



UNIVERZITET U KRAGUJEVCU
PRIRODNO-MATEMATIČKI FAKULTET

Izudin H. Redžepović

**KOMPARATIVNO ISPITIVANJE MOLEKULSKIH
DESKRIPTORA ZASNOVANIH NA SOPSTVENIM
VREDNOSTIMA**

doktorska disertacija

Kragujevac, 2021.



UNIVERSITY OF KRAGUJEVAC
FACULTY OF SCIENCE

Izudin H. Redžepović

**COMPARATIVE INVESTIGATION OF
EIGENVALUE-BASED MOLECULAR
DESCRIPTORS**

Doctoral Dissertation

Kragujevac, 2021.

Identifikaciona stranica doktorske disertacije

Autor	
Ime i prezime: Izudin Redžepović	
Datum i mesto rođenja: 26. 12. 1993. godine, Novi Pazar	
Sadašnje zaposlenje: Istraživač–saradnik na Prirodno–matematičkom fakultetu u Kragujevcu	
Doktorska disertacija	
Naslov: Komparativno ispitivanje molekulskih deskriptora zasnovanih na sopstvenim vrednostima	
Broj stranica: 78	
Broj slika: 38	
Broj bibliografskih podataka: 183	
Ustanova i mesto gde je rad izrađen: Prirodno–matematički fakultet, Kragujevac	
Naučna oblast (UDK): Fizička hemija (544)	
Mentor: Dr Boris Furtula, vanredni profesor, Prirodno–matematički fakultet, Univerzitet u Kragujevcu	
Ocena i odbrana	
Datum prijave teme: 24. 06. 2020. godine.	
Broj odluke i datum prihvatanja teme doktorske disertacije: IV-01-715/5, 14. 10. 2020. godine.	
Komisija za ocenu naučne zasnovanosti teme i ispunjenosti uslova kandidata: <ol style="list-style-type: none">1. Dr Boris Furtula, vanredni profesor, Prirodno–matematički fakultet, Univerzitet u Kragujevcu2. Dr Ivan Gutman, profesor emeritus i redovni član SANU, Prirodno–matematički fakultet, Univerzitet u Kragujevcu3. Dr Svetlana Marković, redovni profesor, Prirodno–matematički fakultet, Univerzitet u Kragujevcu4. Dr Slavko Radenković, vanredni profesor, Prirodno–matematički fakultet, Univerzitet u Kragujevcu5. Dr Biljana Arsić, naučni saradnik, Prirodno–matematički fakultet, Univerzitet u Nišu	
Komisija za ocenu i odbranu doktorske disertacije: <ol style="list-style-type: none">1. Dr Ivan Gutman, profesor emeritus i redovni član SANU, Prirodno–matematički fakultet, Univerzitet u Kragujevcu2. Dr Svetlana Marković, redovni profesor, Prirodno–matematički fakultet, Univerzitet u Kragujevcu3. Dr Biljana Arsić, viši naučni saradnik, Prirodno–matematički fakultet, Univerzitet u Nišu	
Datum odbrane disertacije:	



Rezultati predstavljeni u okviru ove doktorske disertacije ishod su istraživanja koja su rađena na Prirodno–matematičkom fakultetu u Kragujevcu u periodu od 2017. do 2021. godine. Ovom prilikom se zahvaljujem svom mentoru, prof. dr Borisu Furtuli, na iskrenoj i svesrdnoj pomoći tokom rada na disertaciji. Veliku zahvalnost dugujem i prof. dr Svetlani Marković. Na kraju, beskrajnu zahvalnost dugujem svojoj porodici na podršci koju mi pruža.

Autor

Sažetak

Molekulski deskriptori su brojevi ili nizovi brojeva koji se koriste za kvantifikovanje molekulske strukture. Posebna klasa molekulskih deskriptora su grafovske invarijante. Poznate su i kao topološki molekulski deskriptori. Izvođenje ovih deskriptora omogućeno je zamenom molekula molekulskim grafom. Mnoge korisne matematičke veličine mogu se izračunati iz molekulskog grafa, npr. sopstvene vrednosti. Stoga je postalo moguće konstruisati molekulske deskriptore koji se zasnivaju na sopstvenim vrednostima. Oni se nazivaju topološki molekulski deskriptori zasnovani na sopstvenim vrednostima. Danas ih ima mnoštvo. Samo nekoliko njih koristi sopstvene vrednosti dobijene iz klasične matrice susedstva. Među njima se ističu energija grafa, Estradin indeks i rezolventna energija. U okviru ove doktorske disertacije izvršeno je uporedno ispitivanje ovih deskriptora.

