegavati.



Slika 4.29. Rangiranje modela na osnovu srednjih predviđenih $\log_{k_{0,90}}$ vrednosti uzetih za referentni rang



Verovatnoća da su QSRR modeli nasumičnog karaktera je i u ovom slučaju veoma mala, što potvrđuju podaci dati u *tabeli 4.16*, koja osim verovatnoće pokazuje i apsolutne SRD vrednosti. Primenom srednjih predviđenih vrednosti kao referentnog ranga, zapravo se pokazuje udaljenost modela od „centra“ definisanog srednjim vrednostima, što predstavlja neparametarsko merilo sličnosti ili razlika između modela. U ovom slučaju može se uočiti da je većina MLR modela locirana u blizini referentnog ranga, uključujući i PCR i PLS modele (modeli prikazani iznad eksperimentalnih (EXP) vrednosti u *tabeli 4.16*). ANN modeli se nalaze na nešto većoj udaljenosti od idealnog ranga, ali u blizini EXP vrednosti.

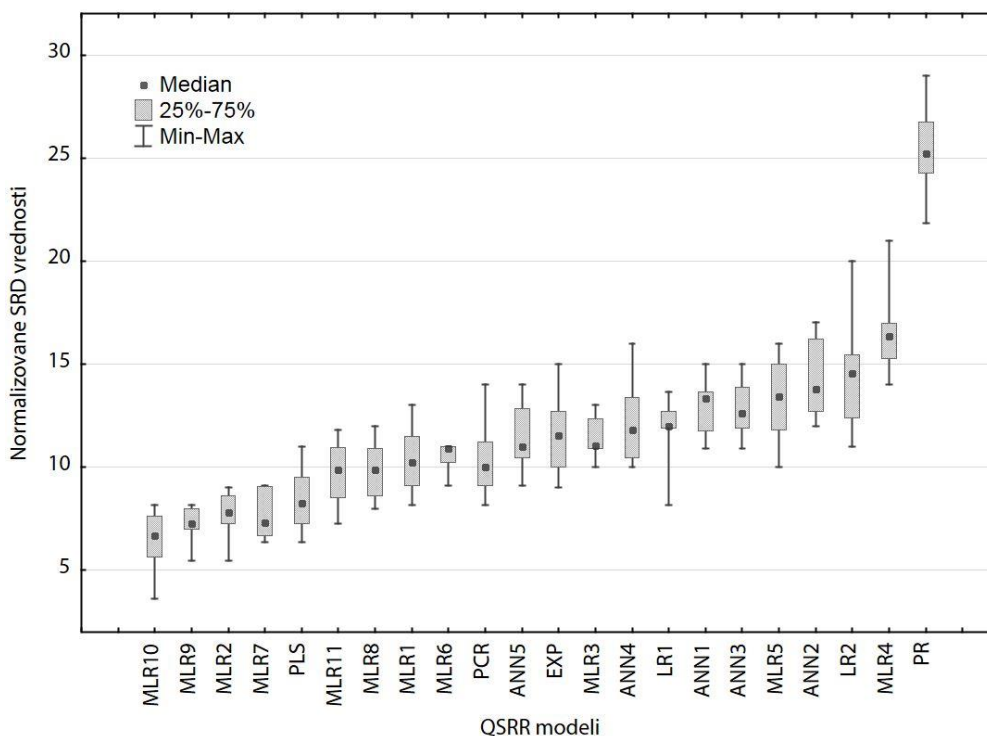
Tabela 4.16. Rezultati SRD analize QSRR modela prema srednjim predviđenim $\log_{k_{0,90}}$ vrednostima kao referentnom rangu i verovatnoća da su modeli slučajnog (nasumičnog) karaktera ($p\%$)

QSRR model	Apsolutna SRD vrednost	Verovatnoća ($p\%$)	
		$x_1 < \text{SRD} \leq x_2$	
MLR10	18	$3,82 \cdot 10^{-10}$	$6,55 \cdot 10^{-10}$
MLR9	20	$6,55 \cdot 10^{-10}$	$1,07 \cdot 10^{-9}$
MLR2	22	$1,07 \cdot 10^{-9}$	$1,91 \cdot 10^{-9}$
MLR7	22	$1,07 \cdot 10^{-9}$	$1,91 \cdot 10^{-9}$
PLS	24	$1,91 \cdot 10^{-9}$	$3,26 \cdot 10^{-9}$
MLR8	28	$5,20 \cdot 10^{-9}$	$9,05 \cdot 10^{-9}$
MLR11	28	$5,20 \cdot 10^{-9}$	$9,05 \cdot 10^{-9}$
MLR1	30	$9,05 \cdot 10^{-9}$	$1,53 \cdot 10^{-8}$
MLR6	30	$9,05 \cdot 10^{-9}$	$1,53 \cdot 10^{-8}$
PCR	30	$9,05 \cdot 10^{-9}$	$1,53 \cdot 10^{-8}$
EXP	32	$1,53 \cdot 10^{-8}$	$2,45 \cdot 10^{-8}$
MLR3	32	$1,53 \cdot 10^{-8}$	$2,45 \cdot 10^{-8}$
ANN5	32	$1,53 \cdot 10^{-8}$	$2,45 \cdot 10^{-8}$
LR1	34	$2,45 \cdot 10^{-8}$	$4,04 \cdot 10^{-8}$
ANN4	34	$2,45 \cdot 10^{-8}$	$4,04 \cdot 10^{-8}$
ANN1	36	$4,04 \cdot 10^{-8}$	$6,78 \cdot 10^{-8}$
ANN3	36	$4,04 \cdot 10^{-8}$	$6,78 \cdot 10^{-8}$
MLR5	38	$6,78 \cdot 10^{-8}$	$1,08 \cdot 10^{-7}$
ANN2	40	$1,08 \cdot 10^{-7}$	$1,71 \cdot 10^{-7}$
LR2	42	$1,71 \cdot 10^{-7}$	$2,84 \cdot 10^{-7}$
MLR4	48	$6,86 \cdot 10^{-7}$	$1,12 \cdot 10^{-6}$
PR	72	$1,13 \cdot 10^{-4}$	$1,60 \cdot 10^{-4}$
XX1	148	4,72	5,50
Q1	174	23,11	25,65
Medijana (Med)	190	48,91	52,04
Q3	208	72,68	75,12
XX19	232	94,72	95,57



Modeli koji su bliži referentnom rangu u odnosu na EXP vrednosti, mogu biti obeleženi kao „profitovani“, što se može dokazati u slučajevima da se SRD analiza izvodi posebno na kalibracionom i eksternom test setu. Različito rangiranje modela prema kalibracionom i test setu bilo bi siguran indikator profitovanja. Međutim, ukoliko modeli imaju bolji rang u odnosu na EXP podatke, to ne mora zasigurno da indicira profitovanost tih modela, već može biti posledica fenomena slučajnog šuma (*random noise*) [151]. U ovom slučaju, testiranje profitovanosti modela na ovaj način nije bilo izvodljivo, s obzirom na to da su kod ANN modela korišćena tri seta podataka, za razliku od linearnih modela. S obzirom na to da su ANN modeli favorizovani kako sa statističkog aspekta, tako i sa aspekta SRD analize prema eksperimentalnim vrednostima, SRD analiza prema prosečnim vrednostima nije dovela u sumnju ispravnost ANN modela. Svi ANN modeli se u tabeli 4.16. nalaze ispod EXP vrednosti, te provera njihove profitovanosti nije neophodna.

Kao i u prethodnom slučaju, validacija SRD analize je izvedena CRRN pristupom i 7-fold unakrsnom validacijom (slika 4.30.). Rezultati unakrsne validacije ukazuju na ispravnost izvedene SRD analize QSRR modela prema srednjim predviđenim $\log k_{0,90}$ vrednostima kao referentnom rangu.



Slika 4.30. Box-Whisker grafikon SRD vrednosti QSRR modela dobijenih pomoću 7-fold unakrsne validacije SRD analize zasnovane na srednjim predviđenim $\log k_{0,90}$ vrednostima kao referentnom rangu

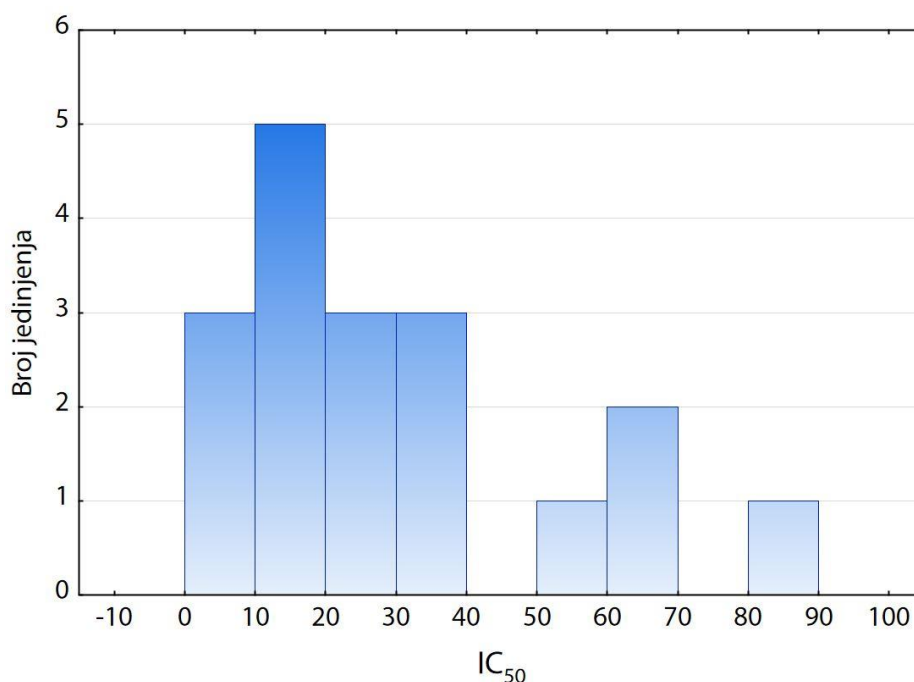
Nakon izvedene statističke i SRD analize QSRR modela, može se izvesti zaključak da modeli formirani primenom veštačkih neuronskih mreža imaju prednost u pogledu boljeg predviđanja $\log k_{0,90}$ vrednosti analiziranih 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata. Model zasnovan na polinomskoj regresiji treba definitivno izbegavati, ukoliko bi cilj njegove primene bio precizno predviđanje hromatografskog parametra.



4.5. QSAR analiza antikancerogene aktivnosti androstanskih derivata

4.5.1. Odabir molekulskih deskriptora kao ulaznih varijabli QSAR modela

QSAR analiza je izvedena na skupu od osamnaest 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata. Cilj ove analize je formiranje modela koji imaju sposobnost predviđanja antiproliferativne (citotoksične) aktivnosti androstanskih derivata prema ćelijama kancera prostate (AR-neg. PC-3). Kako je ranije napomenuto, broj jedinjenja korišćenih u QSAR analizi je manji od broja jedinjenja analiziranih QSRR pristupom, s obzirom na to da je IC_{50} vrednost jedinjenja **2**, **3**, **12**, **13**, **17** i **23** zanemarljiva ($IC_{50} > 100$). Histogram IC_{50} vrednosti jedinjenja koja su analizirana QSAR pristupom prikazan je na slici 4.31.



Slika 4.31. Histogram IC_{50} vrednosti analiziranih 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata

Predstavljeni histogram ukazuje na to da veći broj androstanskih derivata ispoljava zadovoljavajuću, odnosno umerenu antiproliferativnu aktivnost, što upravo opravdava potrebu za primenom QSAR modela koji bi omogućili precizno predviđanje IC_{50} vrednosti strukturno sličnih androstanskih derivata, i samim tim pomogli u potrazi za jedinjenjima sa izrazito jakom antiproliferativnom aktivnošću prema AR-neg. PC-3 ćelijskoj liniji.

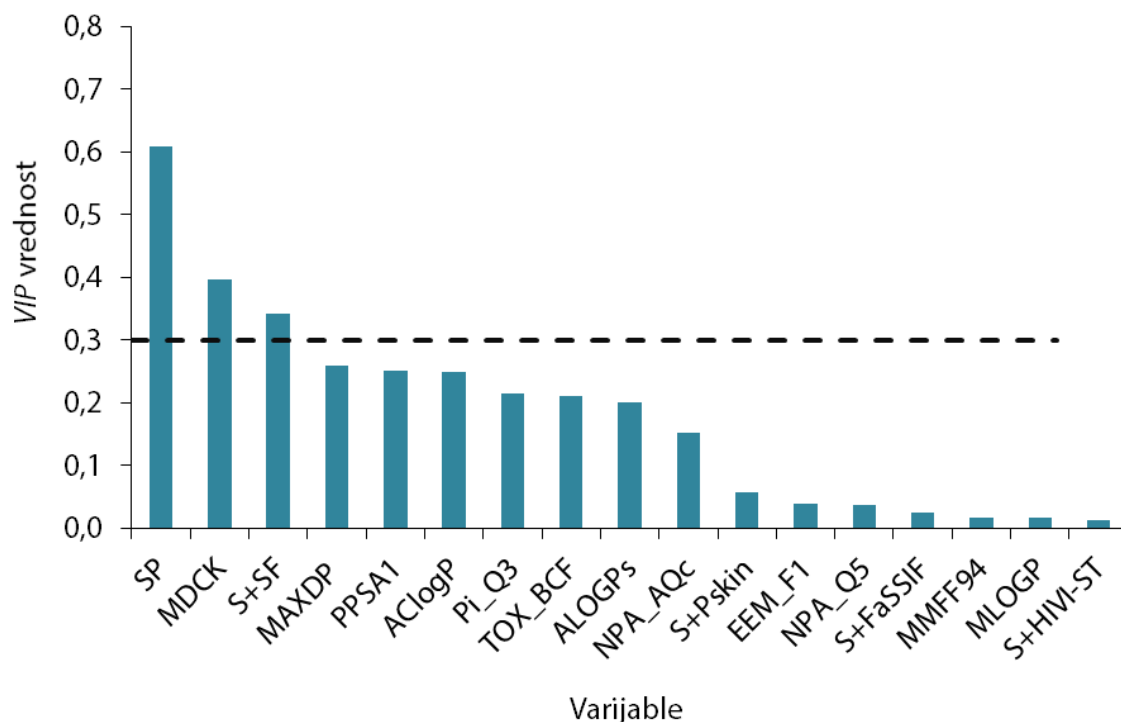


U prvom koraku QSAR analize formiran je set od 250 topoloških, fizičko-hemijskih, elektrostatičkih, ADMET molekulskih deskriptora i deskriptora lipofilnosti. S obzirom na to da nijedna od primenjenih linearnih regresionih tehnika (LR, MLR, PCR, PLS) nije rezultovala matematičkim modelom prihvatljivim za predviđanje IC_{50} vrednosti, nelinearno modelovanje je primenjeno kao moguće rešenje. Međutim, da bi se formirao manji set deskriptora koji bi poslužili kao ulazne varijable za formiranje nelinearnih QSAR modela pomoću veštačkih neuronskih mreža, primenjena je SS metoda odabira varijabli primenom NCSS 2007 programa. Pomoću ove metode, iz seta od polaznih 250 deskriptora izdvojeno je njih sedamnaest. Nivo značajnosti za dodavanje varijable u model iznosio je 0,05, a za uklanjanje varijable iz modela 0,20. Maksimalan broj ponavljanja iznosio je 500. Rezultati SS analize dati su u tabeli 4.17.

Tabela 4.17. Rezultati odabira varijabli pomoću SS metode

Broj ponavljanja	Akcija	Varijabla	R^2	RMSE
1	Dodato	SP	0,1663	23,0233
2	Dodato	MLOGP	0,4903	18,5924
3	Dodato	MDCK	0,6799	15,2501
4	Dodato	PPSA1	0,8228	11,7764
5	Dodato	MAXDP	0,8757	10,2638
6	Dodato	NPA_Q5	0,9056	9,3436
7	Dodato	TOX_BCF	0,9265	8,6451
8	Dodato	S+Pskin	0,9510	7,4415
9	Dodato	S+FaSSIF	0,9769	5,4252
10	Dodato	Pi_Q3	0,9896	3,8883
11	Dodato	NPA_AQc	0,9965	2,4289
12	Dodato	S+SF	0,9990	1,4311
13	Dodato	MMFF94	0,9997	0,8766
14	Dodato	ALOGPs	1,0000	0,1031
15	Dodato	S+HIVI_ST	1,0000	0,0291
16	Dodato	EEM_F1	1,0000	0,0003
17	Dodato	AClogP	1,0000	0,0000
18	Nepromenjeno	-	1,0000	0,0000

Sedamnaest ulaznih varijabli dovelo bi do formiranja kompleksnih ANN modela, koji bi verovatno imali i slabu moć predviđanja. Stoga, primenom PLS-VIP metode selekcije sa NIPALS algoritmom (kriterijum konvergencije je iznosio 0,0001, a maksimalan broj ponavljanja 50) pomoću programa Statistica 10.0, od sedamnaest varijabli izdvojena je podgrupa od tri deskriptora (SP, MDCK i S+SF), čija je VIP vrednost veća od 0,30, što se može uočiti na slici 4.32. Ovi deskriptori su upotrebljeni kao nezavisno promenljive varijable prilikom formiranja QSAR modela zasnovanih na veštačkim neuronskim mrežama.



Slika 4.32. VIP vrednosti deskriptora (odabranih pomoću SS metode) dobijene pomoću PLS-VIP pristupa odabiru varijabli

SP deskriptor predstavlja humanu transdermalnu propustljivost jedinjenja određenu *in vitro*. Izražava se pomoću logaritma konstante K_p [cm/h]. MDCK deskriptor [nm/s] opisuje *in vitro* oralnu apsorpciju jedinjenja pri fiziološkoj vrednosti pH. S+SF deskriptor definiše odnos ravnotežne rastvorljivosti jonizabilne soli ispitivanog jedinjenja i njene intrinzične rastvorljivosti. Ovi deskriptori su u bliskoj vezi sa lipofilnošću supstance i oslikavaju njenu mogućnost dostizanja mesta delovanja u organizmu.

4.5.2. Nelinearno QSAR modelovanje

QSAR modelovanje je izvedeno primenom veštačkih neuronskih mreža, s obzirom na neuspešne pokušaje definisanja linearnih QSAR modela. Formiranje ANN modela je izvedeno primenom programa Statistica 10 pomoću *Automated Network Search (ANS)* opcije. Višeslojne (MLP) nerekurentne (*feedforward*) mreže trenirane su primenom BFGS algoritma kombinacijom sinusne (*Sine*), logističke (*Lgt*), hiperbolične tangens (*Tanh*), eksponencijalne (*Exp*) funkcije i funkcije identiteta (*Idt*), kao aktivacionih funkcija skrivenog i izlaznog sloja. Neuronske težine su određene pomoću primenjenog softvera.

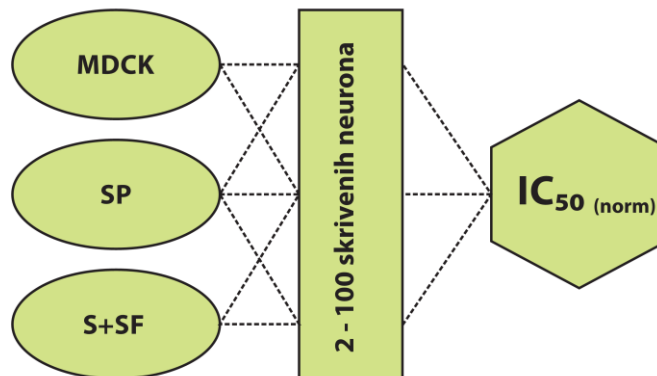


Trenirano je ukupno 10 000 mreža, od kojih su najbolji modeli odabrani na osnovu maksimalnih R_{kalibr} , R_{test} , R_{valid} i minimalnih $RMSE_{kalibr}$, $RMSE_{test}$, $RMSE_{valid}$ vrednosti kalibracionog, test i validacionog seta, uključujući i $RMSE$ i R vrednosti zavisnosti eksperimentalno-predviđeno. U tabeli 4.18. prikazani su podaci na osnovu kojih je obavljeno treniranje ANN modela.