Prvi deo poglavlja Rezultati i diskusija izveštava o rezultatima u vezi sa istraživanjem relacija između energije grafa, Estradinog indeksa i rezolventne energije. Tri topološka molekulska deskriptora zasnovana na sopstvenim vrednostima upoređena su pomoću nekoliko skupova alkana i benzenoidnih ugljovodonika. Otkrivene su i diskutovane relacije među njima. Identifikovani su strukturni parametri koji upravljaju ovim odnosima i dobijene su odgovarajuće formule zasnovane na višestrukoj linearnoj regresiji. Pokazalo se da sva tri istražena indeksa kodiraju gotovo iste strukturne informacije o molekulu. Oni se razlikuju samo po stepenu osetljivosti na grananje molekula i po broju nevezivnih molekulskih orbitala.

Dalja analiza energije grafa, Estradinog indeksa i rezolventne energije vezana je za degenerativnost ovih deskriptora. Da bi se testirao diskriminativni potencijal ovih deskriptora, korišćeno je nekoliko klasa izomera hemijskih stabala. U ovim skupovima broj atoma ugljenika se kretao od 9 do 20. Kvantifikovanje degenerativnosti je urađeno pomoću dobro utvrđene veličine. Rezultati pokazuju da energija grafa i Estradin indeks imaju sličan nivo degenerativnosti. Nagla promena degenerativnosti rezolventne energije u slučaju hemijskih stabala zahtevala je dodatno ispitivanje. Dobijeni rezultati su pokazali da postoji mnogo hemijskih stabala sa istom rezolventnom energijom. Ona se zovu r -ekvienergetska hemijska stabla. Zatim su predstavljeni podaci vezani za pretraživanje rezolventnih ekvienergetskih hemijskih stabala.

Treći deo poglavlja Rezultati i diskusija donosi rezultate o strukturnoj osetljivosti energije grafa, Estradinog indeksa i rezolventne energije na nekoliko serija katakondenzovanih i perikondenzovanih izomernih benzenoidnih ugljovodonika. Strukturna osetljivost je jedno od najvažnijih i najmanje istraženih svojstava grafovskih invarijanti. Nedavno je predstavljena nova metoda za procenu strukturne osetljivosti topoloških molekulskih deskriptora. Ovaj algoritam se sastoji od nekoliko različitih koraka. Zasnovan je na Tanimoto indeksu i Morganovim kružnim fingerprintovima. Utvrđeno je da energija grafa, Estradin indeks i rezolventna energija imaju sličnu strukturnu osetljivost na katakondenzovane izomere. Energija grafa je najosetljivija na male promene u perikondenzovanim benzenoidnim ugljovodonicima. Pored toga, osetljivost ovih deskriptora testirana je na katakondenzovanim izomerima sa različitim brojem zaliva, uvala i fjordova. Otkriveno je da se vrednost ovih deskriptora postepeno menja sa postepenim povećanjem broja ovih strukturnih detalja. Estradin indeks i rezolventna energija se slično ponašaju, i u nekim slučajevima pokazuju istu strukturnu osetljivost. To se može pripisati visokoj korelaciji između njih.

U četvrtom delu poglavlja Rezultati i diskusija predstavljeni su rezultati ispitivanja uticaja cikla na vrednost energije grafa, Estradinog indeksa i rezolventne energije. Naime, pokazano

je da indeksi dobro opisuju fine strukturne detalje, te se može pretpostaviti da ukoliko znamo kako je deskriptor koreliran sa strukturom onda možemo da saznamo i kako osobine zavise od strukture. U cilju ispitivanja uticaja cikla na vrednost molekulskih deskriptora zasnovanih na sopstvenim vrednostima dizajnirana su tri *in silico* eksperimenta.

Poslednji deo ovog poglavlja predstavlja rezultate potencijalne hemijske primenljivosti naših deskriptora. Tačnije, ispitan je potencijal predviđanja fizičko–hemijskih osobina. Energija grafa, Estradin indeks i rezolventna energija testirani su kao oruđe za predviđanje tačke ključanja, toplote obrazovanja i koeficijenta raspodele oktanol/voda alkana. Pokazano je da se molekulski deskriptor zasnovan na sopstvenim vrednostima ne može pojedinačno koristiti za uspešno predviđanje ovih fizičko–hemijskih osobina. Prvi zagrebački indeks, broj nula u spektru i broj metil grupa, takođe, moraju biti uključeni u modele. Dobijene statističke veličine pokazuju da su modeli konstruisani pomoću Estradinog indeksa i rezolventne energije znatno bolji od modela sa energijom grafa. Takav trend je još izraženiji u slučaju koeficijenta raspodele oktanol/voda alkana.

Ključne reči: energija grafa, Estradin indeks, rezolventna energija, molekulska struktura, matrica susedstva, grafovske invarijante.

Abstract

Molecular descriptors are numbers or series of numbers used for quantification of molecular structure. A special class of molecular descriptors are graph invariants. They are also known as topological molecular descriptors. The derivation of these descriptors has been enabled by the substitution of molecule by a molecular graph. Many useful mathematical quantities may be calculated from a molecular graph, *e.g.*, eigenvalues. Therefore, it became possible to construct molecular descriptors that are using eigenvalues. These are called eigenvalue-based topological molecular descriptors. Today, there are plethora of them. Only few of them are using eigenvalues obtained from the ordinary adjacency matrix. The graph energy, Estrada index, and resolvent energy are the most prominent among them. Within this doctoral dissertation comparative investigation of these descriptors have been performed.

The first part of Results and discussion chapter reports results regarding investigation of relationships among graph energy, Estrada index, and resolvent energy. Three eigenvalue-based topological molecular descriptors are compared using several datasets of alkanes and benzenoid hydrocarbons. The relations among them are found and discussed. Structural parameters that govern these relations are identified and the corresponding formulae based on multiple linear regression have been obtained. It has been shown that all three investigated indices are encoding almost the same structural information of a molecule. They differ only by the extent of the sensitivity on a structural branching of a molecule and on the number of non-bonding molecular orbitals.

Further analysis of the graph energy, the Estrada index, and the resolvent energy is concerned with the degeneracy of these descriptors. To test discriminative potential of these descriptors, several classes of chemical-tree-isomers have been employed. In these sets number of carbon atoms ranged from 9 up to 20. The quantification of degeneracy has been done using well-established measure. The results show that graph energy and Estrada index exert similar degeneracy level. The specious degeneracy of the resolvent energy in the case of chemical trees is discussed. Obtained results indicated that there are many chemical trees with the same resolvent energy. These are called *r*-equienergetic chemical trees. Then, the results on searching for resolvent equienergetic chemical trees are given.

The third part of Results and discussion chapter brings results on structural sensitivity of the graph energy, the Estrada index, and the resolvent energy on several series of catacondensed and pericondensed isomeric benzenoid hydrocarbons. Structural sensitivity is one of the most important and the least investigated property of graph invariants. Recently, a novel method for assessing the structural sensitivity of topological molecular descriptors was put forward. This algorithm consists of several different steps. It is based on Tanimoto index and Morgan circular fingerprints. It was found that graph energy, Estrada index, and resolvent energy exert similar structural sensitivity on catacondensed isomers. The graph energy showed best performance on pericondensed molecules. Additionally, the sensitivities of these descriptors were tested on the catacondensed isomers with the increasing number of bays, coves, and fjords. It was revealed that these descriptors gradually change with the increasing number of these structural features. The Estrada index and resolvent energy perform similarly and in some cases with the same structural sensitivity. This may be attributed to the high correlation between them.

The fourth part of the Results and discussion chapter presents the results of the examination of the influence of the cycle on the value of the graph energy, the Estrada index, and the

resolvent energy. Namely, it has been shown that indices describe fine structural details well, so it can be assumed that if we know how the descriptor correlates with the structure then we can also find out how the properties depend on the structure. In order to examine the influence of a cycle on the value of molecular descriptors based on the eigenvalues, three *in silico* experiments were designed.

The last part of this chapter presents results of potential chemical applicability of our descriptors. More precisely, predictive potential of eigenvalue-based topological molecular descriptors was examined. The graph energy, the Estrada index, and the resolvent energy were tested as parameters for the prediction of boiling points, heats of formation, and octanol/water partition coefficients of alkanes. It was shown that an eigenvalue-based molecular descriptor cannot be individually used for successful prediction of these physico-chemical properties. The first Zagreb index, the number of zeros in the spectrum and the number of methyl groups must be also involved in the models. Performed statistics showed that the models constructed using the Estrada index and the resolvent energy are significantly better than ones with the graph energy. Such trend is even more noticeable in the case of octanol/water partition coefficients of alkanes.