Tabela 4.18. Podaci upotrebljeni za formiranje ANN-QSAR modela za predviđanje citotoksične aktivnosti androstanskih derivata prema ćelijama kancera prostate

Jedinjenje	Set	MDCK [nm/s]	SP (log K_p)	S+SF	IC ₅₀ (eksperimentalno) [μM] [41-43]	IC ₅₀ (normalizovano)
1	Kalibracioni	33,4997	-1,9734	1070,0	6,30	0,0738
4	Kalibracioni	7,6154	-3,1941	1250,0	65,50	0,7304
5	Validacioni	1,7836	-4,0654	94,8	33,60	0,3766
6	Kalibracioni	34,1743	-1,9359	1070,0	22,30	0,2513
7	Validacioni	0,0963	-1,5486	1910,0	31,30	0,3511
8	Kalibracioni	11,4686	-2,5335	1110,0	4,30	0,0516
9	Test	28,1355	-2,5557	1030,0	25,60	0,2879
10	Test	46,7956	-1,1013	2010,0	24,70	0,2779
11	Kalibracioni	28,7219	-2,9417	506,0	52,10	0,5818
14	Kalibracioni	65,8760	-1,7981	2590,0	12,90	0,1470
15	Kalibracioni	53,9390	-2,5124	2550,0	14,30	0,1625
16	Kalibracioni	3,0115	-1,8666	303,0	10,10	0,1159
18	Test	0,0605	-1,8442	356,0	0,55	0,0100
19	Kalibracioni	30,4541	-1,7017	3300,0	34,10	0,3821
20	Validacioni	65,8410	-1,6423	1960,0	10,10	0,1159
21	Kalibracioni	2,6234	-1,8892	3700,0	66,20	0,7382
22	Kalibracioni	25,2490	-2,8980	2080,0	88,90	0,9900
24	Kalibracioni	69,5311	-1,0540	3440,0	19,90	0,2246

Od 10 000 treniranih mreža, na osnovu statističkih parametara izdvojeno je sedamnaest ANN modela kao najboljih. Njihova opšta arhitektura prikazana je na slici 4.33. Neophodno je istaći da su najbolji ANN-QSAR modeli dobijeni primenom prirodnih ulaznih varijabli i normalizovane izlazne varijable (IC₅₀ vrednosti), za razliku od QSRR modela koji su formirani primenom normalizovanih ulaznih varijabli i prirodne izlazne varijable.



Slika 4.33. Opšta arhitektura formiranih ANN modela

Statistički parametri odabranih mreža dati su u *tabeli 4.19*. Na osnovu predstavljenih statističkih parametara može se uočiti da su modeli sličnog kvaliteta, odnosno da nema drastične razlike između njih. Međutim, mreže MLP 3-4-1, MLP 3-15-1 i MLP 3-71-1 mogu se izdvojiti od ostalih kao modeli sa najboljim statističkim karakteristikama.

Procena prediktivne sposobnosti ova tri modela izvedena je poređenjem eksperimentalnih i predviđenih IC_{50} vrednosti (*slika 4.34*) i analizom ostataka (*slika 4.35*). Poređenjem eksperimentalnih i predviđenih IC_{50} vrednosti pokazano je veoma malo rasipanje tačaka oko linearne zavisnosti kod sva tri modela. U prilog kvalitetu ova tri modela govore i koeficijenti korelacije i nagibi zavisnosti eksperimentalno-predviđeno, koji su veoma bliski jedinici, kao i odsečki veoma bliski nuli. Analiza rasporeda ostataka oko $y = 0$ ose ukazuje na njihovu nasumičnost, koja je potvrđena Vald-Volfovici testom (*tabela 4.20*).

Ukoliko se uporede R i $RMSE$ parametri svih sedamnaest ANN modela, koji su računati na osnovu predviđenih i eksperimentalnih vrednosti uzimajući u obzir sva tri seta podataka, na osnovu *slike 4.36.(1)* uočava se da se upravo modeli MLP 3-4-1, MLP 3-15-1 i MLP 3-71-1 izdvajaju kao najbolji na osnovu najviših R i najnižih $RMSE$ vrednosti. Ovo je još jedna potvrda njihove značajne sposobnosti predviđanja antiproliferativne aktivnosti 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata. Ova tri modela su okarakterisana i najmanjom amplitudom ostataka (*slika 4.36.(2)*). Vrednosti razlika između eksperimentalnih i predviđenih vrednosti za ova tri modela nalaze se u opsegu od -8,5 do 8,0.


Tabela 4.19. Statističke karakteristike formiranih ANN-QSAR modela

ANN model (arhitektura)	Skrivena aktivaciona funkcija	Izlazna aktivaciona funkcija	R_{kalib}	R_{test}	R_{valid}	$RMSE_{kalib}$	$RMSE_{test}$	$RMSE_{valid}$	F-vrednost (eksperimentalno-predviđeno)	p-vrednost (eksperimentalno-predviđeno)	Koef. varijacije (CV)
1. MLP 3-16-1	Tanh	Sine	0,9654	0,8387	0,9398	0,0033	0,0027	0,0035	164,3	0,0000	0,25
2. MLP 3-20-1	Tanh	Sine	0,9425	0,9918	0,9837	0,0051	0,0005	0,0065	120,3	0,0000	0,32
3. MLP 3-14-1	Tanh	Exp	0,9767	0,9708	0,8554	0,0023	0,0016	0,0029	236,9	0,0000	0,21
4. MLP 3-29-1	Tanh	Sine	0,9807	0,8347	0,9600	0,0017	0,0030	0,0011	288,2	0,0000	0,19
5. MLP 3-8-1	Tanh	Sine	0,9777	0,9976	0,9937	0,0020	0,0001	0,0009	360,0	0,0000	0,17
6. MLP 3-10-1	Lgt	Idt	0,9783	0,9999	0,9808	0,0019	0,0001	0,0008	372,5	0,0000	0,17
7. MLP 3-18-1	Lgt	Sine	0,9912	0,9884	0,9786	0,0008	0,0005	0,0075	301,2	0,0000	0,20
8. MLP 3-15-1	Tanh	Idt	0,9932	0,9833	0,9605	0,0006	0,0005	0,0015	762,1	0,0000	0,12
9. MLP 3-27-1	Tanh	Sine	0,9792	1,0000	0,9924	0,0019	0,0011	0,0002	369,4	0,0000	0,17
10. MLP 3-14-1	Tanh	Sine	0,9912	0,9609	0,9948	0,0008	0,0010	0,0022	543,8	0,0000	0,14
11. MLP 3-7-1*	Lgt	Sine	0,9844	0,9843	0,9946	0,0016	0,0003	0,0018	454,5	0,0000	0,15
12. MLP 3-17-1	Lgt	Tanh	0,9925	0,9632	0,9795	0,0007	0,0009	0,0020	588,5	0,0000	0,14
13. MLP 3-24-1	Exp	Idt	0,9812	0,9998	1,0000	0,0017	0,0002	0,0025	353,9	0,0000	0,18
14. MLP 3-71-1	Exp	Sine	0,9909	0,9920	0,9912	0,0008	0,0003	0,0010	753,3	0,0000	0,12
15. MLP 3-4-1	Lgt	Idt	0,9927	0,9937	0,9945	0,0007	0,0002	0,0012	837,4	0,0000	0,11
16. MLP 3-13-1	Lgt	Sine	0,9837	0,9948	0,9935	0,0015	0,0001	0,0045	333,5	0,0000	0,19
17. MLP 3-7-1**	Lgt	Sine	0,9842	1,0000	0,9991	0,0014	0,0000	0,0009	503,1	0,0000	0,15

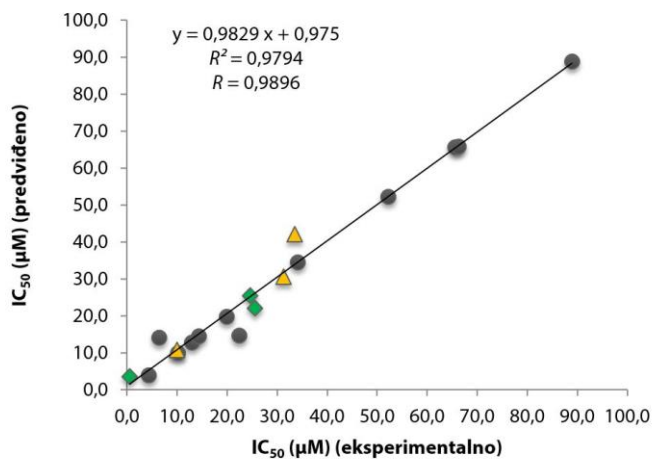
*** Obeležene mreže se međusobno razlikuju po broj ciklusa treniranja nakon kojih je dobijena mreža optimalne arhitekture. Mreža 11 je dobijena nakon 37, a mreža 17 nakon 29 ciklusa treniranja.

Tabela 4.20. Rezultati Vald-Volfovica testa nasumičnosti ostataka

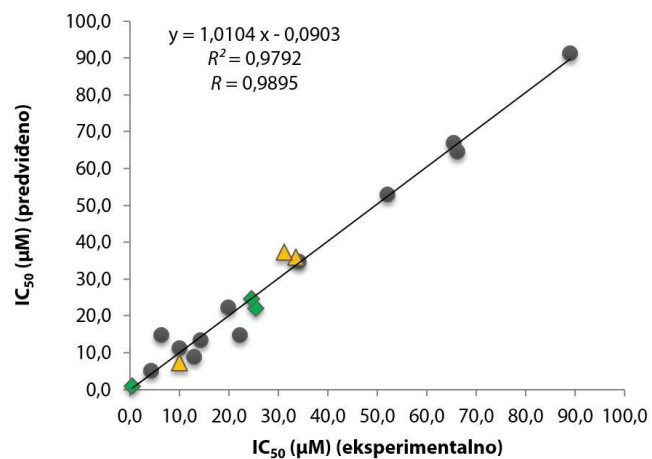
H ₀ : Nasumična raspodela ostataka ($\alpha = 0,05$)			
ANN model	r	r_{kr}	Rezultati testiranja H ₀
3-15-1	8	5,7	H ₀ se prihvata ($r > r_{kr}$)
3-71-1	11	5,9	H ₀ se prihvata ($r > r_{kr}$)
3-4-1	12	5,9	H ₀ se prihvata ($r > r_{kr}$)



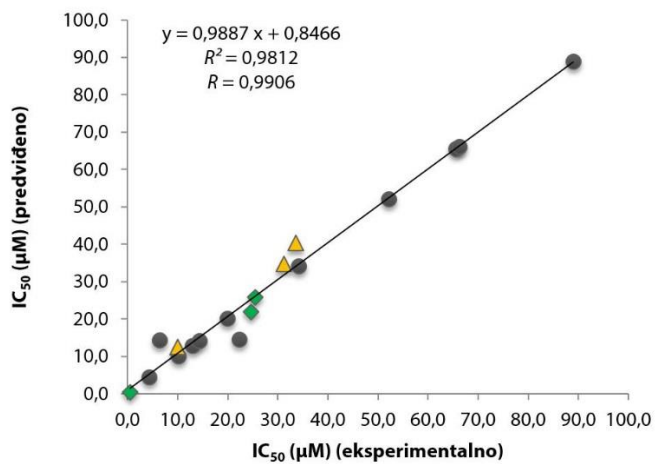
MLP 3-15-1



MLP 3-71-1



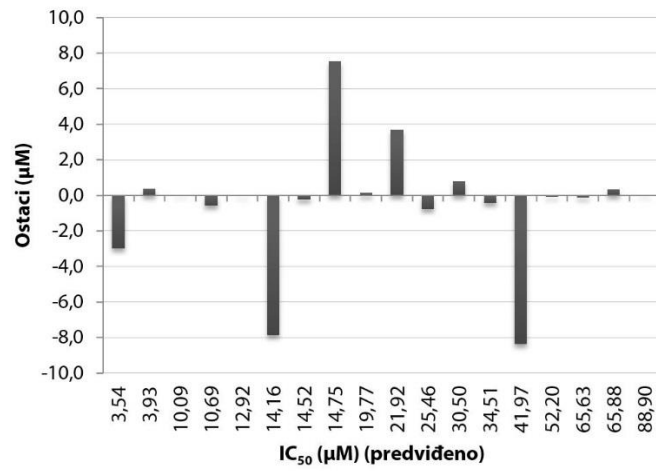
MLP 3-4-1



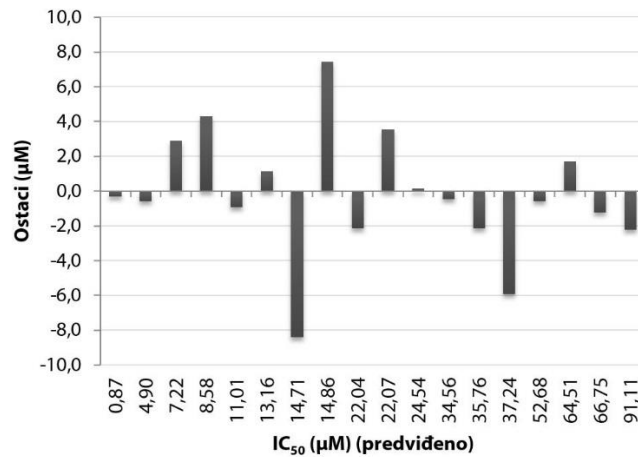
Slika 4.34. Poređenje eksperimentalnih i predviđenih IC₅₀ vrednosti (● - kalibracioni set, ◆ - test set, ▲ - validacioni set)



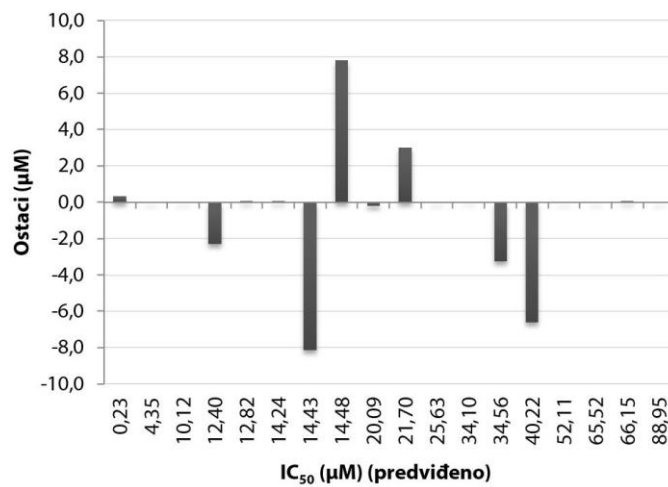
MLP 3-15-1



MLP 3-71-1



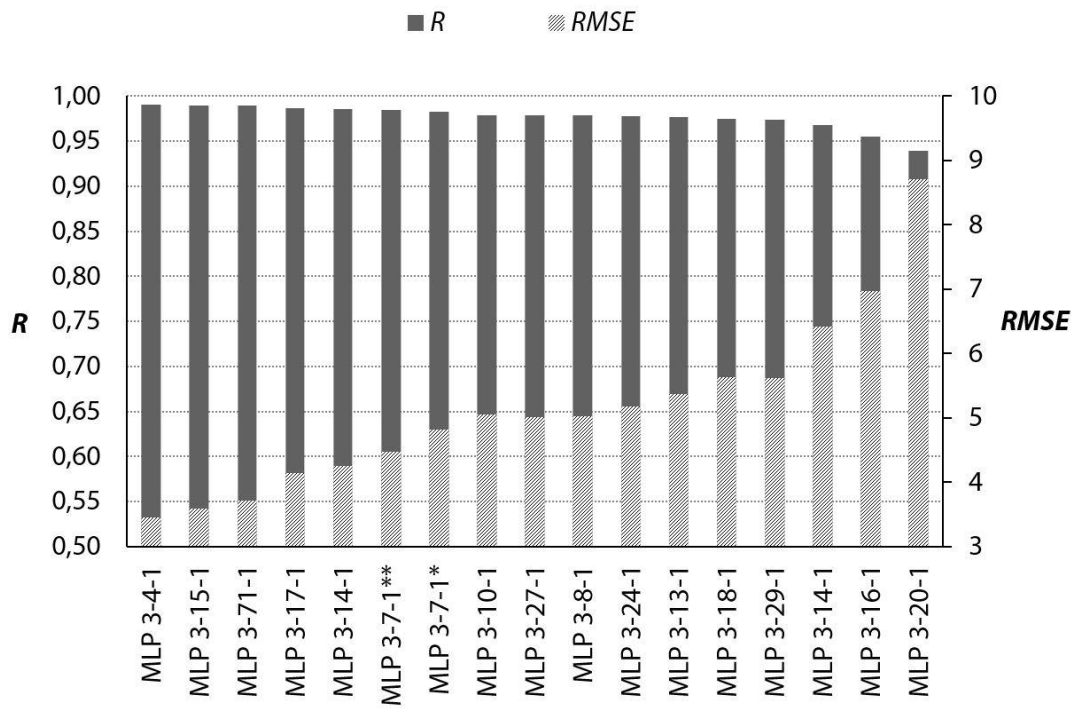
MLP 3-4-1



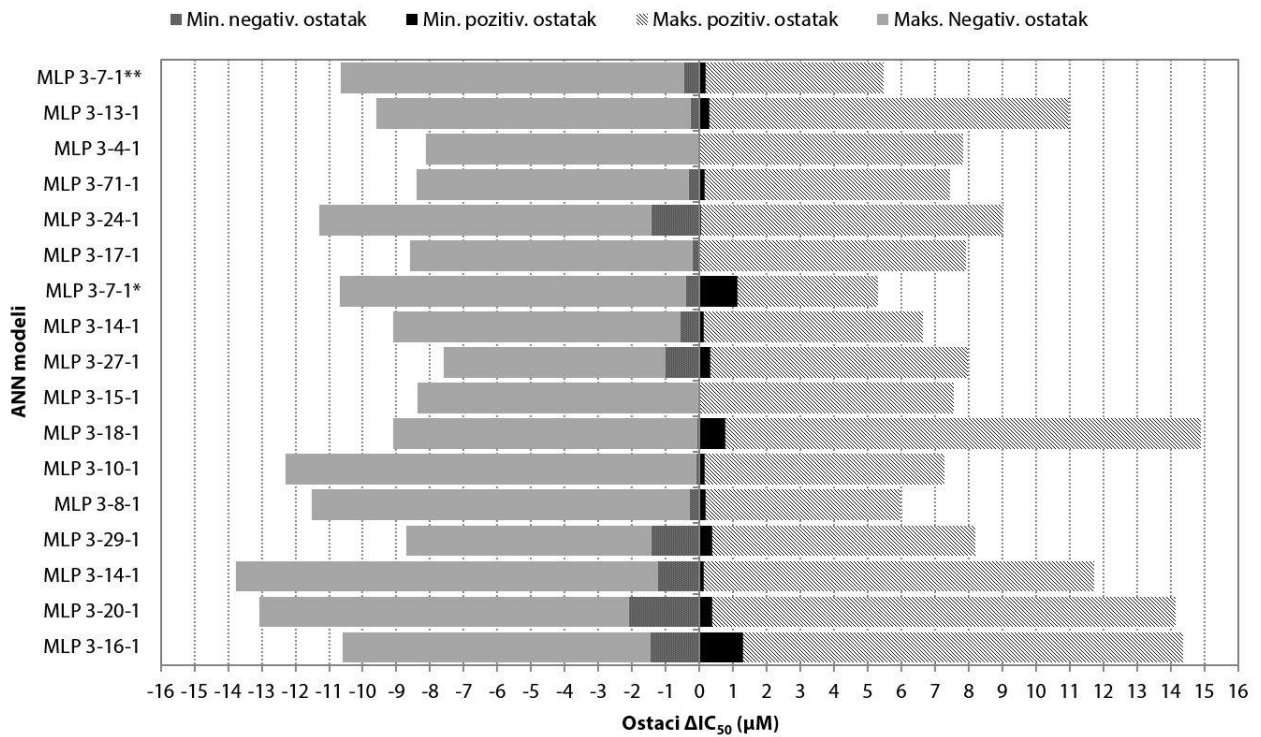
Slika 4.35. Poređenje predviđenih IC₅₀ vrednosti i ostataka