Keywords: graph energy, Estrada index, resolvent energy, molecular structure, adjacency matrix, graph invariants.

Sadržaj

Spisak slika	ix
Spisak tabela	xi
1 OPŠTI DEO	1
1.1 Molekulski deskriptori	1
1.2 Topološki molekularni deskriptori	3
1.2.1 Hemijska teorija grafova	3
1.2.2 Sopstvene vrednosti matrice susedstva	4
1.2.2.1 Saksova teorema	5
1.2.3 Spektralni momenti grafa	6
1.2.4 Podela topoloških molekularnih deskriptora	6
1.2.4.1 Deskriptori zasnovani na rastojanju čvorova	7
1.2.4.2 Deskriptori zasnovani na stepenu čvora	8
1.3 Energija grafa, Estradin indeks i rezolventna energija	9
1.3.1 Hikelova molekularno-orbitalna (HMO) teorija	9
1.3.2 Energija grafa	10
1.3.3 Estradin indeks	11
1.3.4 Rezolventna energija	12
1.4 Strukturna osetljivost molekularnih deskriptora	13
1.5 Efekat cikla	15
1.5.1 Uticaj na energiju grafa	15
1.5.2 Uticaj na Estradin indeks	17
1.5.3 Uticaj na rezolventnu energiju	17
1.6 Predviđanje fizičko-hemijskih osobina jedinjenja	18
1.6.1 Metode za predviđanje osobina	18
1.6.2 QSPR	19
1.6.2.1 Višestruka linearna regresija	21
1.6.2.2 Valdov test	21
1.6.2.3 Validacija modela	22
1.6.2.4 Osnovne validacione veličine	22
2 CILJEVI DISERTACIJE	25
3 REZULTATI I DISKUSIJA	26
3.1 Veze između molekularnih deskriptora zasnovanih na sopstvenim vrednostima	27
3.1.1 Veza između energije grafa i rezolventne energije	27
3.1.2 Veza između Estradinog indeksa i rezolventne energije	29
3.1.3 Veza između energije grafa i Estradinog indeksa	33

3.2	Degenerativnost molekulskih deskriptora zasnovanih na sopstvenim vrednostima i r -ekvienergetska hemijska stabla	35
3.2.1	Degenerativnost	35
3.2.2	r -Ekvienergetska hemijska stabla	36
3.3	Strukturna osetljivost molekulskih deskriptora zasnovanih na sopstvenim vrednostima	41
3.3.1	Katakondenzovani i perikondenzovani benzenoidni ugljovodonici	41
3.3.2	Broj zaliva, uvala i fjordova	43
3.4	Efekat cikla na vrednost molekulskih deskriptora zasnovanih na sopstvenim vrednostima	46
3.4.1	Uticao veličine cikla	46
3.4.2	Uticao veličine molekula	48
3.4.3	Uticao tipa anelacije	50
3.5	Predikcioni potencijal molekulskih deskriptora zasnovanih na sopstvenim vrednostima	52
3.5.1	Tačka ključanja i entalpija obrazovanja	53
3.5.2	Koeficijent $\log P$	61
4	ZAKLJUČCI	65
	Literatura	67
	Indeks	78
A	Publikovani radovi u kojima su prezentovani rezultati doktorske disertacije	79
B	Biografija autora	87

Spisak slika

Slika 1.1	Podela molekulskih deskriptora prema dimenzionalnosti molekulske strukture iz koje se izvode	2
Slika 1.2	A–Strukturna formula acetilsalicilne kiseline; B–Molekulski graf acetilsalicilne kiseline; C–Matrica susedstva izvedena iz grafa acetilsalicilne kiseline	4
Slika 1.3	Saksovi grafovi izopropilciklopropana (označeni crvenom bojom) . . .	5
Slika 1.4	Konstrukcija linijskog grafa $L(G)$ iz grafa G	11
Slika 1.5	Energetski efekti ciklova (obojeni plavom bojom) fenantrena i bifeni- lena, izraženi u jedinicama HMO integrala rezonancije β	16
Slika 1.6	Generalna šema konstrukcije QSPR modela	20
Slika 1.7	Šematski prikaz validacije matematičkog predikcionog modela pomoću butstrap tehnike	24
Slika 3.1	Korelacija između energije grafa i rezolventne energije na primeru hemijskih stabala sa $n = 10$	27
Slika 3.2	Korelacija između rezolventne energije i modela zasnovanog na jednačini 3.1 za hemijska stabla sa $n = 10$	28
Slika 3.3	Korelacija između Estradinog indeksa i rezolventne energije na primeru hemijskih stabala sa $n = 10$	29
Slika 3.4	Zavisnost parametra α od broja čvorova kod hemijskih stabala	30
Slika 3.5	Korelacija između normalizovanog Estradinog indeksa i normalizovane rezolventne energije u grupi koja se sastoji od 5572 katakondenzovana benzenoidna ugljovodonika	32
Slika 3.6	Korelacija između normalizovanog Estradinog indeksa i normalizovane rezolventne energije u grupi koja se sastoji od 25 perikondenzovana benzenoidna ugljovodonika	32
Slika 3.7	Korelacija između Estradinog indeksa i energije grafa na primeru hemijskih stabala sa $n = 10$	33
Slika 3.8	Korelacija između energije grafa i modela zasnovanog na jednačini 3.8 za hemijska stabla sa $n = 10$	34
Slika 3.9	Promena procenta degenerativnosti/senzitivnosti energije grafa, Estradinog indeksa i rezolventne energije kod hemijskih stabala sa porastom broja čvorova	36
Slika 3.10	Neizospektralna hemijska stabla ($n = 19$) koja imaju istu vrednost rezolventne energije do 20 decimalnih mesta	39
Slika 3.11	Strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije kod katakondenzovanih molekula sa 10 heksagona. Plavi stubići predstavljaju \overline{SS} , dok zeleni stubići predstavljaju \overline{Abr}	42

Slika 3.12	Strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije kod perikondenzovanih molekula sa 10 heksagona. Plavi stubići predstavljaju \overline{SS} , dok zeleni stubići predstavljaju \overline{Abr}	43
Slika 3.13	Strukturni detalji katakondenzovanih benzenoidnih ugljovodonika . .	43
Slika 3.14	Strukturna senzitivnost E , EE i ER kod katakondenzovanih molekula sa različitim brojem zaliva (A), uvala (B) i fjordova (C). Plavi stubići predstavljaju \overline{SS} , dok zeleni stubići predstavljaju \overline{Abr}	44
Slika 3.15	Korelacija između E' , EE' i ER' i B , C i F	45
Slika 3.16	Grafik zavisnosti ef_{EE} od veličine cikla	47
Slika 3.17	Grafik zavisnosti ef_{ER} od veličine cikla	47
Slika 3.18	Grafik zavisnosti ef_E od veličine cikla	48
Slika 3.19	Zavisnost ef_{EE} od veličine molekula	49
Slika 3.20	Zavisnost ef_E od veličine molekula	49
Slika 3.21	Zavisnost ef_{ER} od veličine molekula	50
Slika 3.22	Tipovi anelacije benzenovog prstena	50
Slika 3.23	Benzenoidni ugljovodonik koji sadrži sve tipove prstenova prikazanih na slici 3.22	51
Slika 3.24	Korelacija između rezolventne energije i tački ključanja izomera oktana	53
Slika 3.25	Korelacija između rezolventne energije i entalpija obrazovanja izomera oktana	54
Slika 3.26	Korelacija između eksperimentalnih i izračunatih tački ključanja i entalpija obrazovanja izomera oktana	56
Slika 3.27	Distribucija vrednosti koeficijenata A , B , C i D iz modela 3.13 za tačku ključanja. Vertikalna linija označava srednju vrednost	60
Slika 3.28	Distribucija vrednosti koeficijenata A , B i C iz modela 3.12 za entalpiju obrazovanja. Vertikalna linija označava srednju vrednost	61
Slika 3.29	Korelacija između rezolventne energije i koeficijenta $\log P$ izomera oktana	62
Slika 3.30	Korelacija između eksperimentalnih vrednosti iz tabele 3.13 i izračunatih vrednosti koeficijenta $\log P$	63
Slika 3.31	Korelacija između eksperimentalnih vrednosti iz tabele 3.19 i izračunatih vrednosti koeficijenta $\log P$	64