1



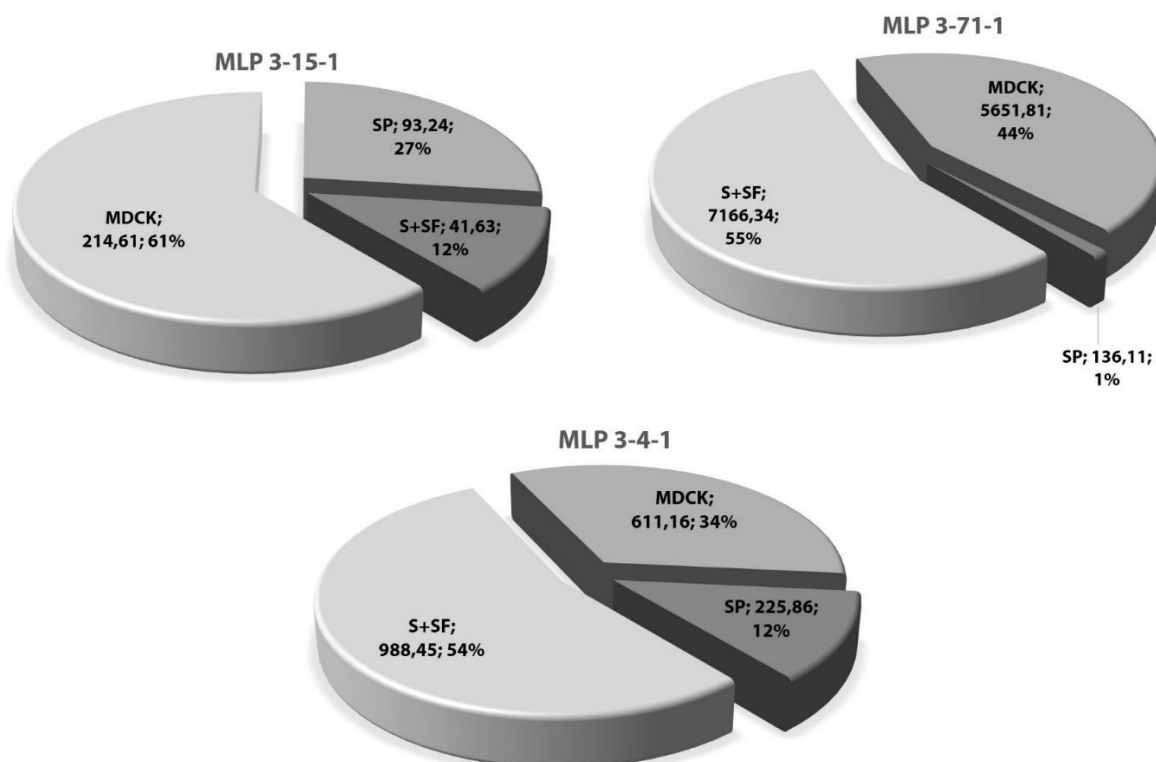
2



Slika 4.36. Poređenje *RMSE* i *R* vrednosti (1) i analiza maksimalnih i minimalnih ostataka (2) za ANN-QSAR modele



Pomoću analize globalne osetljivosti procenjen je uticaj svake ulazne varijable na varijacije vrednosti parametara dobijenih ANN-QSAR modela, odnosno na vrednosti izlazne varijable (IC_{50}). GSA koeficijenti MDCK, SP i S+SF varijabli prikazani su na *slici 4.37*. Visoke vrednosti GSA koeficijenta, značajno veće od 1, ukazuju na velik uticaj sve tri varijable na vrednosti izlazne varijable. MDCK i S+SF varijable se posebno ističu po najvećim vrednostima GSA koeficijenta.



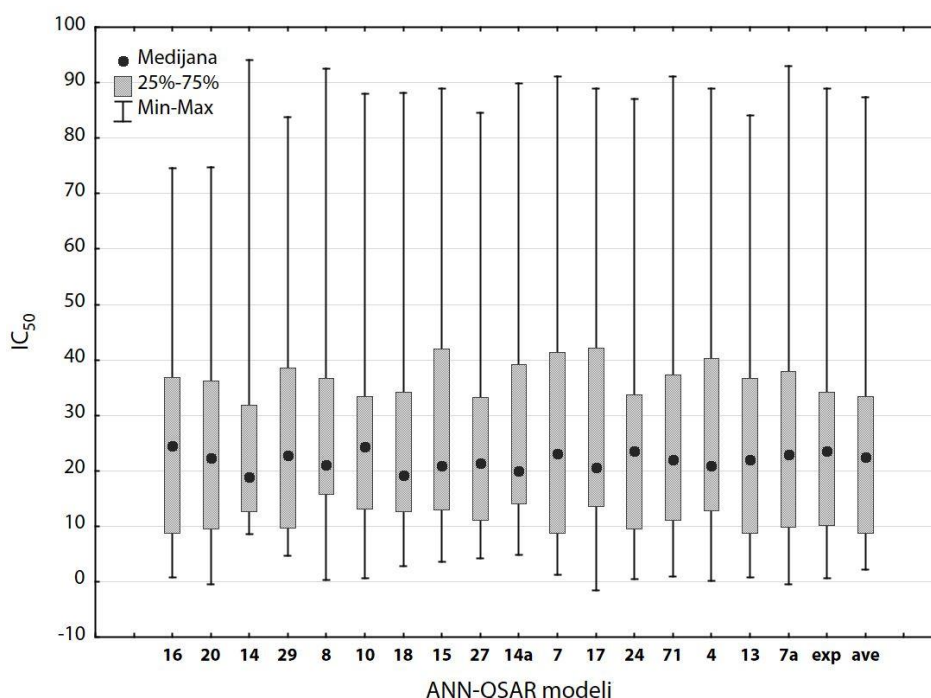
Slika 4.37. GSA koeficijenti ulaznih varijabli

Na osnovu sprovedene statističke analize formiranih ANN modela može se zaključiti da je postavljen cilj QSAR analize ostvaren. Nelinearnim modelovanjem pomoću veštačkih neuronskih mreža dobijeni su ANN modeli sa značajnom sposobnosti predviđanja antiproliferativne aktivnosti analiziranih 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata prema ćelijama kancera prostate pomoću odabranih ADMET deskriptora. Primenom predstavljenih ANN modela (MLP 3-4-1, MLP 3-15-1 i MLP 3-71-1) potraga za novim 17α -pikolil i $17(E)$ -pikoliniliden androstanskim derivatima sa jakom antiproliferativnom aktivnošću prema AR-neg. PC-3 ćelijskoj liniji može biti olakšana.



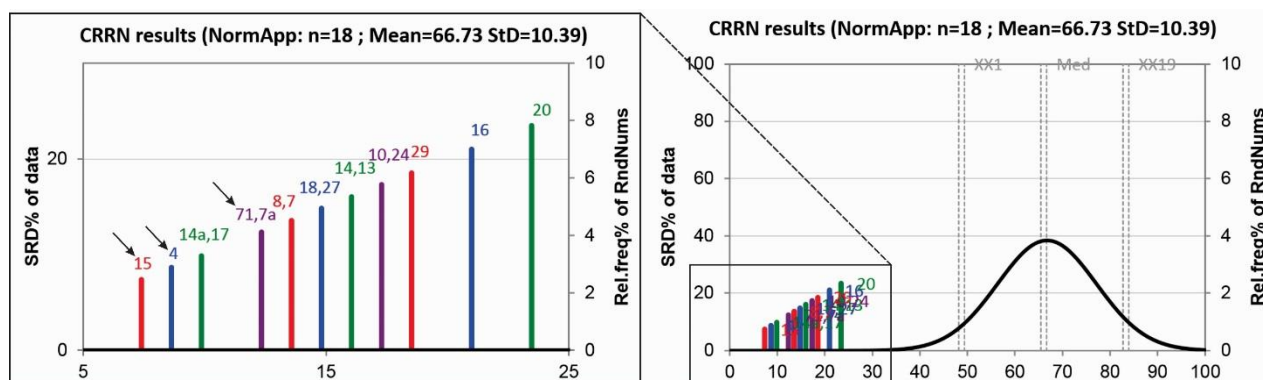
4.5.3. Rangiranje QSAR modela

Pored odabira kvalitetnih ANN-QSAR modela klasičnim poređenjem statističkih parametara, izvedeno je njihovo rangiranje primenom SRD metode. *Box-Whisker* grafik (*slika 4.38.*) ne otkriva sličnosti i razlike između modela, niti ističe bilo koji model po njegovim dobrim ili lošim karakteristikama, kako u odnosu na eksperimentalne (exp), tako i u odnosu na prosečne (ave) IC_{50} vrednosti. Da bi se utvrdile potencijalne razlike između ANN modela, SRD analiza je izvedena na osnovu eksperimentalnih i srednjih predviđenih IC_{50} vrednosti uzetih za referentni (idealni) rang. ANN-QSAR modeli su u rezultatima SRD analize obeleženi brojem neurona zastupljenih u skrivenom sloju.



Slika 4.38. *Box-Whisker* grafik eksperimentalnih IC_{50} vrednosti (exp), IC_{50} vrednosti predviđenih pomoću formiranih ANN-QSAR modela i prosečnih IC_{50} vrednosti (ave)

Rangiranjem ANN-QSAR modela u odnosu na eksperimentalne IC_{50} vrednosti, koje su uzete za referentni rang (*slika 4.39.*), može se uočiti da su modeli 15 i 4, koji su i bili preporučeni kao najbolji na osnovu statističkih parametara, zapravo najbliži referentnom rangu, što govori u prilog njihovom dobrom fitovanju podataka. Model 71 se u odnosu na referentni rang nalazi na nešto većoj, ali ipak prihvatljivoj, udaljenosti u poređenju sa modelima 15 i 4. Model 20 se nalazi na najvećoj udaljenosti od referentnog ranga, te se njegova primena treba izbegavati. Kako u slučaju rangiranja QSRR modela, tako i u ovom slučaju rangiranje u odnosu na eksperimentalne vrednosti nije izvedeno sa ciljem opisivanja greške predviđanja, nego procene fitovanja podataka. Validacija SRD metode je izvedena na osnovu teorijske raspodele slučajnih brojeva (CRRN), koja je približna normalnoj raspodeli s obzirom na to da je $n > 13$ (*slika 4.39.*), i na osnovu unakrsne validacije. Apsolutne SRD vrednosti i verovatnoća koja opisuje nasumičnost ANN-QSAR modela prikazani su u *tabeli 4.21.* Svi ANN-QSAR modeli su okarakterisani veoma malom verovatnoćom, te nisu slučajnog karaktera.



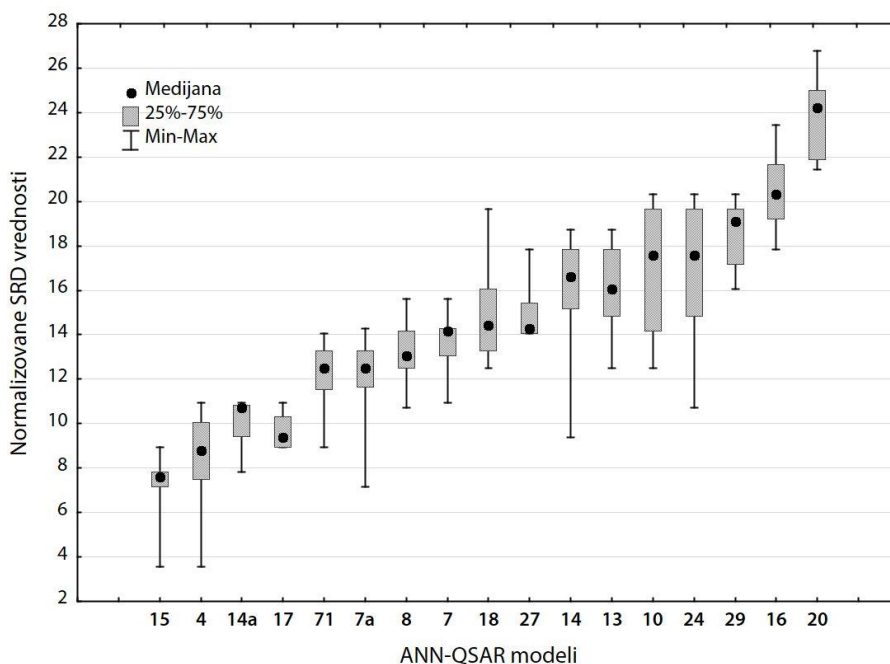
Slika 4.39. Rangiranje ANN-QSAR modela na osnovu eksperimentalnih IC₅₀ vrednosti uzetih za referentni rang (normalizovane SRD vrednosti od 0 do 100 nalaze se na apscisi)

Tabela 4.21. Rezultati SRD analize ANN-QSAR modela prema eksperimentalnim IC₅₀ vrednostima kao referentnom rang i verovatnoća da su modeli slučajnog (nasumičnog) karaktera (*p*%)

ANN model	Apsolutna SRD vrednost	Verovatnoća (<i>p</i> %)	
		$x_1 < SRD \leq x_2$	
15	12	$3,75 \cdot 10^{-7}$	$7,69 \cdot 10^{-7}$
4	14	$7,69 \cdot 10^{-7}$	$1,52 \cdot 10^{-6}$
14a	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
17	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
71	20	$5,68 \cdot 10^{-6}$	$1,10 \cdot 10^{-5}$
7a	20	$5,68 \cdot 10^{-6}$	$1,10 \cdot 10^{-5}$
8	22	$1,10 \cdot 10^{-5}$	$2,06 \cdot 10^{-5}$
7	22	$1,10 \cdot 10^{-5}$	$2,06 \cdot 10^{-5}$
18	24	$2,06 \cdot 10^{-5}$	$3,78 \cdot 10^{-5}$
27	24	$2,06 \cdot 10^{-5}$	$3,78 \cdot 10^{-5}$
14	26	$3,78 \cdot 10^{-5}$	$6,84 \cdot 10^{-5}$
13	26	$3,78 \cdot 10^{-5}$	$6,84 \cdot 10^{-5}$
10	28	$6,84 \cdot 10^{-5}$	$1,25 \cdot 10^{-4}$
24	28	$6,84 \cdot 10^{-5}$	$1,25 \cdot 10^{-4}$
29	30	$1,25 \cdot 10^{-4}$	$2,24 \cdot 10^{-4}$
16	34	$3,90 \cdot 10^{-4}$	$6,64 \cdot 10^{-4}$
20	38	$1,15 \cdot 10^{-3}$	$1,95 \cdot 10^{-3}$
XX1	80	4,86	5,91
Q1	96	24,31	27,42
Medijana (Med)	108	49,10	52,94
Q3	120	74,26	77,26
XX19	136	94,67	95,64

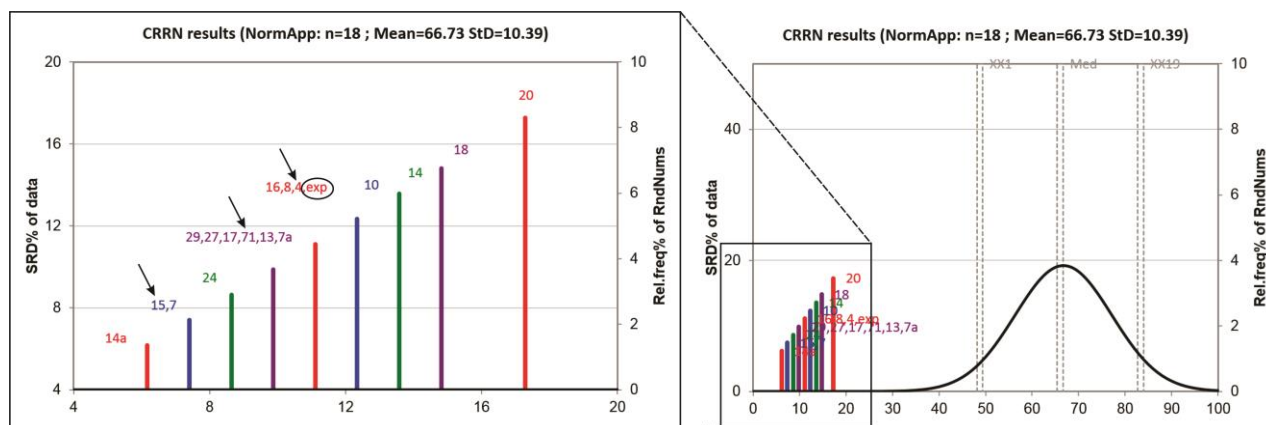


Rezultati *7-fold* unakrsne validacije prikazani su na *slici 4.40.* u formi *Box-Whisker* grafika. Za svaki ANN-QSAR model izračunato je osam SRD vrednosti, uključujući sedam SRD vrednosti dobijenih unakrsnom validacijom i jednu SRD vrednost dobijenu na osnovu celokupnog skupa podataka. Rezultati unakrsne validacije su u skladu sa rezultatima rangiranja modela predstavljenih na *slici 4.39.* i u *tabeli 4.21.* Unakrsnom validacijom je potvrđena blizina modela 15 i 4 idealnom rangu, kao i najveće odstupanje modela 20 od idealnog ranga (najviše SRD vrednosti), te se njegova primena treba definitivno izbegavati.



Slika 4.40. *Box-Whisker* grafik SRD vrednosti ANN-QSAR modela dobijenih pomoću *7-fold* unakrsne validacije SRD analize zasnovane na eksperimentalnim IC_{50} vrednostima kao referentnom rangu

Rangiranje modela u odnosu na prosečne IC_{50} vrednosti uzete za referentni rang prikazano je na *slici 4.41.* Model 14 je najbliži referentnom rangu, dok je i sa ovog aspekta model 20 najdalji, što zapravo predstavlja potvrdu prethodno navedene preporuke da se primena ovog modela treba izbegavati. Ukoliko se fokusiramo na rang modela 4, 15 i 71, koji su preporučeni na osnovu kvaliteta statističkih parametara, može se uočiti da model 4 ima isti rang kao eksperimentalne vrednosti (exp) i od ova tri modela nalazi se na najvećoj udaljenosti od referentnog ranga, dok je model 15 najbliži referentnom rangu. Ukoliko uvažimo pretpostavku da modeli koji su bliži referentnom rangu u odnosu na eksperimentalne vrednosti mogu biti obeleženi kao „profitovani“, model 4 se može smatrati generalno ispravnim, s obzirom na isti rang sa eksperimentalnim vrednostima. Verovatnoća da su ANN-QSAR modeli nasumičnog karaktera je i u ovom slučaju veoma mala, na šta ukazuju podaci dati u *tabeli 4.22.*



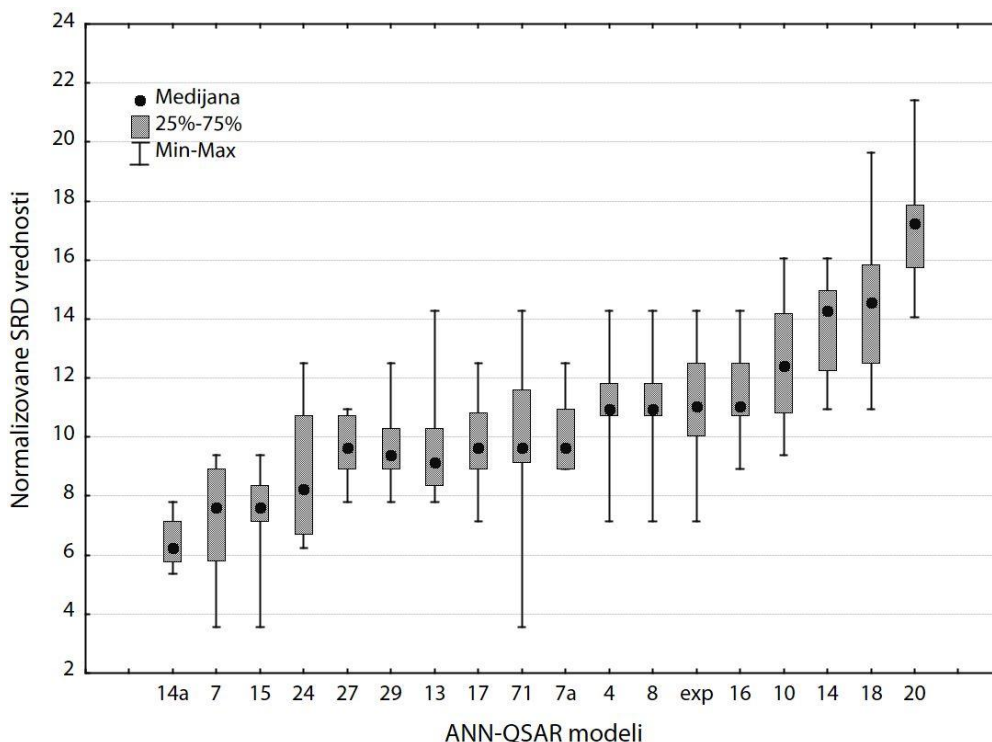
Slika 4.41. Rangiranje ANN-QSAR modela na osnovu srednjih predviđenih IC_{50} vrednosti uzetih za referentni rang

Tabela 4.22. Rezultati SRD analize ANN-QSAR modela prema srednjim predviđenim IC_{50} vrednostima kao referentnom rangu i verovatnoća da su modeli slučajnog (nasumičnog) karaktera ($p\%$)

ANN model	Apsolutna SRD vrednost	Verovatnoća ($p\%$) $x_1 < SRD \leq x_2$	
14a	10	$1,78 \cdot 10^{-7}$	$3,75 \cdot 10^{-7}$
15	12	$3,75 \cdot 10^{-7}$	$7,69 \cdot 10^{-7}$
7	12	$3,75 \cdot 10^{-7}$	$7,69 \cdot 10^{-7}$
24	14	$7,69 \cdot 10^{-7}$	$1,52 \cdot 10^{-6}$
29	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
27	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
17	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
71	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
13	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
7a	16	$1,52 \cdot 10^{-6}$	$2,93 \cdot 10^{-6}$
exp	18	$2,93 \cdot 10^{-6}$	$5,68 \cdot 10^{-6}$
16	18	$2,93 \cdot 10^{-6}$	$5,68 \cdot 10^{-6}$
8	18	$2,93 \cdot 10^{-6}$	$5,68 \cdot 10^{-6}$
4	18	$2,93 \cdot 10^{-6}$	$5,68 \cdot 10^{-6}$
10	20	$5,68 \cdot 10^{-6}$	$1,10 \cdot 10^{-6}$
14	22	$1,10 \cdot 10^{-6}$	$2,06 \cdot 10^{-5}$
18	24	$2,06 \cdot 10^{-5}$	$3,78 \cdot 10^{-5}$
20	28	$6,84 \cdot 10^{-5}$	$1,25 \cdot 10^{-4}$
XX1	80	4,86	5,91
Q1	96	24,31	27,42
Medijana (Med)	108	49,10	52,94
Q3	120	74,26	77,26
XX19	136	94,67	95,64



Box-Whisker grafik (slika 4.42.) predstavlja rezultate 7-fold unakrsne validacije SRD analize i ukazuje na ispravnost rangiranja ANN-QSAR modela primenom SRD analize prema srednjim predviđenim IC_{50} vrednostima kao referentnom rang.



Slika 4.42. Box-Whisker grafik SRD vrednosti ANN-QSAR modela dobijenih pomoću 7-fold unakrsne validacije SRD analize zasnovane na srednjim predviđenim IC_{50} vrednostima kao referentnom rang

Razmatranjem statističkih parametara i rezultata SRD analize ANN-QSAR modela, model 4 se može preporučiti kao optimalno rešenje za predviđanje antiproliferativne aktivnosti 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata prema ćelijskoj liniji kancera prostate (AR-neg. PC-3) izražene pomoću IC_{50} vrednosti. Odabir ovog modela kao optimalnog zapravo predstavlja kompromis između različitih kriterijuma odabira modela. Ovaj model pravi najmanju grešku predviđanja, ima najoptimalnije statističke parametre, a na osnovu položaja u odnosu na eksperimentalne vrednosti kao referentni rang može se zaključiti i da se primenom ovog modela ostvaruje veoma dobro fitovanje podataka.





5. Zaključak

Rezultati hromatografske analize, izvedene pomoću RP-HPLC sistema, i hemometrijske analize definisanog retencionog ponašanja (hromatografske lipofilnosti) 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata, predstavljeni u ovoj doktorskoj disertaciji, navode na sledeće zaključke:

- Ispitivani derivati androstana imaju veću retenciju u sistemu sa mobilnom fazom koja sadrži manji udeo metanola ($\varphi = 0,70$). Primenom ove mobilne faze može se uočiti značajna razlika u retencionom ponašanju grupe 17 α -pikolil androstanskih derivata od grupe 17(E)-pikoliniliden androstanskih derivata, što je pokazano Vald-Volfovici testom. Prema tome, kapacitetni faktor (k) se u tom slučaju može smatrati diskriminacionim faktorom ove dve grupe jedinjenja.
- Analizom uticaja supstituenata na retenciju uočava se da zamena 17 α -pikolil i 17 β -hidroksi funkcionalnih grupa 17(E)-pikoliniliden grupom dovodi do značajnog povećanja retencije jedinjenja u primenjenom hromatografskom sistemu. Veću retenciju, a samim tim i veću hromatografsku lipofilnost, ima grupa 17(E)-pikoliniliden androstanskih derivata.
- Računarskim modelovanjem hidrofilno-lipofilnih (HILI) površina, može se uočiti da je kod većine jedinjenja lipofilni karakter zastupljeniji u odnosu na hidrofilni, što govori u prilog umerenoj polarnosti ovih jedinjenja.
- Na osnovu izračunatih deskriptora lipofilnosti ($\log P$, $\log D$, $\log S$) može se zaključiti da su analizirani derivati umereno do jako lipofilni. Ustanovljena je jaka korelacija između deskriptora lipofilnosti i retencionog parametra $\log k$, što predstavlja potvrdu da se dati retencioni parametar može smatrati alternativnim parametrom lipofilnosti obe grupe ispitivanih androstanskih derivata.
- Klaster analizom i analizom glavnih komponenta dobijeni su rezultati koji ukazuju na to da se grupa 17 α -pikolil androstanskih derivata razlikuje u određenom stepenu od grupe 17(E)-pikoliniliden androstanskih derivata, pogotovo prema deskriptorima lipofilnosti. Međutim, ova razlika nije drastična, te su ove dve grupe jedinjenja posmatrane kao celina prilikom formiranja QSRR i QSAR modela.
- Linearno i nelinearno QSRR modelovanje rezultovalo je LR, PR, MLR, PCR, PLS i ANN modelima, od kojih su ANN modeli preporučeni kao najbolji za predviđanje $\log k$ parametra strukturno sličnih androstanskih derivata. Kod većine QSRR modela najznačajniji uticaj na retenciju pokazali su deskriptori lipofilnosti. Odabir najkvalitetnijih QSRR modela izveden je na osnovu statističkih parametara i pomoću SRD analize.



Na osnovu rezultata QSAR analize citotoksičnosti 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata prema ćelijskoj liniji kancera prostate (AR-neg. PC-3) mogu se izvesti sledeći zaključci:

- Adekvatnim odabirom deskriptora izdvojena je podgrupa od tri ADMET deskriptora (SP, MDCK i S+SF). Ovi deskriptori su u bliskoj vezi sa lipofilnošću supstance i oslikavaju njenu mogućnost dostizanja mesta delovanja u organizmu.
- Nelinearnim modelovanjem pomoću veštačkih neuronskih mreža dobijeni su ANN modeli koji omogućavaju predviđanje citotoksičnosti analiziranih 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata prema ćelijama kancera prostate pomoću odabranih ADMET deskriptora.
- Primenom predstavljenih ANN modela (MLP 3-4-1, MLP 3-15-1 i MLP 3-71-1) potraga za novim 17α -pikolil i $17(E)$ -pikoliniliden androstanskim derivatima sa jakom antiproliferativnom aktivnošću prema AR-neg. PC-3 ćelijskoj liniji može biti olakšana.
- Razmatranjem statističkih parametara i rezultata SRD analize ANN-QSAR modela, model MLP 3-4-1 se može preporučiti kao optimalno rešenje za predviđanje citotoksične aktivnosti 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata prema AR-neg. PC-3 ćelijskoj liniji. Odabir ovog modela kao optimalnog zapravo predstavlja kompromis između različitih kriterijuma odabira modela.
- Formirani ANN-QSAR modeli mogu poslužiti kao smernice za sintezu novih 17α -pikolil i $17(E)$ -pikoliniliden androstanskih derivata sa jakom antikancerogenom aktivnošću prema ćelijama kancera prostate.

Prikazani rezultati ukazuju na to da su ciljevi ove doktorske disertacije, predstavljeni u uvodu, u potpunosti ispunjeni.



Conclusion

The results of RP-HPLC chromatographic analysis and chemometric analysis of the chromatographic behavior (chromatographic lipophilicity) of a series of 17 α -picolyl and 17(*E*)-picolinylidene androstane derivatives, presented in this doctoral dissertation, lead to the following conclusions:

- The analyzed androstane derivatives have higher retention in chromatographic system with the mobile phase with methanol concentration of $\varphi = 0,70$ v/v, than in the system with methanol concentration of $\varphi = 0,90$ v/v. The significant difference between the retention behavior of the group of 17 α -picolyl androstane derivatives and retention behavior of the group of 17(*E*)-picolinylidene androstane derivatives is noticeable in the mobile phase with lower methanol concentration. This was confirmed by Wald-Wolfowitz test. Therefore, in that case the capacity factor (*k*) can be considered as discriminating factor of these groups of androstane derivatives.
- The analysis of the influence of substituents on the retention behavior showed that the substitution of 17 α -picolyl and 17 β -hydroxy functional groups with 17(*E*)-picolinylidene group leads to the significant increase of the retention in the applied chromatographic system. The highest retention (chromatographic lipophilicity) has the group of 17(*E*)-picolinylidene androstane derivatives.
- Based on the results of computational modeling of hydrophilic-lipophilic (HILI) surfaces it can be seen that lipophilic character is dominant in most analyzed androstane derivatives. This is in accordance with the assumption that the analyzed 17 α -picolyl and 17(*E*)-picolinylidene androstane derivatives have moderate polarity.
- Based on calculated lipophilicity descriptors ($\log P$, $\log D$, $\log S$) the analyzed compounds have moderate or high lipophilicity. Strong correlation between the lipophilicity descriptors and retention parameter $\log k$ was determined. This confirms the assumption that the $\log k$ can be considered as the alternative lipophilicity parameter of 17 α -picolyl and 17(*E*)-picolinylidene androstane derivatives.
- The results of cluster analysis and principal components analysis indicate the differences between the 17 α -picolyl and 17(*E*)-picolinylidene groups of androstane derivatives, particularly on the basis of lipophilicity descriptors. However, these differences are not drastic, so QSRR and QSAR models are based on both groups of androstane derivatives.



- Linear and non-linear QSRR modeling resulted in LR, PR, MLR, PCR, PLS and ANN models. The ANN models have the best statistical characteristics and can be recommended for prediction of $\log k$ parameters of structurally similar androstane derivatives. The lipophilicity descriptors exhibited the most significant influence on the retention in majority of the established QSRR models. The selection of the best quality QSRR models was achieved on the basis of statistical parameters and SRD analysis.

Based on the results of QSAR analysis of cytotoxicity of 17α -picolyl and $17(E)$ -picolinylidene androstane derivatives toward prostate cancer cell line (AR-neg. PC-3) the following conclusions can be made:

- The subgroup of three ADMET (SP, MDCK i S+SF) descriptors was derived by appropriate descriptors-selection methods. These descriptors are in closely related to lipophilicity of a substance and reflect its ability to reach the site of action in an organism.
- Non-linear modeling, which was based on artificial neural networks, resulted in several ANN models which can be used for prediction of cytotoxicity of 17α -picolyl and $17(E)$ -picolinylidene androstane derivatives toward prostate cancer cell line on the basis of selected ADMET descriptors.
- The application of selected ANN models (MLP 3-4-1, MLP 3-15-1 i MLP 3-71-1) could facilitate the search for new 17α -picolyl and $17(E)$ -picolinylidene androstane derivatives with strong antiproliferative activity toward AR-neg. PC-3 cell line.
- Evaluation of statistical parameters and results of SRD analysis of ANN-QSAR models showed that the model MLP 3-4-1 can be recommended as the optimal solution for prediction of cytotoxic activity of 17α -picolyl and $17(E)$ -picolinylidene androstane derivatives. The selection of this ANN model as the optimal one is a compromise between different model selection criteria.
- The obtained ANN-QSAR models could be used as guidelines for further syntheses of new 17α -picolyl and $17(E)$ -picolinylidene androstane derivatives with significant cytotoxic activity toward prostate cancer cells.

The presented results imply that the goals of this doctoral dissertation, which are listed in the Introduction section, are completely achieved.



6. Literatura

- [1] Koolman J, Klaus-Heinrich R, *Color atlas of biochemistry, second edition, revised and enlarged*. Thieme, Stuttgart, New York, **2005**.
- [2] Arsenijević S, *Organska hemija*, 8. dopunjeno izdanje. Partenon, Beograd, **2001**.
- [3] Karlson P, *Biokemija za studente kemije i medicine*. Školska knjiga, Zagreb, **1993**.
- [4] Lončar E, *Molekulska struktura i retencija u tečnoj hromatografiji*. Monografija, Univerzitet u Novom Sadu, Tehnološki fakultet, Novi Sad, **2010**.
- [5] Petrovic SM, Loncar E, Kolarov Lj, Pejanovic V, *Correlation between retention and 1-octanol-water partition coefficients of some estrane derivatives in reversed-phase thin-layer chromatography*. Journal of Chromatographic Science, **2002**, 40, 170-175.
- [6] Perišić-Janjić N, Djaković-Sekulić T, Stojanović S, Penov-Gaši K, *Evaluation of the lipophilicity of some dehydroepiandrosterone derivatives using RP-18 HPTLC chromatography*. Chromatographia, **2004**, 60, 201-205.
- [7] Poša M, Sebenji A, Trifunović J, *Influence of temperature on retention parameter of bile acids in normal phase thin-layer chromatography: the role of steroid skeleton*. Acta Chimica Slovenica, **2013**, 60, 151-158.
- [8] Milošević NP, Stojanović SZ, Penov-Gaši K, Perišić-Janjić N, Kaliszan R, *Reversed- and normal-phase liquid chromatography in quantitative structure retention-property relationships of newly synthesized seco-androstene derivatives*. Journal of Pharmaceutical and Biomedical Analysis, **2014**, 88, 636-642.
- [9] Cazes J, editor, *Encyclopedia of chromatography, update supplement*. Marcel Dekker Inc., New York, **2004**.
- [10] Locatelli M, Governatori L, Carlucci G, Genovese S, Mollica A, Epifano F, *Recent application of analytical methods to phase I and phase II drugs development: a review*. Biomedical Chromatography, **2012**, 26, 283-300.
- [11] Nord LI, Fransson D, Jacobsson SP, *Prediction of liquid chromatographic retention times of steroids by three-dimensional structure descriptors and partial least squares modeling*. Chemometrics and Intelligent Laboratory Systems, **1998**, 44, 257-269.
- [12] Liu X, Tanaka H, Yamauchi A, Testa B, Chuman H, *Determination of lipophilicity by reversed-phase high-performance liquid chromatography, influence of 1-octanol in the mobile phase*. Journal of Chromatography A, **2005**, 1091, 51-59.
- [13] Benhaim D, Grushka E, *Effect of n-octanol in the mobile phase on lipophilicity determination by reversed phase high-performance liquid chromatography on a modified silica column*. Journal of Chromatography A, **2008**, 1209, 111-119.
- [14] Poša M, Pilipović A, Lalić M, Popović J, *Hydrophobicity and retention coefficient of selected bile acid oxo derivatives*. Acta Chimica Slovenica, **2010**, 57, 828-835.
- [15] Poša M, Pilipović A, Lalić M, Popović J, *Determination and importance of temperature dependence of retention coefficient (RPHPLC) in QSAR model of nitrazepam's partition coefficient in bile acid micelles*. Talanta, **2011**, 83, 1634-1642.



- [16] Rouessac F, Rouessac A, *Chemical analysis, modern instrumentation methods and techniques, 2nd edition*. John Wiley & Sons Ltd, UK, **2007**.
- [17] Mi X, Li S, Wang K, Zhu D, Chen G, *Quantitative determination of 26 steroids in eggs from various species using liquid chromatography–triple quadrupole–mass spectrometry*. Journal of Chromatography A, **2014**, 1356, 54-63.
- [18] Wu L, Hu C, Dung Y, Wu T, *In-capillary formation of polymer/surfactant complexes–assisted reversed-migration micellar electrokinetic chromatography for facile analysis of neutral steroids*. Talanta, **2013**, 107, 389-395.
- [19] Nováková L, Chocholouš P, Solich P, *Ultra-fast separation of estrogen steroids using subcritical fluid chromatography on sub-2-micron particles*. Talanta, **2014**, 121, 178-186.
- [20] Witting M, Rudloff H, Thondamal M, Aguilaniu H, Schmitt-Kopplin P, *Fast separation and quantification of steroid hormones Δ^4 - and Δ^7 -dafachronic acid in *Caenorhabditis elegans**. Journal of Chromatography B, **2015**, 978/979, 118-121.
- [21] Hansen M, Jacobsen NW, Nielsen FK, Björklund E, Styrihave B, Halling-Sørensen B, *Determination of steroid hormones in blood by GC-MS/MS*. Analytical and Bioanalytical Chemistry, **2011**, 400, 3409-3417.
- [22] Figueroa-Valverde L, Díaz-Cedillo F, García-Cervera E, Pool-Gómez E, López-Ramos M, Rosas-Nexticapa M, Hau-Heredia L, Sarabia-Alcocer B, *Synthesis and antibacterial activity evaluation of two androgen derivatives*. Steroids, **2015**, 93, 8-15.
- [23] Cui J, Liu L, Zhao D, Gan C, Huang X, Xiao Q, Qi B, Yang L, Huang Y, *Synthesis, characterization and antitumor activities of some steroidal derivatives with side chain of 17-hydrazone aromatic heterocycle*. Steroids, **2015**, 95, 32-38.
- [24] Richmond V, Careaga VP, Sacca P, Calvo JC, Maier MS, *Synthesis and cytotoxic evaluation of four new 6E-hydroximinosteroids*. Steroids, **2014**, 84, 7-10.
- [25] Savić MP, Djurendić EA, Petri ET, Čelić A, Klisurić OR, Sakač MN, Jakimov DS, Kojić VV, Penov Gaši KM, *Synthesis, structural analysis and antiproliferative activity of some novel D-homo lactone androstane derivatives*. RSC Advances, **2013**, 3, 10385-10395.
- [26] Shingate BB, Hazra BG, Salunke DB, Pore VS, Shirazi F, Deshpande MV, *Synthesis and antimicrobial activity of novel oxysterols from lanosterol*. Tetrahedron, **2013**, 69, 11155-11163.
- [27] Yu B, Shi X, Zheng Y, Fang Y, Zhang E, Yu D, Liu H, *A novel [1,2,4] triazolo [1,5-a] pyrimidine-based phenyl-linked steroid dimer: Synthesis and its cytotoxic activity*. European Journal of Medicinal Chemistry, **2013**, 69, 323-330.
- [28] Nickisch K, Nair HB, Kesavaram N, Das B, Garfield R, Shi S, Bhaskaran SS, Grimm SL, Edwards DP, *Synthesis and antiprogesterone properties of novel 17-fluorinated steroids*. Steroids, **2013**, 78, 909-919.
- [29] Kakati D, Sarma RK, Saikia R, Barua NC, Sarma JC, *Rapid microwave assisted synthesis and antimicrobial bioevaluation of novel steroidal chalcones*. Steroids, **2013**, 78, 321-326.
- [30] Shingate BB, Hazra BG, Salunke DB, Pore VS, Shirazi F, Deshpande MV, *Stereoselective synthesis and antimicrobial activity of steroidal C-20 tertiary alcohols with thiazole/pyridine side chain*. European Journal of Medicinal Chemistry, **2011**, 46, 3681-3689.
- [31] Savić M, *Sinteza i antiproliferativna aktivnost novih D-homo i D-seko derivata androstana*. Doktorska disertacija, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Novi Sad, **2012**.
- [32] Belitz HD, Grosch W, Schieberle, *Food chemistry, 4th revised and extended edition*. Springer-Verlag, Berlin, **2009**.
- [33] Internet prezentacija Svetske zdravstvene organizacije (WHO - World Health Organization), *Cancer prevention and control*, <http://www.who.int/nmh/a5816/en/> (08.02.2015.)
- [34] Ellem SJ, Risbridger GP, *Treating prostate cancer: a rationale for targeting local oestrogens*. Nature Reviews Cancer, **2007**, 7, 621-627.