Spisak tabela

Tabela 3.1	Vrednosti koeficijenata iz jednačine 3.1 za hemijska stabla sa $n = 6 - 20$. R i ARE predstavljaju koeficijent korelacije i prosečnu relativnu grešku, respektivno	28
Tabela 3.2	Vrednosti parametra α iz jednačine 3.5	30
Tabela 3.3	Koeficijenti korelacije (R) između Estradinog indeksa i rezolventne energije kod benzenoidnih ugljovodonika	31
Tabela 3.4	Vrednosti koeficijenata iz jednačine 3.8 za hemijska stabla sa $n = 6 - 20$	34
Tabela 3.5	Procenat degenerativnosti energije grafa, Estradinog indeksa i rezolventne energije	35
Tabela 3.6	Familije hemijskih stabala čije se rezolventne energije razlikuju za manje od 10^{-20} . # <i>familija</i> označava broj familija date veličine i tipa. <i>Tip a + b + ...</i> označava familiju koja se sastoji od a izospektralnih stabala, b izospektralna stabla nisu izospektralna sa stablima a , i tako dalje	37
Tabela 3.7	Strukturna osetljivost E , EE i ER kod katakondenzovanih i perikondenzovanih benzenoidnih ugljovodonika	41
Tabela 3.8	Strukturna senzitivnost E , EE i ER kod katakondenzovanih molekula sa različitim brojem zaliva, uvala i fjordova	44
Tabela 3.9	Koeficijenti korelacije između standardizovane energije grafa, standardizovanog Estrading indeksa, i standardizovane rezolventne energije i B , C i F	45
Tabela 3.10	Vrednosti efekta cikla na energiju grafa, Estradin indeks i rezolventnu energiju za ciklove veličine od 3 do 20 čvorova	46
Tabela 3.11	Efekat terminalnog benzenovog prstena na vrednosti energije grafa, Estradinog indeksa i rezolventne energije u klasi poliacena do 50 ugljenikovih atoma	48
Tabela 3.12	Vrednosti ef_E , ef_{EE} i ef_{ER} za sve tipove anelacije benzenovog prstena. Vrednosti za ef_{ER} su pomnožene sa 10^{12}	51
Tabela 3.13	Eksperimentalni podaci za izomere oktana	52
Tabela 3.14	Rezultati ispitivanja valjanosti fitovanja modela zasnovanih na jednačinama 3.11, 3.12 i 3.13	55
Tabela 3.15	Vrednosti koeficijenata za najbolje predikcione modele	55
Tabela 3.16	Eksperimentalni podaci korišćeni za butstrap ispitivanje predikcionih modela	57
Tabela 3.17	Prosečne vrednosti koeficijenata fitovanja i koeficijenata determinacije predikcionih modela	60
Tabela 3.18	Vrednosti koeficijenata fitovanja i koeficijenata determinacije predikcionih modela	62
Tabela 3.19	Eksperimentalne vrednosti koeficijenta $\log P$	63

1

OPŠTI DEO

1.1 Molekulski deskriptori

Ideja da informacije, koje su sadržane u molekulskoj strukturi, određuju fizičko–hemijske osobine datog jedinjenja zauzima značajno mesto u hemiji. Iako veza između molekulske strukture i neke osobine može da izgleda logično i jednostavno, neretko je takvu vezu teško identifikovati i objasniti. Ekstrakcija i manipulacija informacijama, izvedenim iz molekulske strukture, omogućena je upotrebom molekulskih deskriptora. Najčešće se koristi ova definicija za molekulski deskriptor [1]:

"Molekulski deskriptor je krajnji rezultat logičkih i matematičkih procedura koje transformišu hemijsku informaciju, koja je sadržana u simboličkoj reprezentaciji molekula, u koristan broj ili u rezultat standardizovanog eksperimenta".