- [35] Berman DM, Rodriguez R, Veltri RW, *Development, molecular biology and physiology of the prostate (Chapter 90)*. In: *Campbell-Walsh Urology, 10th edition* (Wein AJ, Kavoussi LR, Novick AC, Partin AW, Peters CA, editors). Saunders, USA, **2012**.
- [36] Saylor PJ, *Prostate cancer: The androgen receptor remains front and centre*. *Nature Reviews Clinical Oncology*, **2013**, 10, 126-128.
- [37] Ajduković J, *Sinteza i biološka aktivnost 17-supstituisanih androstanskih derivata*. Doktorska disertacija, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Novi Sad, **2013**.
- [38] Mohareb RM, Abbas NS, Abdelaziz MA, *Heterocyclic ring extension of androstenedione: Synthesis and cytotoxicity of fused pyran, pyrimidine and thiazole derivatives*. *Steroids*, **2014**, 86, 45-55.
- [39] Ajduković JJ, Penov Gaši KM, Jakimov DS, Klisurić OR, Jovanović-Šanta SS, Sakač MN, Aleksić LD, Djurendić EA, *Synthesis, structural analysis and antitumor activity of novel 17 α -picolyl and 17(E)-picolinylidene A-modified androstane derivatives*. *Bioorganic & Medicinal Chemistry*, **2015**, 23, 1557-1568.
- [40] Penov-Gaši KM, Djurendić Brenesel MDj, Djurendić EA, Sakač MN, Čanadi JJ, Daljev JJ, Armbruster T, Andrić S, Sladić DM, Božić TT, et al. *Synthesis and biological evaluation of some 17-picolyl and 17-picolinylidene androst-5-ene derivatives*. *Steroids*, **2007**, 72, 31-40.
- [41] Djurendić E, Daljev J, Sakač M, Čanadi J, Jovanović Šanta S, Andrić S, Klisurić O, Kojić V, Bogdanović G, Djurendić-Brenesel M, et al. *Synthesis of some epoxy and/or N-oxy 17-picolyl and 17-picolinylidene-androst-5-ene derivatives and evaluation of their biological activity*. *Steroids*, **2008**, 73, 129-138.
- [42] Djurendić EA, Ajduković JJ, Sakač MN, Čanadi JJ, Kojić VV, Bogdanović GM, Penov Gaši KM, *17-Picolinylidene-substituted steroid derivatives and their antiaromatase and cytotoxic activity*. *ARKIVOC*, **2009**, 13, 311-323.
- [43] Djurendić EA, Ajduković JJ, Sakač MN, Čanadi JJ, Kojić VV, Bogdanović GM, Penov Gaši KM, *Synthesis and cytotoxic activity of some 17-picolyl and 17-picolinylidene androstane derivatives*. *European Journal of Medicinal Chemistry*, **2012**, 54, 784-792.
- [44] Ajduković JJ, Djurendić EA, Petri ET, Klisurić OR, Čelić AS, Sakač MN, Jakimov DS, Penov Gaši KM, *17(E)-picolinylidene androstane derivatives as potential inhibitors of prostate cancer cell growth: antiproliferative activity and molecular docking studies*. *Bioorganic & Medicinal Chemistry*, **2013**, 21, 7257-7266.
- [45] Jensen F, *Introduction to computational chemistry, 2nd edition*. John Wiley & Sons, Ltd. UK, **2007**.
- [46] Cramer CJ, *Essentials of computational chemistry – theories and models, 2nd edition*. John Wiley & Sons, Ltd. UK, **2004**.
- [47] Lewars E, *Computational chemistry – Introduction to the theory and applications of molecular and quantum mechanics*. Kluwer Academic Publishers, New York, **2004**.
- [48] Cambridge Crystallographic Database, <http://www.ccdc.cam.ac.uk/> (**10.02.2015.**)
- [49] Brookhaven Protein Data Bank, <http://www.rcsb.org/> (**10.02.2015.**)
- [50] Sadowski J, *Three-dimensional structure generation*. In: *Molecular drug properties – Measurement and prediction* (Mannhold R, Kubinyi H, Folkers G, editors). Wiley-VCH, Weinheim, Germany, **2008**.
- [51] Todeschini R, Consonni V, *Handbook of molecular descriptors*. In: *Methods and principles in medicinal chemistry* (Mannhold R, Kubinyi H, Timmerman H, editors). Wiley-VCH, Germany, Weinheim, **2000**.
- [52] Katritzky AR, Kuanar M, Slavov S, Hall CD, *Quantitative correlation of physical and chemical properties with chemical structure: Utility of prediction*. *Chemical Reviews*, **2010**, 110, 5714-5789.
- [53] Kovačević SZ, Jevrić LR, Podunavac Kuzmanović SO, Lončar ES, *Chemometric estimation of the RP TLC retention behaviour of some estrane derivatives by using multivariate regression analysis*. *Central European Journal of Chemistry*, **2013**, 11, 2031-2039.
- [54] Podunavac-Kuzmanović SO, *Analiza zavisnosti antimikrobne aktivnosti od strukture derivata benzimidazola*. Monografija, Univerzitet u Novom Sadu, Tehnološki fakultet, Novi Sad, **2009**.
- [55] Sangster J, *Octanol-water partition coefficient of simple organic compounds*. *Journal of Physical and Chemical Reference Data*, **1989**, 18, 1111-1229.
- [56] Balaban AT, *Chemical application of graph theory*. Academic Press, London, UK, **1976**.



- [57] Roy K, *Topological descriptors in drug design and modeling studies*. Molecular Diversity, **2004**, 8, 321-323.
- [58] Karelson M, Lobanov VS, Katritzky AR, *Quantum-chemical descriptors in QSAR/QSPR studies*. Chemical Reviews, **1996**, 96, 1027-1044.
- [59] Molina E, Estrada E, Nodarse D, Torres LA, González H, Uriarte E, *Quantitative structure-antibacterial activity relationship modeling using a combination of piecewise linear regression-discriminant analysis (I): Quantum chemical, topographic, and topological descriptors*. International Journal of Quantum Chemistry, **2008**, 108, 1856-1871.
- [60] CambridgeSoft Corporation, PerkinElmer Inc., ChemBio3D software version 14.0, **2014**, <http://www.cambridgesoft.com/>
- [61] Biologics Suite 2015-1, BioLuminate version 1.7, Schrödinger, LLC, New York, NY, **2015**, <http://www.schrodinger.com/>
- [62] Yamashita S, Furubayashi T, Kataoka M, Sakane T, Sezaki H, Tokuda H, *Optimized conditions for prediction of intestinal drug permeability using Caco-2 cells*. European Journal of Pharmaceutical Sciences, **2000**, 10, 195-204.
- [63] Irvine JD, Takahashi L, Lockhart K, Cheong J, Tolan JW, Selick HE, Grove JR, *MDCK (Madin-Darby canine kidney) cells: A tool for membrane permeability screening*. Journal of Pharmaceutical Sciences, **1999**, 88, 28-33.
- [64] Zhao YH, Le J, Abraham MH, Hersey A, Eddershaw PJ, Luscombe CN, Butina D, Beck G, Sherborne B, Cooper I, Platts JA, *Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors*. Journal of Pharmaceutical Sciences, **2001**, 90, 749-784.
- [65] Liu R, Sun H, So S, *Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 2. Blood-brain barrier penetration*. Journal of Chemical Information and Modeling, **2001**, 41, 1623-1632.
- [66] Hou TJ, Xu XJ, *ADME evaluation in drug discovery. 3. Modeling blood-brain barrier partitioning using simple molecular descriptors*. Journal of Chemical Information and Modeling, **2003**, 43, 2137-2152.
- [67] Singh S, Singh J, *Transdermal drug delivery by passive diffusion and iontophoresis: a review*. Medicinal Research Reviews, **1993**, 13, 569-621.
- [68] Wang J, Krudy G, Xie X, Wu C, Holland G, *Genetic algorithm-optimized QSPR models for bioavailability, protein binding, and urinary excretion*. Journal of Chemical Information and Modeling, **2006**, 46, 2674-2683.
- [69] Rodgers SL, Davis AM, Tomkinson NP, *QSAR modeling using automatically updating correction libraries: application to a human plasma protein binding model*. Journal of Chemical Information and Modeling, **2007**, 47, 2401-2407.
- [70] PreADMET software, <http://preadmet.bmdrc.org/>
- [71] Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, Lee PW, Tang Y, *AdmetSAR: a comprehensive source and free tool for evaluating chemical ADMET properties*. Journal of Chemical Information and Modeling, **2012**, 52, 3099-3105.
- [72] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ, *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Advanced Drug Delivery Reviews, **1997**, 23, 3-25.
- [73] Carbonell P, Carlsson L, Faulon JL, *Stereo signature molecular descriptor*. Journal of Chemical Information and Modeling, **2013**, 53, 887-897.
- [74] Rabal O, Oyarzabal J, *Using novel descriptor accounting for ligand-receptor interactions to define and visually explore biologically relevant chemical space*. Journal of Chemical Information and Modeling, **2012**, 52, 1086-1102.
- [75] Poša M, Rašeta M, Kuhajda K, *A contribution to the study of hydrophobicity (lipophilicity) of bile acids with an emphasis on oxo derivatives of 5 β -cholanoic acid*. Hemijska industrija, **2011**, 65, 115-121.



- [76] Wold S, *Chemometrics; What do we mean with it, and what do we want from it*. Chemometrics and Intelligent Laboratory Systems, **1995**, 30, 109-115.
- [77] Brereton R, *Applied chemometrics for scientists*. John Wiley & Sons, Ltd. UK, **2007**.
- [78] Kovačević SZ, Podunavac-Kuzmanović SO, Jevrić LR, Djurendić EA, Ajduković JJ, *Non-linear assessment of anticancer activity of 17-picoyl and 17-picolinylidene androstane derivatives – Chemometric guidelines for further syntheses*. European Journal of Pharmaceutical Sciences, **2014**, 62, 258-266.
- [79] Kovačević S, Podunavac-Kuzmanović SO, Jevrić LR, *Multivariate regression modelling of antifungal activity of some benzoxazole and oxazolo[4,5-b]pyridine derivatives*. Acta Chimica Slovenica, **2013**, 60, 756-762.
- [80] Podunavac-Kuzmanović SO, Cvetković DD, Barna DJ, *QSAR Analysis of 2-amino or 2-methyl-1-substituted benzimidazoles against Pseudomonas aeruginosa*. International Journal of Molecular Sciences, **2009**, 10, 1670-1682.
- [81] Podunavac-Kuzmanović S, Markov S, Barna D, *Relationship between the lipophilicity and antifungal activity of some benzimidazole derivatives*. Journal of Theoretical and Computational Chemistry, **2007**, 6, 687-698.
- [82] Podunavac-Kuzmanović SO, Cvetković DD, *Antibacterial evaluation of some benzimidazole derivatives and their zinc(II) complexes*. Journal of the Serbian Chemical Society, **2007**, 75, 459-466.
- [83] Gadzuric SB, Podunavac Kuzmanovic SO, Jokic AI, Vranes MB, Ajdukovic N, Kovacevic SZ, *Chemometric estimation of post-mortem interval based on Na⁺ and K⁺ concentrations from human vitreous humour by linear least squares and artificial neural networks modelling*. Australian Journal of Forensic Sciences, **2014**, 46, 166-179.
- [84] Jokić S, Podunavac Kuzmanović S, Jevrić L, Sudar R, Vidović S, Aladić K, Kovačević S, *HPLC retention behavior of triacylglycerols extracted from soybean oil by supercritical CO₂*. Croatica Chemica Acta, **2014**, 87, 261-269.
- [85] Jokić S, Vidović S, Zeković Z, Podunavac Kuzmanović S, Jevrić L, Marić B, *Chemometric analysis of tocopherols content in soybean oil obtained by supercritical CO₂*. The Journal of Supercritical Fluids, **2012**, 72, 305-311.
- [86] Malbaša R, Jevrić L, Lončar E, Vitas J, Podunavac-Kuzmanović S, Milanović S, Kovačević S, *Chemometric approach to texture profile analysis of kombucha fermented milk products*. Journal of Food Science and Technology, Mysore, **2015**, in press. DOI: 10.1007/s13197-014-1648-4
- [87] Kovačević SZ, Jevrić LR, Podunavac Kuzmanović SO, Lončar ES, *Prediction of in-silico ADME properties of 1,2-O-isopropylidene aldohexose derivatives*. Iranian Journal of Pharmaceutical Research, **2014**, 13, 899-907.
- [88] Dodić J, Grahovac J, Kalajdžija N, Kovačević S, Jevrić L, Podunavac Kuzmanović S, *Chemometric approach to prediction of antibacterial agents production by Streptomyces hygroscopicus*. Applied Biochemistry and Biotechnology, **2014**, 174, 534-541.
- [89] Kovačević S, *Ispitivanje korelacije između hemijske strukture, fizičko-hemijskih i retencionih parametara derivata aldopentoza i aldoheksoza hromatografijom na normalnim fazama*. Master rad, Univerzitet u Novom Sadu, Tehnološki fakultet, Novi Sad, **2012**.
- [90] Kovačević SZ, Jevrić LR, Podunavac Kuzmanović SO, Kalajdžija ND, Lončar ES, *Quantitative structure-retention relationship analysis of some xylofuranose derivatives by linear multivariate method*. Acta Chimica Slovenica, **2013**, 60, 420-428.
- [91] Jevrić LR, Podunavac Kuzmanović SO, Švarc-Gajić JV, Tepić AN, Kovačević SZ, Kalajdžija ND, *QSRR modeling of retention behavior of some s-triazine derivatives*. Acta Chimica Slovenica, **2013**, 60, 732-742.
- [92] Trifković J, Andrić F, Ristivojević P, Andrić D, Tešić ŽLj, Milojković-Opsenica DM, *Structure-retention relationship study of arylpiperazines by linear multivariate modeling*. Journal of Separation Science, **2010**, 33, 2619-2628.
- [93] Karadžić M, Jevrić L, Podunavac Kuzmanović S, Kovačević S, Lončar E, *Structure-retention analysis of some 1,6-anhydrohexose and D-aldopentose derivatives by linear multivariate approach*. Journal of Liquid Chromatography & Related Technologies, **2015**, 38, 662-669.



- [94] Kovačević SZ, Podunavac Kuzmanović SO, Jevrić LR, Lončar ES, *Assessment of chromatographic lipophilicity of some anhydro-D-aldose derivatives on different stationary phases by QSRR approach*. Journal of Liquid Chromatography & Related Technologies, **2015**, 38, 492-500.
- [95] Perišić-Janjić NU, Podunavac-Kuzmanović SO, Balaž JS, Vlaović Dj, *Chromatographic behaviour and lipophilicity of some benzimidazole derivatives*. Journal of Planar Chromatography, **2000**, 13, 123-129.
- [96] Can A, *Quantitative structure-toxicity relationship (QSTR) studies on the organophosphate insecticides*. Toxicology Letters, **2014**, 230, 434-443.
- [97] Farkas O, Jakus J, Héberger K, *Quantitative structure - antioxidant activity relationships of flavonoid compounds*. Molecules, **2004**, 9, 1079-1088.
- [98] Wang SR, Chen C, Xiong MJ, Wu LP, Ye LM, *Quantitative retention-activity relationship models of angiotensin converting enzyme inhibitors using biopartitioning micellar chromatography*. Journal of Chromatographic Science, **2010**, 48, 134-139.
- [99] Kaliszan R, *QSRR: Quantitative structure-(chromatographic) retention relationships*. Chemical Reviews, **2007**, 107, 3212-3246.
- [100] Héberger K, *Quantitative structure-(chromatographic) retention relationships*. Journal of Chromatography A, **2007**, 1158, 273-305.
- [101] Kaliszan R, Wiczling P, Markuszewski MJ, Al-Haj MA, *Thermodynamic vs. extrathermodynamic modeling of chromatographic retention*. Journal of Chromatography A, **2011**, 1218, 5120-5130.
- [102] Didaoui L, Touabet A, Badjah Hadj Ahmed AY, Meklati BY, *Evaluation of dead time calculation in reversed-phase liquid chromatography using a multiparametric mathematical method*. Journal of High Resolution Chromatography, **1999**, 22, 559-564.
- [103] Kaliszan R, Kaliszan A, Noctor TAG, *Mechanism of retention of benzodiazepines in affinity, reversed-phase and adsorption high-performance liquid chromatography in view of quantitative structure-retention relationships*. Journal of Chromatography, **1992**, 609, 69-81.
- [104] Cirera-Domènech E, Estrada-Tejedor R, Broto-Puig F, Teixidó J, Gassiot-Matas M, Comellas L, Lliberia JL, Méndez A, Paz-Estivill S, Delgado-Ortiz MR, *Quantitative structure-retention relationships applied to liquid chromatography gradient elution method for the determination of carbonyl-2,4-dinitrophenylhydrazone compounds*. Journal of Chromatography A, **2013**, 1276, 65-77.
- [105] Nasal A, Siluk D, Kaliszan R, *Chromatographic retention parameters in medicinal chemistry and molecular pharmacology*. Current Medicinal Chemistry, **2003**, 10, 381-426.
- [106] Ahmed Al-Haj M, Kaliszan R, Nasal A, *Test analytes for studies of the molecular mechanism of chromatographic separations by quantitative structure-retention relationships*. Analytical Chemistry, **1999**, 71, 2976-2985.
- [107] Ahmed Al-Haj M, Kaliszan R, Buszewski B, *Quantitative structure-retention relationships with model analytes as a means of an objective evaluation of chromatographic columns*. Journal of Chromatographic Science, **2001**, 39, 29-38.
- [108] Odović JV, Trbojević-Stanković JB, *Hromatografske metode u proučavanju lipofilnosti biološki aktivnih supstanci*. Hemijska industrija, **2009**, 63, 33-37.
- [109] Ayouni L, Cazorla G, Chaillou D, Herbreteau B, Rudaz S, Lantéri P, Carrupt PA, *Fast determination of lipophilicity by HPLC*. Chromatographia, **2005**, 62, 251-255.
- [110] Soczewiński E, Wachtmeister CA, *The relation between the composition of certain ternary two-phase solvent systems and R_M values*. Journal of Chromatography, **1962**, 7, 311-320.
- [111] Snyder LR, Dolan JW, Gant JR, *Gradient elution in high-performance liquid chromatography: I. Theoretical basis for reversed-phase systems*. Journal of Chromatography, **1979**, 165, 3-30.
- [112] Baczek T, Markuszewski M, Kaliszan R, van Straten MA, Claessens HA, *Linear and quadratic relationships between retention and organic modifier content in eluent in reversed phase high-performance liquid chromatography: A systematic comparative statistical study*. Journal of High Resolution Chromatography, **2000**, 23, 667-676.



- [113] Kowalska S, Krupczyńska K, Buszewski B, *Some remarks on characterization and application of stationary phases for RP-HPLC determination of biologically important compounds*. Biomedical Chromatography, **2006**, 20, 4-22.
- [114] Goryński K, Bojko B, Nowaczyk A, Buciński A, Pawliszyn J, Kaliszan R, *Quantitative structure-retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: endogenous metabolites and banned compounds*. Analytica Chimica Acta, **2013**, 797, 13-19.
- [115] Golubović J, Protić A, Zečević M, Otašević B, Mikić M, Živanović L, *Quantitative structure-retention relationships of azole antifungal agents in reversed-phase high performance liquid chromatography*. Talanta, **2012**, 100, 329-337.
- [116] Kaczmarek M, Buciński A, Marszał MP, Badura A, Kaliszan R, *Thermodynamic and QSRR modeling of HPLC retention on modern stationary phases*. Journal of Liquid Chromatography & Related Technologies, **2015**, 38, 62-67.
- [117] Hemmateenejad B, Shamsipur M, Safavi A, Sharghi H, Amiri AA, *Reversed-phase high performance liquid chromatography (RP-HPLC) characteristics of some 9,10-anthraquinone derivatives using binary acetonitrile-water mixtures as mobile phase*. Talanta, **2008**, 77, 351-359.
- [118] Perisic-Janjic N, Kaliszan R, Wiczling P, Milosevic N, Uscumlic G, Banjac N, *Reversed-phase TLC and HPLC retention data in correlation studies with in silico molecular descriptors and druglikeness properties of newly synthesized anticonvulsant succinimide derivatives*. Molecular Pharmaceutics, **2011**, 8, 555-563.
- [119] Jevrić L, *Ispitivanje korelacije između hemijske strukture, fizičko-hemijskih i retencionih parametara u hromatografiji na obrnutim fazama novosintetisanih derivata s-triazina*. Doktorska disertacija, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Novi Sad, **2009**.
- [120] Tosti T, Natić M, Dabić D, Milić D, Milojković-Opsenica D, Tešić Ž, *Structure-retention relationship study of polyoxygenated steroids*. Journal of Separation Science, **2012**, 35, 2693-2698.
- [121] Valkó K, Slégel P, *New chromatographic hydrophobicity index (ϕ_0) based on the slope and the intercept of the log k' versus organic phase concentration plot*. Journal of Chromatography A, 631, **1993**, 49-61.
- [122] Bieganowska ML, Szopa AD, Petruczynik A, *The retention behavior of some sulfonamides on different TLC plates. 2. Comparison of the selectivity of the systems and quantitative determination of hydrophobicity parameters*. Journal of Planar Chromatography, **1995**, 8, 122-128.
- [123] Hansch C, Maloney P, Fujita T, Muir R, *Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients*. Nature, **1962**, 194, 178-180.
- [124] Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, *et al.* QSAR modeling: Where have you been? Where are you going to? Journal of Medicinal Chemistry, **2014**, 57, 4977-5010.
- [125] Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN, *Virtual screening in drug discovery – A computational perspective*. Current Protein and Peptide Science, **2007**, 8, 329-351.
- [126] Harvey AL, Edrada-Ebel R, Quinn RJ, *The re-emergence of natural products for drug discovery in the genomics era*. Nature Reviews Drug Discovery, **2015**, 14, 111-129.
- [127] Andrade CH, Pasqualoto KFM, Ferreira EI, Hopfinger AJ, *4D-QSAR: Perspectives in drug design*. Molecules, **2010**, 15, 3281-3294.
- [128] Vedani A, Dobler M, Lill MA, *Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor*. Journal of Medicinal Chemistry, **2005**, 48, 3700-3703.
- [129] Oliveira FG, Sant'Anna CM, Caffarena ER, Dardenne LE, Barreiro EJ, *Molecular docking study and development of an empirical binding free energy model for phosphodiesterase 4 inhibitors*. Bioorganic & Medicinal Chemistry, **2006**, 14, 6001-6011.
- [130] Cramer RD, Patterson DE, Bunce JD, *Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins*. Journal of the American Chemical Society, **1988**, 110, 5959-5967.



- [131] Selassie CD, *History of quantitative structure-activity relationships*. In: *Burger's medicinal chemistry and drug discovery, 6th edition* (Abraham DJ, editor). John Wiley & Sons Inc., UK, **2003**.
- [132] Thomas G, *Fundamentals of Medicinal Chemistry*. John Wiley & Sons Ltd, UK, **2003**.
- [133] Podunavac-Kuzmanović SO, Cvetković DD, Jevrić LR, Uzelac NJ, *quantitative structure-activity relationship (QSAR) study of a series of benzimidazole derivatives as inhibitors of Saccharomyces cerevisiae*. Acta Chimica Slovenica, **2013**, 60, 26-33.
- [134] Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, Kruger FA, Light Y, Mak L, McGlinchey S, *et al. The ChEMBL bioactivity database: an update*. Nucleic Acids Research, **2014**, 42, 1083-1090. ChEMBL Database, <https://www.ebi.ac.uk/chembl/>
- [135] Podunavac-Kuzmanović SO, Cvetković DD, *Lipophilicity and antifungal activity of some 2-substituted benzimidazole derivatives*. Chemical Industry & Chemical Engineering Quarterly, **2011**, 17, 9-15.
- [136] Podunavac-Kuzmanović SO, Cvetković DD, *QSAR modeling of antibacterial activity of some benzimidazole derivatives*. Chemical Industry and Chemical Engineering Quarterly, **2011**, 17, 33-38.
- [137] Kovačević SZ, Podunavac Kuzmanović SO, Jevrić LR, Kalajdžija ND, *Neural network modelling of antifungal activity of a series of oxazole derivatives based on in silico pharmacokinetic parameters*. Acta Periodica Technologica, **2013**, 44, 249-258.
- [138] Polster J, Schieberle P, *Structure-odor correlations in homologous series of alkanethiols and attempts to predict odor thresholds by 3D-QSAR studies*. Journal of Agricultural and Food Chemistry, **2015**, 63, 1419-1432.
- [139] Wu J, Wang Y, Shen Y, *Molecular docking and QSAR analysis on maleimide derivatives selective inhibition against human monoglyceride lipase based on various modeling methods and conformations*. Chemometrics and Intelligent Laboratory Systems, **2014**, 131, 22-30.
- [140] Gissi A, Lombardo A, Roncaglioni A, Gadaleta D, Mangiatordi GF, Nicolotti O, Benfenati E, *Evaluation and comparison of benchmark QSAR models to predict a relevant REACH endpoint: The bioconcentration factor (BCF)*. Environmental Research, **2015**, 137C, 398-409.
- [141] Ivan D, Crisan L, Funar-Timofei S, Mracec M, *A quantitative structure-activity relationships study for the anti-HIV-1 activities of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives using the multiple linear regression and partial least squares methodologies*. Journal of the Serbian Chemical Society, **2013**, 78, 495-506.
- [142] da Silva DL, Silva Terra B, Ribeiro Lage M, Lúcia Tasca Góis Ruiz A, Capeletti da Silva C, Ernesto de Carvalho J, Walkimar de Mesquita Carneiro J, Terra Martins F, Antonio Fernandes S, de Fátima Â, *Xanthenones: calixarenes-catalyzed syntheses, anticancer activity and QSAR studies*. Organic & Biomolecular Chemistry, **2015**, in press. DOI: 10.1039/c4ob02611j
- [143] Dehmer M, Varmuza K, Bonchev D (editors), *Statistical modelling of molecular descriptors in QSAR/QSPR*. Wiley-VCH Verlag & Co. KGaA, Weinheim, Germany, **2012**.
- [144] Miller JN, Miller JC, *Statistics and chemometrics for analytical chemistry, 6th edition*. Pearson Education Limited, Harlow, UK, **2010**.
- [145] Esbensen KH, *Multivariate data analysis – in practice, 5th edition*. CAMO Software AS, Oslo, Norway, **2009**.
- [146] Varmuza K, Filzmoser P, *Introduction to multivariate statistical analysis in chemometrics*. Taylor & Francis Group, CRC Press, Florida, USA, **2009**.
- [147] Héberger K, *Sum of ranking differences compares methods or models fairly*. Trends in Analytical Chemistry, **2010**, 29, 101-109.
- [148] Héberger K, Kollár-Hunek K, *Sum of ranking differences for method discrimination and its validation: comparison of ranks with random numbers*. Journal of Chemometrics, **2011**, 25, 151-158.
- [149] Kollár-Hunek K, Héberger K, *Method and model comparison by sum of ranking differences in cases of repeated observations (ties)*. Chemometrics and Intelligent Laboratory Systems, **2013**, 127, 139-146.
- [150] Vračko M, Minovski N, Héberger K, *Ranking of QSAR models to predict minimal inhibitory concentrations toward Mycobacterium tuberculosis for a set of fluoroquinolones*. Acta Chimica Slovenica, **2010**, 57, 586-590.



- [151] Héberger K, Škrbić B, *Ranking and similarity for quantitative structure-retention relationship models in predicting Lee retention indices of polycyclic aromatic hydrocarbons*. *Analytica Chimica Acta*, **2012**, 716, 92-100.
- [152] Vastag G, Apostolov S, Perišić-Janjić N, Matijević B, *Multivariate analysis of chromatographic retention data and lipophilicity of phenylacetamide derivatives*. *Analytica Chimica Acta*, **2013**, 767, 44-49.
- [153] Andrić F, Héberger K, *Chromatographic and computational assessment of lipophilicity using sum of ranking differences and generalized pair-correlation*. *Journal of Chromatography A*, **2015**, 1380, 130-138.
- [154] Montgomery DC, Runger GC, *Applied statistics and probability for engineers, 6th edition*. John Wiley & Sons, Inc., USA, **2014**.
- [155] Topliss JG, Costello RJ, *Chance correlations in structure-activity studies using multiple regression analysis*. *Journal of Medicinal Chemistry*, **1972**, 15, 1066-1068.
- [156] Wehrens R, *Chemometrics with R – Multivariate data analysis in the natural sciences and life sciences*. Springer-Verlag, Berlin, Germany, **2011**.
- [157] Eriksson L, Johansson E, Kettaneh-Wold N, Wikström C, Wold S, *Design of experiments – Principles and applications*. Umetrics Academy, Umeå, Sweden, **2008**.
- [158] Farkas O, Zenkevich IG, Stout F, Kalivas JH, Héberger K, *Prediction of retention indices for identification of fatty acid methyl esters*. *Journal of Chromatography A*, **2008**, 1198-1199, 188-195.
- [159] Dudek AZ, Arodz T, Gálvez, *Computational methods in developing quantitative structure-activity relationships (QSAR): A review*. *Combinatorial Chemistry & High Throughput Screening*, **2006**, 9, 213-228.
- [160] Gonçalves Maltarollo V, Honório KM, Borges Ferreira da Silva A, *Applications of artificial neural networks in chemical problems*. *INTECH*, **2013**, 203-223, DOI: 10.5772/51275
- [161] Jayalakshmi T, Santhakumaran A, *Statistical normalization and back propagation for classification*. *International Journal of Computer Theory and Engineering*, **2011**, 3, 89-93.
- [162] Priddy KL, Keller PE, *Artificial neural networks – an introduction*. SPIE, USA, **2005**.
- [163] Richard Bowen W, Jones MG, Yousef HNS, *Dynamic ultrafiltration of proteins – A neural network approach*. *Journal of Membrane Science*, **1998**, 146, 225-235.
- [164] Sarkar B, Sengupta A, De S, DasGupta S, *Prediction of permeate flux during electric field enhanced cross-flow ultrafiltration – A neural network approach*. *Separation and Purification Technology*, **2009**, 65, 260-268.
- [165] Therón R, De Paz JF, *Visual sensitivity analysis for artificial neural networks*. *Intelligent Data Engineering and Automated Learning – IDEAL, 7th International Conference, Conference Proceeding, Burgos, Spain, Lecture Notes in Computer Science*, **2006**, 4224, pp 191-198, DOI: 10.1007/11875581_23
- [166] Shojaeefard MH, Akbari M, Tahani M, Farhani F, *Sensitivity analysis of the artificial neural network outputs in friction stir lap joining of aluminium to brass*. *Advances in Materials Science and Engineering*, **2013**, 2013, 1-7.
- [167] Garkani-Nejad Z, *Use of self-training artificial neural networks in a QSRR study of a diverse set of organic compounds*. *Chromatographia*, **2009**, 70, 869-874.
- [168] Fragkaki AG, Farmaki E, Thomaidis N, Tsantili-Kakoulidou A, Angelis YS, Koupparis M, Georgakopoulos C, *Comparison of multiple linear regression, partial least squares and artificial neural networks for prediction of gas chromatographic relative retention times of trimethylsilylated anabolic androgenic steroids*. *Journal of Chromatography A*, **2012**, 1256, 232-239.
- [169] Myint KZ, Xie XQ, *Recent advances in fragment-based QSAR and multi-dimensional QSAR methods*. *International Journal of Molecular Sciences*, **2010**, 11, 3846-3866.
- [170] Shahlaei M, *Descriptor selection methods in quantitative structure-activity relationship studies: A review study*. *Chemical Reviews*, **2013**, 113, 8093-8103.
- [171] Mehmood T, Hovde Liland K, Snipen L, Sæbø S, *A review of variable selection methods in partial least squares regression*. *Chemometrics and Intelligent Laboratory Systems*, **2012**, 118, 62-69.



- [172] Chong IG, Jun CH, *Performance of some variable selection methods when multicollinearity is present*. Chemometrics and Intelligent Laboratory Systems, **2005**, 78, 103-112.
- [173] Golbraikh A, Tropsha A, *Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection*. Journal of Computer-Aided Molecular Design, **2002**, 16, 357-369.
- [174] Golbraikh A, Tropsha A, *Beware of q^2 !* Journal of Molecular Graphics and Modelling, **2002**, 20, 269-276.
- [175] Lin LI, *A concordance correlation coefficient to evaluate reproducibility*. Biometrics, **1989**, 45, 255-268.
- [176] Shi LM, Fang H, Tong W, Wu J, Perkins R, Blair RM, Branham WS, Dial SL, Moland CL, Sheehan DM, *QSAR models using a large diverse set of estrogens*. Journal of Chemical Information and Computer Sciences, **2001**, 41, 186-195.
- [177] Schüürmann G, Ebert R, Chen J, Wang B, Kühne R, *External validation and prediction employing the predictive squared correlation coefficients test set activity mean vs training set activity mean*. Journal of Chemical Information and Modeling, **2008**, 48, 2140-2145.
- [178] Podunavac-Kuzmanović SO, Velimirović SD, *Correlation between the lipophilicity and antifungal activity of some benzoxazole derivatives*. Acta Periodica Technologica, **2010**, 41, 177-185.
- [179] Chirico N, Gramatica P, *Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient*. Journal of Chemical Information and Modeling, **2011**, 51, 2320-2335.
- [180] Chirico N, Gramatica P, *Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection*. Journal of Chemical Information and Modeling, **2012**, 51, 2320-2335.
- [181] Daljev J, *Sinteza novih 17-pikolil i 17-pikoliniliden derivata 5-androstena i njihova biološka aktivnost*. Magistarski rad, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Novi Sad, **2008**.
- [182] Penov Gaši KM, Gaković AR, Ajduković JJ, Djurendić-Brenesel MDj, Djurendić EA, Savić MP, Sakač MN, *Reactivity of 17 β -hydroxy-17 α -substituted androstane derivatives*. Acta Periodica Technologica, **2010**, 41, 169-176.
- [183] Jović BD, Ajduković JJ, Djurendić EA, Nikolić AD, *FTIR investigation of solvent-induced carbonyl band shifts of 17 β -hydroxy-17 α -picolyl-androst-4-en-3-one*. Acta Periodica Technologica, **2014**, 45, 191-199.
- [184] MarvinSketch 6.1, **2013**, ChemAxon, <http://www.chemaxon.com>
- [185] Simulations Plus, Inc. <http://www.simulations-plus.com/>
- [186] Molinspiration Cheminformatics, online program, <http://www.molinspiration.com/>
- [187] Virtual Computational Chemistry Laboratory, Parameter Client online program, <http://www.vclab.org/>
- [188] ADRIANA.Code software - Calculation of Molecular Descriptors, <https://www.molecular-networks.com/>
- [189] Virtual Computational Chemistry Laboratory ALOGPS 2.1 online program, <http://146.107.217.178/lab/alogps/start.html>
- [190] HyperChem 8.0, Hypercube Inc, <http://www.hyper.com/>
- [191] Avogadro software 1.1.1, <http://avogadro.cc/wiki/>
- [192] StatSoft Inc, STATISTICA (data analysis software system), version 10, **2011**, www.statsoft.com
- [193] Hintze J, NCSS 2007, NCSS, LLC. Kaysville, Utah, USA, **2007**, www.ncss.com
- [194] Microsoft. Microsoft Excel. Computer Software. Redmond, Washington, **2013**.
- [195] OriginPro 9.1, OriginLab Corporation, Northampton, MA, **2013**, www.originlab.com
- [196] Minitab Inc, MINITAB 16.1.1, **2010**, www.minitab.com
- [197] MATLAB R2013a, The MathWorks Inc, Natick, Massachusetts, United States, **2013**, www.mathworks.com
- [198] CRRN (Compare Ranks with Random Numbers) VBA program, <http://knight.kit.bme.hu/CRRN/>
- [199] Bhar L, *Nonparametric tests*. http://www.iasri.res.in/ebook/EBADAT/3-Diagnostics%20and%20Remedial%20Measures/4-Non-Parametric_test.pdf (04.03.2015.)
- [200] Weiner PK, Langridge R, Blaney JM, Schaefer R, Kollman PA, *Electrostatic potential molecular surfaces*. Proceedings of the National Academy of Sciences, **1982**, 79, 3754-3758.



7. Prilog

Formule za računanje parametara unakrsne validacije

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (\text{P-7.1.})$$

$$PRESS = \sum_{i=1}^n (y_i - y_{i,p})^2 \quad (\text{P-7.2.})$$

$$R_{CV}^2 = 1 - \frac{PRESS}{TSS} \quad (\text{P-7.3.})$$

$$SD_{PRESS} = \sqrt{\frac{PRESS}{n}} \quad (\text{P-7.4.})$$

y_i – eksperimentalne vrednosti

\bar{y} – srednja vrednost eksperimentalnih vrednosti

$y_{i,p}$ – predviđena (modelovana) vrednost

n – broj objekata u setu podataka

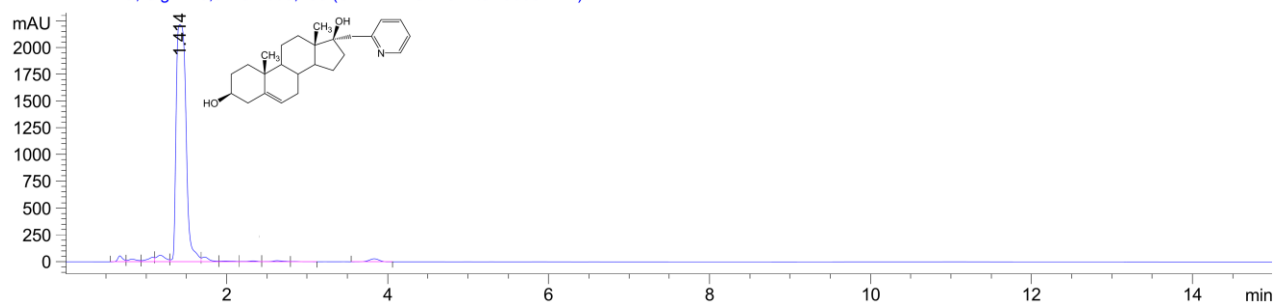
**Tabela P-7.1.** Rezultati određivanja vrednosti kapacitetnog faktora androstanskih derivata pomoću C18-RP-HPLC sistema

Mobilna faza:	Metanol/voda (90/10 v/v)		Metanol/voda (70/30 v/v)	
Jedinjenje	<i>k</i> (±SD)	log <i>k</i> (±SD)	<i>k</i> (±SD)	log <i>k</i> (±SD)
1	1,116 (±0,002)	0,048 (±0,001)	1,323 (±0,002)	0,122 (±0,001)
2	2,996 (±0,003)	0,476 (±0,000)	4,397 (±0,001)	0,643 (±0,000)
3	1,109 (±0,004)	0,045 (±0,002)	1,383 (±0,003)	0,141 (±0,001)
4	1,238 (±0,004)	0,093 (±0,001)	1,473 (±0,003)	0,168 (±0,001)
5	0,371 (±0,003)	-0,431 (±0,003)	0,469 (±0,001)	-0,329 (±0,001)
6	0,744 (±0,004)	-0,128 (±0,002)	1,177 (±0,001)	0,071 (±0,000)
7	2,747 (±0,005)	0,439 (±0,001)	4,478 (±0,001)	0,651 (±0,000)
8	0,408 (±0,001)	-0,389 (±0,001)	1,025 (±0,002)	0,011 (±0,001)
9	0,387 (±0,001)	-0,413 (±0,001)	0,901 (±0,002)	-0,045 (±0,001)
10	5,624 (±0,004)	0,750 (±0,000)	9,368 (±0,001)	0,972 (±0,000)
11	0,378 (±0,003)	-0,423 (±0,004)	0,723 (±0,001)	-0,141 (±0,000)
12	6,498 (±0,008)	0,813 (±0,001)	12,579 (±0,001)	1,100 (±0,000)
13	2,817 (±0,003)	0,450 (±0,001)	3,452 (±0,002)	0,538 (±0,000)
14	1,887 (±0,004)	0,276 (±0,001)	2,369 (±0,003)	0,375 (±0,001)
15	1,845 (±0,003)	0,266 (±0,001)	2,045 (±0,001)	0,311 (±0,000)
16	0,886 (±0,002)	-0,053 (±0,001)	1,641 (±0,002)	0,215 (±0,000)
17	1,251 (±0,002)	0,097 (±0,001)	1,882 (±0,001)	0,275 (±0,000)
18	2,352 (±0,004)	0,371 (±0,001)	5,283 (±0,001)	0,723 (±0,000)
19	2,237 (±0,006)	0,350 (±0,001)	2,496 (±0,001)	0,397 (±0,000)
20	2,592 (±0,006)	0,414 (±0,001)	3,239 (±0,001)	0,510 (±0,000)
21	5,886 (±0,002)	0,770 (±0,000)	10,307 (±0,001)	1,013 (±0,000)
22	0,843 (±0,001)	-0,074 (±0,000)	1,781 (±0,001)	0,251 (±0,000)
23	0,698 (±0,003)	-0,156 (±0,002)	1,485 (±0,004)	0,172 (±0,001)
24	15,636 (±0,003)	1,194 (±0,000)	19,892 (±0,001)	1,299 (±0,000)



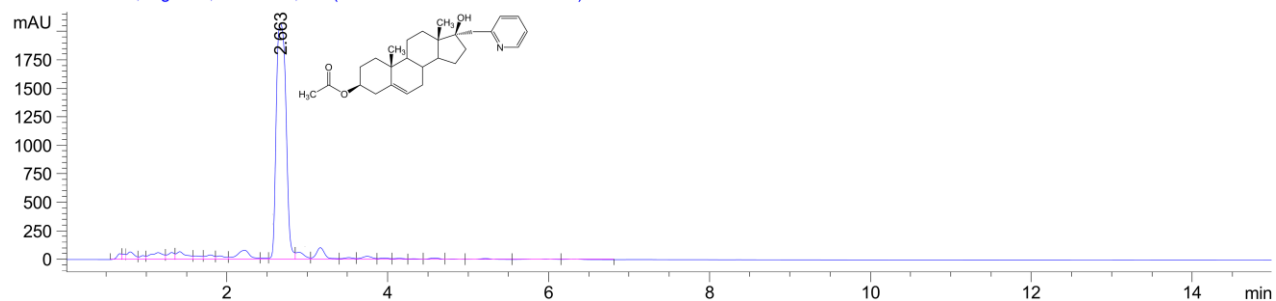
Jedinjenje 1

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000021.D)



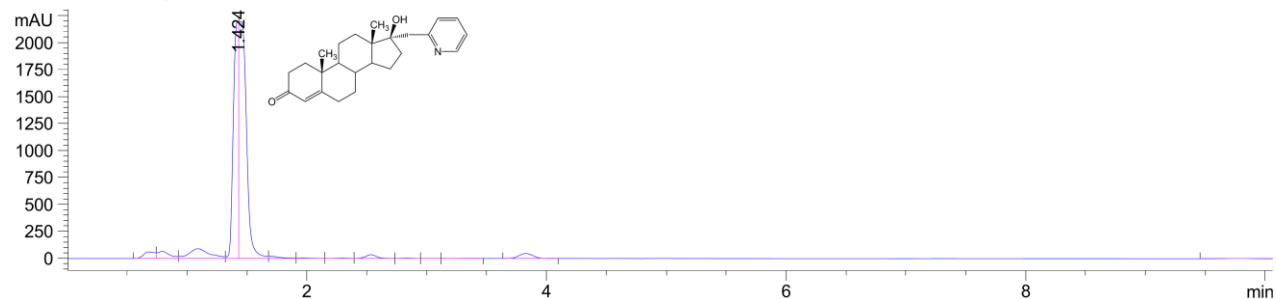
Jedinjenje 2

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000008.D)



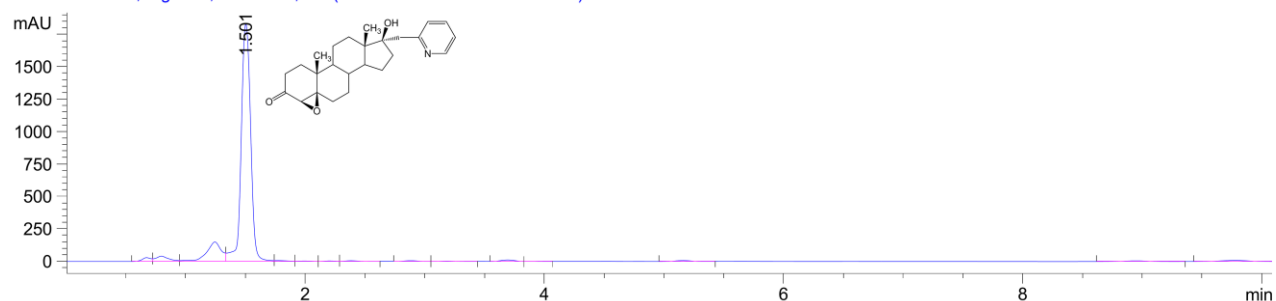
Jedinjenje 3

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000022.D)



Jedinjenje 4

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000023.D)

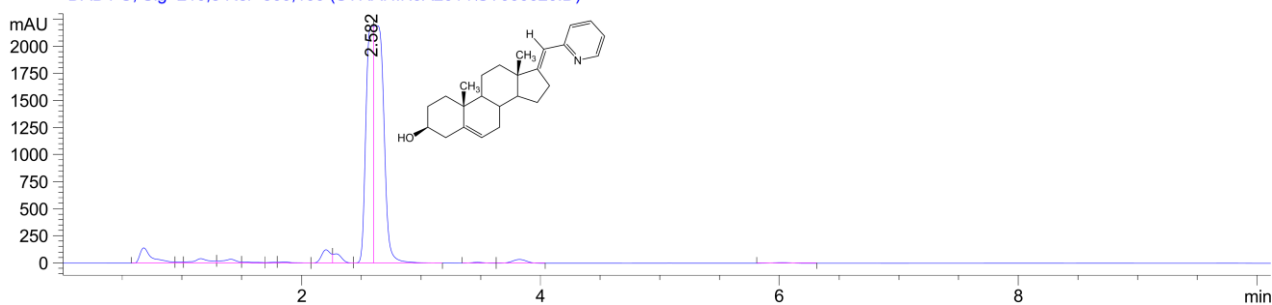


Slika P-7.1. Hromatogrami odabranih 17α-pikolil androstanskih derivata ($\varphi = 0,90$)



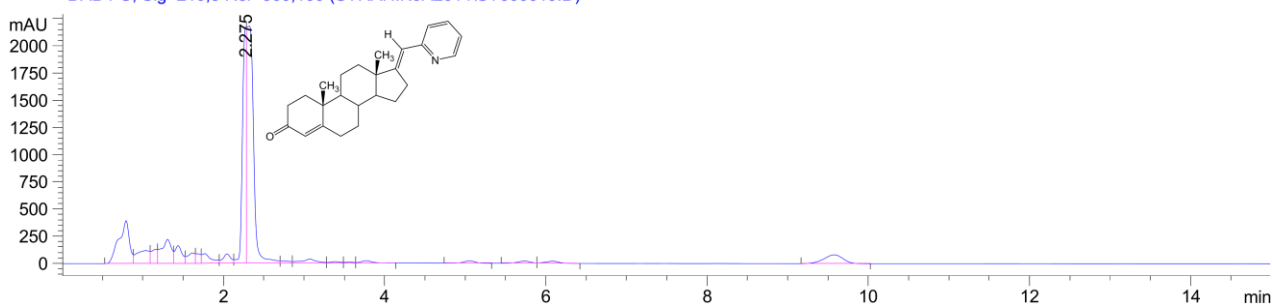
Jedinjenje 13

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000020.D)



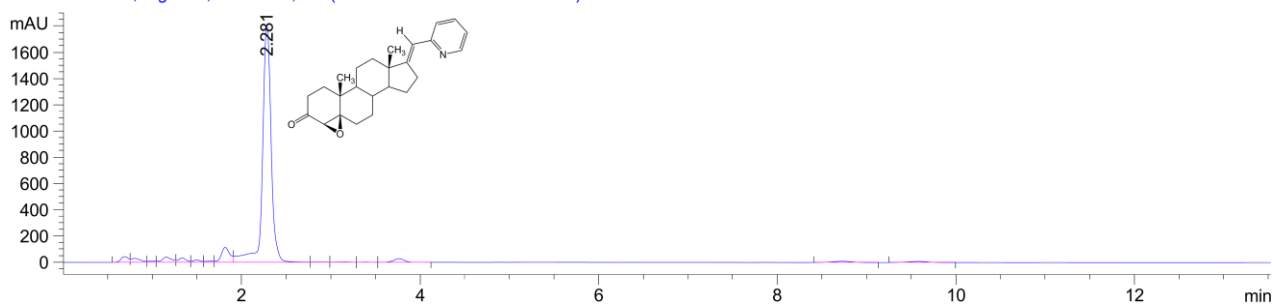
Jedinjenje 14

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000019.D)



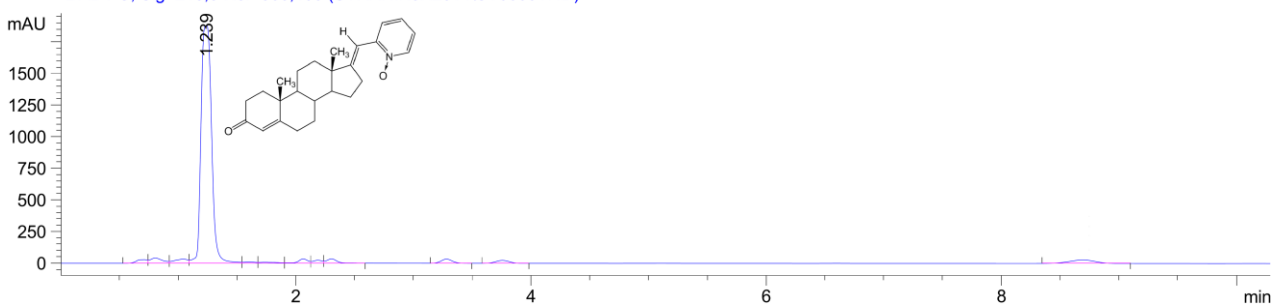
Jedinjenje 15

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000018.D)



Jedinjenje 16

DAD1 C, Sig=210,8 Ref=360,100 (STRAHINJA2014\ST000017.D)



Slika P-7.2. Hromatogrami odabranih 17(*E*)-pikoliniliden androstanskih derivata ($\varphi = 0,90$)



Tabela P-7.2. Vrednosti kvantno-hemijskih deskriptora analiziranih androstanskih derivata dobijenih modelovanjem EP površina i HOMO/LUMO orbitala (E_{min} , E_{max} , E_{sr} – minimalna, maksimalna i srednja vrednost elektrostatickog potencijala, $\Delta E = E_{max} - E_{min}$, E_{HOMO} i E_{LUMO} – energije HOMO i LUMO orbitala, E_{GAP} – energetska praznina)

Jedinjenje	E_{min} [eV]	E_{max} [eV]	ΔE [eV]	E_{sr} [eV]	E_{HOMO} [eV]	E_{LUMO} [eV]	E_{GAP} [eV]
1	-32,8914	20,5245	53,4159	-6,1835	-8,4810	0,8350	9,3160
2	-28,9100	14,3180	43,2280	-7,2960	-8,4930	0,8540	9,3470
3	-23,9038	9,1332	33,0370	-7,3853	-8,3830	-4,0520	4,3310
4	-29,4555	10,9479	40,4034	-9,2538	-8,5070	-2,4960	6,0110
5	-27,9069	12,7736	40,6805	-7,5667	-12,0710	0,8440	12,9150
6	-25,7551	17,2959	43,0510	-4,2296	-8,3500	0,7830	9,1330
7	-24,7850	7,8921	32,6771	-8,4465	-8,4470	0,0450	8,4920
8	-23,1188	20,5366	43,6554	-1,2911	-8,3900	-7,5130	0,8770
9	-23,5547	13,9145	37,4692	-4,8201	-8,4190	-5,4090	3,0100
10	-23,9294	21,5052	45,4346	-1,2121	-8,4150	-2,7550	5,6600
11	-31,2091	29,0187	60,2278	-1,0952	-8,4910	0,8580	9,3490
12	-19,9968	12,3674	32,3642	-3,8147	-7,9870	-2,1620	5,8250
13	-22,3089	23,7128	46,0217	0,7020	-7,9900	-2,1640	5,8260
14	-19,7805	8,7498	28,5303	-5,5154	-7,9800	-4,0670	3,9130
15	-19,2984	9,8107	29,1091	-4,7439	-7,9860	-2,4550	5,5310
16	-19,7723	12,6601	32,4324	-3,5561	-11,2510	-4,0120	7,2390
17	-18,4626	19,6741	38,1367	0,6058	-7,9880	-4,7230	3,2650
18	-17,8046	12,8184	30,6230	-2,4931	-11,6810	-2,1630	9,5180
19	-20,6962	11,8608	32,5570	-4,4177	-7,9840	-3,8850	4,0990
20	-27,8664	13,6236	41,4900	-7,1214	-7,9780	-2,1360	5,8420
21	-21,1495	9,3695	30,5190	-5,8900	-7,9730	-2,1390	5,8340
22	-13,5652	11,5387	25,1039	-1,0133	-7,9980	-7,5330	0,4650
23	-17,0841	12,2078	29,2919	-2,4382	-8,0090	-5,4300	2,5790
24	-19,5922	9,5346	29,1268	-5,0288	-7,9890	-2,7570	5,2320


Tabela P-7.3. Vrednosti deskriptora lipofilnosti analiziranih androstanskih derivata

Jedinjenje	ALOGPs	AClogP	ALOGP	MLOGP	KOWWIN	XLOGP2	XLOGP3	Average logP	ALOGpS	AClogS	Average logS	miLogP	logP _{vg}	logP _{klop}
1	4,25	4,06	3,99	3,98	4,87	4,34	4,29	4,25	-4,71	-4,48	-4,60	4,134	3,220	3,314
2	4,81	4,55	4,37	4,30	5,47	5,08	4,68	4,75	-5,34	-4,89	-5,12	4,838	3,649	3,829
3	3,91	4,22	4,09	3,89	4,24	4,30	3,92	4,08	-4,99	-4,53	-4,76	3,949	4,068	3,926
4	3,35	3,07	3,19	3,17	4,06	3,49	3,53	3,41	-4,80	-4,34	-4,61	3,382	3,331	3,816
5	2,07	2,64	1,99	1,35	3,05	4,69	2,53	2,62	-4,95	-6,76	-5,85	2,154	0,686	1,682
6	3,97	4,06	4,13	3,98	4,87	4,61	4,15	4,25	-4,75	-4,48	-4,62	4,134	3,365	3,488
7	4,65	4,55	4,50	4,30	5,47	5,35	4,73	4,79	-5,37	-4,89	-5,13	4,838	3,794	4,003
8	3,31	3,56	2,94	3,00	4,21	2,67	3,14	3,26	-4,64	-4,19	-4,41	2,848	3,682	3,289
9	3,41	3,53	3,08	3,00	4,47	4,07	3,70	3,61	-4,57	-4,24	-4,40	3,109	4,215	3,190
10	5,53	4,79	4,78	4,73	5,79	5,01	5,34	5,14	-5,30	-4,65	-4,98	5,265	4,318	4,351
11	3,13	3,22	3,16	3,17	3,33	3,58	2,99	3,23	-4,08	-4,08	-4,08	3,219	2,296	2,349
12	5,90	5,42	5,49	5,03	6,89	4,92	5,70	5,62	-5,80	-5,44	-5,62	5,718	5,048	4,543
13	5,56	4,93	5,11	4,73	5,88	4,18	5,13	5,07	-5,29	-5,03	-5,16	5,014	4,619	4,028
14	4,65	5,10	5,20	4,64	5,66	4,13	5,09	4,93	-5,45	-5,08	-5,27	4,828	5,127	4,640
15	3,91	3,95	4,31	3,89	5,48	4,80	4,55	4,41	-5,38	-4,98	-5,18	4,262	4,730	4,530
16	3,30	4,84	4,36	2,70	3,52	5,40	4,10	4,03	-5,58	-7,47	-6,52	3,415	3,535	3,468
17	4,78	4,80	5,36	4,76	6,07	4,86	5,30	5,13	-5,57	-5,80	-5,68	5,610	4,980	4,199
18	4,04	5,16	4,64	3,11	4,75	6,19	4,71	4,66	-5,87	-7,83	-6,85	4,304	3,456	3,371
19	4,98	4,96	4,70	4,01	5,45	4,16	5,27	4,67	-5,55	-5,14	-5,34	3,825	4,500	4,570
20	5,14	4,93	5,24	4,73	5,88	4,44	5,18	5,08	-5,34	-5,03	-5,19	5,014	4,764	4,202
21	5,75	5,42	5,62	5,03	6,89	5,18	5,75	5,66	-5,84	-5,44	-5,64	5,718	5,193	4,717
22	3,19	4,43	4,06	3,73	5,63	3,99	4,17	4,27	-5,14	-4,73	-4,94	3,727	5,081	4,002
23	4,29	4,41	4,20	3,73	5,48	4,12	4,73	4,42	-4,96	-4,79	-4,87	3,988	5,614	3,904
24	6,49	5,67	5,90	5,52	7,21	5,69	6,37	6,12	-6,03	-5,20	-5,62	6,145	5,717	5,065


Tabela P-7.3. Vrednosti deskriptora lipofilnosti analiziranih androstanskih derivata (*nastavak*)

Jedinjenje	$\log P_{\text{phys}}$	$\log P_{\text{wgt}}$	$\log D_{\text{vg}} (\text{pH} = 1.5)$	$\log D_{\text{klop}} (\text{pH} = 1.5)$	$\log D_{\text{phys}} (\text{pH} = 1.5)$	$\log D_{\text{wgt}} (\text{pH} = 1.5)$	* $T\log D_{\text{vg}} (\text{pH} = 1.5)$	$T\log D_{\text{klop}} (\text{pH} = 1.5)$	$T\log D_{\text{phys}} (\text{pH} = 1.5)$	$T\log D_{\text{wgt}} (\text{pH} = 1.5)$	$T\log D_{\text{vg}} (\text{pH} = 7.4)$	$T\log D_{\text{klop}} (\text{pH} = 7.4)$	$T\log D_{\text{phys}} (\text{pH} = 7.4)$	$T\log D_{\text{wgt}} (\text{pH} = 7.4)$
1	4,108	3,547	1,529	1,841	2,926	2,098	1,524	1,838	2,924	2,096	3,217	3,311	4,105	3,545
2	4,487	3,988	1,958	2,356	3,305	2,540	1,954	2,354	3,304	2,537	3,647	3,826	4,485	3,986
3	4,350	4,114	2,377	2,453	3,168	2,666	2,372	2,451	3,167	2,663	4,065	3,923	4,348	4,112
4	3,663	3,603	1,640	2,344	2,482	2,155	1,635	2,341	2,480	2,152	3,328	3,813	3,661	3,601
5	2,810	1,726	0,517	1,523	2,725	1,567	0,517	1,523	2,725	1,567	0,686	1,682	2,810	1,726
6	4,256	3,703	1,673	2,016	3,075	2,254	1,669	2,013	3,073	2,252	3,362	3,486	4,254	3,701
7	4,636	4,144	2,103	2,531	3,454	2,696	2,098	2,528	3,453	2,693	3,791	4,001	4,633	4,142
8	3,371	3,447	1,991	1,816	2,190	1,999	1,986	1,814	2,188	1,996	3,679	3,286	3,369	3,445
9	3,605	3,234	1,216	1,718	2,423	1,785	2,574	2,680	2,726	2,660	4,267	4,152	3,907	4,109
10	5,047	4,572	2,627	2,879	3,865	3,123	2,622	2,876	3,864	3,121	4,315	4,349	5,045	4,570
11	3,240	2,628	0,605	0,877	2,058	1,180	0,600	0,874	2,057	1,177	2,293	2,347	3,237	2,626
12	5,495	5,029	3,366	3,075	4,316	3,585	3,353	3,068	4,312	3,577	5,047	4,542	5,494	5,028
13	5,116	4,587	2,936	2,560	3,936	3,143	2,923	2,553	3,932	3,136	4,618	4,027	5,115	4,587
14	5,358	5,155	3,784	3,172	4,179	3,711	3,771	3,165	4,175	3,704	5,466	4,639	5,357	5,154
15	4,671	4,644	3,047	3,063	3,492	3,200	3,034	3,055	3,488	3,193	4,729	4,529	4,671	4,643
16	4,953	4,099	3,807	3,405	4,917	4,035	3,807	3,405	4,917	4,035	3,874	3,468	4,953	4,099
17	4,987	4,521	2,695	2,732	3,808	3,078	2,682	2,724	3,804	3,070	4,377	4,198	4,987	4,521
18	5,090	3,972	3,389	3,308	5,055	3,909	3,389	3,308	5,055	3,909	3,456	3,371	5,090	3,972
19	5,170	4,750	2,820	3,110	4,000	3,310	2,800	3,100	3,990	3,300	4,500	4,57	5,170	4,750
20	5,264	4,743	3,081	2,735	4,085	3,299	3,068	2,727	4,082	3,292	4,763	4,201	5,264	4,743
21	5,644	5,185	3,510	3,250	4,465	3,741	3,497	3,242	4,461	3,733	5,192	4,716	5,643	5,184
22	4,379	4,487	3,398	2,535	3,200	3,044	3,385	2,527	3,196	3,036	5,080	4,002	4,379	4,487
23	4,613	4,274	2,623	2,437	3,434	2,830	3,973	3,394	3,734	3,700	5,668	4,868	4,917	5,151
24	6,055	5,612	4,034	3,598	4,876	4,168	4,021	3,590	4,872	4,161	5,716	5,064	6,054	5,611

*oznaka T se odnosi na tautomeriju i rezonancu koje su uzete u obzir prilikom računanja deskriptora



Tabela P-7.4. Vrednosti fizičko-hemijskih, elektrostatičkih i topoloških deskriptora uključenih u linearne QSRR modele

Jedinjenje	AMR	BP	CP	CT	DE	FPSA3	Jhetv	PSA
1	113,277	756,448	17,361	931,361	105,49	0,040076	1,396	53,350
2	113,749	766,126	13,933	945,286	122,59	0,038386	1,285	59,420
3	126,747	752,841	17,019	939,840	102,11	0,036638	1,405	50,190
4	108,930	757,411	17,924	944,466	239,47	0,035049	1,258	62,720
5	111,248	792,156	19,036	970,227	259,75	0,034619	1,253	78,450
6	109,402	756,448	17,361	931,331	108,16	0,039746	1,392	53,350
7	111,720	766,126	13,933	945,286	125,03	0,040656	1,286	59,420
8	112,246	785,573	17,521	975,486	130,05	0,040546	1,433	67,260
9	122,439	805,355	18,724	988,887	118,61	0,046782	1,445	70,420
10	113,210	721,632	17,033	907,690	115,97	0,035043	1,428	33,120
11	112,312	792,788	18,247	970,683	116,25	0,047181	1,396	73,580
12	121,154	742,880	13,185	935,116	119,07	0,030774	1,302	39,190
13	121,000	733,203	16,325	918,979	102,13	0,031754	1,429	33,120
14	120,682	729,595	16,013	937,124	100,68	0,027605	1,439	29,960
15	114,837	734,166	16,838	941,261	212,93	0,025848	1,287	42,490
16	120,529	760,732	16,619	946,166	106,60	0,023122	1,436	42,530
17	110,979	805,254	14,218	987,398	139,04	0,033667	1,344	85,010
18	111,027	774,017	13,636	952,866	126,04	0,026738	1,307	51,760
19	117,199	752,256	14,490	950,236	122,01	0,029453	1,414	39,190
20	110,555	733,203	16,325	918,979	105,58	0,031444	1,424	33,120
21	112,003	742,880	13,185	935,116	122,09	0,033296	1,304	39,190
22	111,849	762,327	16,471	970,615	118,74	0,031813	1,466	47,030
23	111,531	782,109	17,565	974,414	108,12	0,038703	1,480	50,190
24	111,377	698,386	16,026	907,636	104,08	0,026559	1,466	12,890



Tabela P-7.4. Vrednosti fizičko-hemijskih, elektrostatičkih i topoloških deskriptora uključenih u linearne QSRR modele (*nastavak*)

Jedinjenje	MP	MR	S3K	SCAA3	TE	SASA1.4	vdWSA	vdWV
1	717,03	112,004	2,692	-0,025117	54,4316	502,53	620,27	384,04
2	721,08	121,155	3,330	-0,027530	57,3793	586,22	684,59	420,48
3	728,67	110,952	2,619	-0,028073	50,4061	499,88	609,22	377,11
4	786,60	109,328	2,229	-0,033178	98,3034	500,38	629,41	383,34
5	799,85	114,713	2,448	-0,041159	150,1971	510,23	654,77	399,33
6	717,03	111,850	2,692	-0,025214	48,6357	493,91	619,04	384,00
7	721,08	121,001	3,330	-0,026183	54,1804	582,65	684,24	420,58
8	796,89	111,645	2,644	-0,039105	53,9780	493,18	604,81	380,04
9	802,77	114,036	2,583	-0,034068	56,2258	499,56	589,97	378,79
10	661,21	111,451	2,450	-0,016579	55,0592	483,58	583,44	368,05
11	773,61	113,211	2,718	-0,033354	51,4516	505,21	627,61	393,07
12	650,96	120,608	3,340	-0,020164	62,6919	596,62	647,48	403,63
13	646,91	111,456	2,669	-0,016739	59,8418	508,76	582,00	366,89
14	658,55	110,405	2,594	-0,019712	45,4118	506,62	571,13	360,17
15	716,48	108,780	2,184	-0,024880	93,2276	514,00	590,94	365,45
16	683,44	114,387	2,699	-0,034591	87,8390	509,85	587,95	368,83
17	795,37	126,672	3,629	-0,038177	56,8336	597,42	685,91	426,66
18	675,85	124,590	3,433	-0,032535	107,4060	596,45	662,94	412,12
19	704,57	117,210	2,806	-0,018572	53,6355	563,21	618,52	386,89
20	646,91	111,302	2,669	-0,016834	49,2038	501,02	580,88	366,88
21	650,96	120,454	3,340	-0,018756	54,6561	589,40	646,11	403,45
22	726,77	111,098	2,625	-0,031362	49,3311	501,97	567,57	362,52
23	732,65	113,489	2,559	-0,026105	42,3962	503,93	551,42	361,08
24	591,09	110,903	2,429	-0,007519	54,0074	492,30	545,80	350,79



Formula za računanje *VIP* vrednosti

$$VIP_i = \sqrt{\frac{p}{\sum_{m=1}^M SS(b_m t_m)} \sum_{m=1}^M \omega_{mi}^2 SS(b_m t_m)} \quad (\text{P-7.5.})$$

p – broj nezavisno promenljivih varijabli

M – broj latentnih varijabli u PLS modelu

ω_{mi} – udeo i -te varijable u m -toj latentnoj varijabli

$SS(b_m t_m)$ – procenat podaka sadržanih u y -vrednostima objašnjen pomoću m -latentnih varijabli



BIOGRAFIJA



Strahinja Kovačević je rođen u Novom Sadu, 1988. godine. Diplomirao je na Tehnološkom fakultetu, Univerziteta u Novom Sadu, 2011. godine na studijskom programu Prehrambeno inženjerstvo – Kontrola kvaliteta sa prosečnom ocenom 9,41. Diplomске akademske studije – master na istom studijskom programu završava 2012. godine sa prosečnom ocenom 9,80 i iste godine upisuje doktorske akademske studije na smeru Prehrambeno inženjerstvo.

Zaposlen je na Tehnološkom fakultetu, Univerziteta u Novom Sadu, na Katedri za primenjene i inženjerske hemije, u zvanju istraživač-saradnik. Učesnik je projekta Ministarstva prosvete, nauke i tehnološkog razvoja „Pristupi održivosti i zelene hemije u razvoju ekološki pogodnih analitičkih metoda i skladištenju energije (OI 172012)“, kao i međunarodne COST akcije „Understanding Movement and Mechanism in Molecular Machines (CM 1306)“. Koautor je 47 naučnih publikacija, od kojih je 16 naučnih radova objavljenih u međunarodnim časopisima sa SCI liste kategorija M21, M22 i M23, a koji prezentuju primenu različitih hemometrijskih metoda u analizi rezultata iz oblasti analitičke hemije, prehrambenog i farmaceutskog inženjerstva, mikrobiologije i forenzike, uključujući i primenu hromatografskih metoda u analizi biološki aktivnih jedinjenja. Bio je učesnik tri međunarodne letnje škole instrumentalnih metoda analize održanih u Mariboru, na Fakultetu za hemiju i hemijsku tehnologiju (Univerzitet u Mariboru), Skoplju, na Tehnološko-metalurškom fakultetu (Univerzitet u Skoplju) i Sofiji, na Institutu za organsku hemiju sa centrom za fitohemiju, Bugarske akademije nauka.

Pored naučnog rada, angažovan je i u izvođenju nastave na osnovnim i master studijama na predmetima Opšta i neorganska hemija, Laboratorije za kontrolu kvaliteta i Hemometrija. Koautor je pomoćnog udžbeničkog materijala „Laboratorije za kontrolu kvaliteta – radna sveska sa elementima teorije“ za studente 4. godine smeru Kontrola kvaliteta na Tehnološkom fakultetu u Novom Sadu.

Član je Međunarodne unije za čistu i primenjenu hemiju (IUPAC), Međunarodne asocijacije inženjera (IAENG) i Srpskog hemijskog društva (SHD).



KLJUČNA DOKUMENTACIJSKA INFORMACIJA



Univerzitet u Novom Sadu
Tehnološki fakultet Novi Sad



Redni broj:

RBR

Identifikacioni broj:

IBR

Tip dokumentacije:

TD

Monografska publikacija

Tip zapisa:

TZ

Tekstualni štampani materijal

Vrsta rada:

VR

Doktorska disertacija

Autor:

AU

Strahinja Z. Kovačević, mast. inž. tehnol.

Mentor/ko-mentor:

MN

Dr Sanja Podunavac-Kuzmanović, redovni profesor

Naslov rada:

NR

**HEMOMETRIJSKO MODELOVANJE
HROMATOGRAFSKOG PONAŠANJA I
BIOLOŠKE AKTIVNOSTI SERIJE
ANDROSTANSKIH DERIVATA**

Jezik publikacije:

JP

Srpski (latinica)

Jezik izvoda:

Jl

Srpski/Engleski

Zemlja publikovanja:

ZP

Srbija

Uže geografsko područje: UGP	AP Vojvodina
Godina: GO	2015.
Izdavač: IZ	Autorski reprint
Mesto i adresa: MA	Bulevar cara Lazara 1, 21000 Novi Sad, Srbija
Fizički opis rada: FO	Broj poglavlja: 7 Broj strana: 162 Literaturnih citata: 200 Tabela: 30 Slika/grafikona: 64
Naučna oblast: NO	Tehnološko inženjerstvo
Naučna disciplina: ND	Primenjena hemija
Predmetna odrednica/ključne reči: PO	17-pikolil i 17-pikoliniliden androstanski derivati, Hemometrija, QSRR analiza, QSAR analiza, Hromatografska analiza, Antikancerogena aktivnost.
UDK	547.15:543.544+543.068(043.3)
Čuva se: ČU	U biblioteci Tehnološkog fakulteta u Novom Sadu, Bulevar cara Lazara 1, 21000 Novi Sad, Srbija
Važna napomena: VN	–
Izvod: IZ	
Steroidna jedinjenja spadaju u grupu supstanci sa širokim spektrom biološkog delovanja i predstavljaju dobru polaznu osnovu za sintezu mnogih derivata sa željenim biološkim potencijalom. Organskim sintezama se došlo do velikog broja steroidnih derivata, od kojih su neki pokazali značajnu biološku aktivnost, kao što je citotoksičnost prema različitim ćelijskim linijama kancera. Karakterizacija novosintetisanih jedinjenja može se izvesti	

eksperimentalnim i računarskim (*in silico*) metodama. U ovoj doktorskoj disertaciji predstavljeno je eksperimentalno određivanje lipofilnosti 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata primenom visokopritisne tečne hromatografije na obrnutim fazama, a potom hemometrijska analiza hromatografskog ponašanja (hromatografske lipofilnosti) QSRR pristupom. Hemijska struktura analiziranih derivata opisana je numerički, pomoću izračunatih molekulskih deskriptora. U drugom delu doktorske disertacija predstavljena je QSAR analiza citotoksične aktivnosti 17 α -pikolil i 17(E)-pikoliniliden androstanskih derivata prema ćelijama androgen-receptor negativnog kancera prostate (AR-neg. PC-3). Odabir najkvalitetnijih QSRR i QSAR modela obavljen je na osnovu izračunatih statističkih parametara, a njihovo rangiranje izvedeno je primenom metode sume razlika rangova (SRD). Pored regresionih QSRR i QSAR hemometrijskih metoda, primenjene su i klaster analiza i analiza glavnih komponenata sa ciljem utvrđivanja sličnosti ili razlika između analiziranih derivata na osnovu izračunatih molekulskih deskriptora.

Datum prihvatanja teme od strane Senata Univerziteta: **11.09.2014.**
DP

Datum odbrane: **DO**

Članovi komisije:
(naučni stepen/ime i prezime/zvanje/fakultet)
KO

Predsednik:
**Dr Slobodan Gadžurić, vanredni profesor,
Prirodno-matematički fakultet, Univerzitet u
Novom Sadu**

Član - Mentor:
**Dr Sanja Podunavac-Kuzmanović, redovni
profesor, Tehnološki fakultet Novi Sad,
Univerzitet u Novom Sadu**

Član:
**Dr Radomir Malbaša, redovni profesor,
Tehnološki fakultet Novi Sad, Univerzitet u
Novom Sadu**

Član:
**Dr Lidija Jevrić, vanredni profesor,
Tehnološki fakultet Novi Sad, Univerzitet u
Novom Sadu**

KEYWORDS DOCUMENTATION



University of Novi Sad
Faculty of Technology Novi Sad



Accession number:

ANO

Identification number:

INO

Document type:

DT

Monographic publication

Type of record:

TR

Textual material, printed

Contents code:

CC

Doctoral dissertation

Author:

AU

MSc Strahinja Z. Kovačević

Mentor/co-mentor:

MN

Dr. Sanja Podunavac Kuzmanović, full professor

Title:

TI

CHEMOMETRIC MODELING OF CHROMATOGRAPHIC BEHAVIOR AND BIOLOGICAL ACTIVITY OF A SERIES OF ANDROSTANE DERIVATIVES

Language of text:

LT

Serbian (Latin)

Language of abstract:

LA

Serbian/English

Country of publication:

CP

Serbia

Locality of publication:

LP

AP Vojvodina

Publication year: PY	2015
Publisher: PU	Author's reprint
Publication place: PP	Bulevar cara Lazara 1, 21000 Novi Sad, Serbia
Physical description: PD	Chapters: 7 Pages: 162 References: 200 Tables: 30 Figures/Graphs: 64
Scientific field: SF	Technological Engineering
Scientific discipline: SD	Applied Chemistry
Subject/Key words: SKW	17-picolyl and 17-picolinylidene androstane derivatives; Chemometrics; QSRR analysis; QSAR analysis; Chromatographic analysis; Anticancer activity.
UDC	547.15:543.544+543.068(043.3)
Holding data: HD	Library of the Faculty of Technology Novi Sad, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia
Note: N	-
Abstract: AB	
Steroidal compounds belong to the group of substances with wide spectrum of biological activity and represent the basic material for synthesis of many derivatives with preferred biological potential. A grate number of steroidal derivatives have been obtained through organic syntheses, many of which have demonstrated significant biological activity, such as cytotoxicity toward various cancer cell lines. Characterization of newly synthesized compounds can be achieved experimentally and by computational approach (<i>in silico</i>). This doctoral dissertation describes experimental determination of lipophilicity of 17α-picolyl and 17(<i>E</i>)-picolinylidene androstane derivatives applying reversed-phase high pressure	

liquid chromatography followed by quantitative structure-retention relationship (QSRR) chemometric analysis of chromatographic behaviour (chromatographic lipophilicity). Chemical structure of the analyzed derivatives was described numerically by in silico molecular descriptors. The second part of this dissertation describes quantitative structure-activity relationship (QSAR) analysis of cytotoxic activity of 17 α -picolyl and 17(E)-picolinylidene androstane derivatives toward androgen-receptor negative prostate cancer cell line (AR-neg. PC-3). Selection of the best QSRR and QSAR models was carried out based on their statistical parameters, and their ranking was done by sum of ranking differences (SRD) method. Besides the regression QSRR and QSAR chemometric methods, cluster analysis and principal components analysis were conducted in order to reveal possible similarities and dissimilarities among the studied derivatives on the basis of calculated molecular descriptors.

Accepted on Senate on:

11th September 2014

ASB

Defended:

DE

Thesis Defend Board:

(Degree/name/surname/title/faculty)

DB

President:

**Dr. Slobodan Gadžurić, associate professor,
Faculty of Sciences, University of Novi Sad**

Member - Mentor:

**Dr. Sanja Podunavac-Kuzmanović, full
professor, Faculty of Technology Novi Sad,
University of Novi Sad**

Member:

**Dr. Radomir Malbaša, full professor,
Faculty of Technology Novi Sad,
University of Novi Sad**

Member:

**Dr. Lidija Jevrić, associate professor,
Faculty of Technology Novi Sad,
University of Novi Sad**