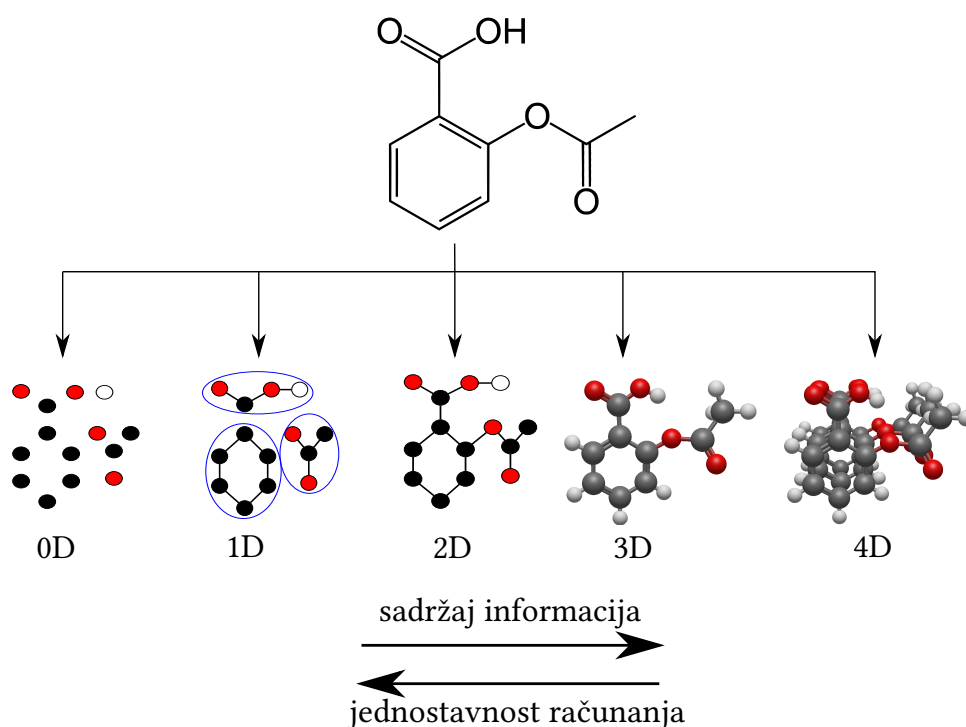
Danas je dostupan ogroman broj ovih deskriptora koji se koriste za različite svrhe. Oni se razlikuju po količini i kompleksnosti informacija koje nose, što utiče na vreme potrebno za njihovo računanje. Generalno, što je deskriptor bogatiji informacijama, to je potrebno više kompjuterskog vremena za njegovo računanje. Na primer, molekulska težina predstavlja jednostavan deskriptor koji se lako i brzo računa, ali pored podatka o masi molekula, on ne nudi dodatne informacije. S druge strane, molekulski deskriptori zasnovani na komplikovanim kvantno–mehaničkim proračunima zahtevaju dosta vremena za računanje, ali vrlo dobro opisuju dati molekul [2]. Koncept molekulskih deskriptora se zasniva na nekoliko različitih teorija. Neke od njih su kvantna hemija, organska hemija, teorija informacija i teorija grafova. Sve one se primenjuju kako bi se omogućilo modeliranje različitih osobina ili aktivnosti molekula, koje su od značaja u različitim naučnim disciplinama, kao npr. u toksikologiji [3], analitičkoj i fizičkoj hemiji [4], farmaciji [5] i ekologiji [6].

Molekulski deskriptori se mogu klasifikovati na različite načine. Najčešća je podela po dimenzionalnosti molekulske strukture iz koje se deskriptor izvodi [7–12], pa tako možemo izdvojiti:

- 0D deskriptore, koji su najjednostavniji i zasnivaju se na jednostavnom prebrojavanju u molekulu (npr. broj azotovih atoma u jedinjenju);
- 1D deskriptore, takođe zasnovani na prebrojavanju, koji npr. nose informaciju o broju određene funkcionalne grupe unutar molekula;

- 2D deskriptore, koji se određuju iz dvodimenzionalne strukture molekula;
- 3D deskriptore, koji se dobijaju iz trodimenzionalne strukture molekula i
- 4D deskriptore, koji se dobijaju kada pored prostornih informacija o molekulu, postoji još jedna dimenzija (npr. vreme) koja dodatno obogaćuje deskriptor podacima.

Na slici 1.1 je grafički prikazana podela molekulskih deskriptora na primeru acetilsalicilne kiseline.



Slika 1.1: Podela molekulskih deskriptora prema dimenzionalnosti molekulske strukture iz koje se izvode

Primene molekulskih deskriptora u hemiji su raznovrsne, a najizraženiji je njihov doprinos u pretraživanju hemijskih baza [13–15] i u konstruisanju matematičkih modela za predviđanje fizičko–hemijskih osobina i aktivnosti jedinjenja [16–18]. U literaturi se danas može naći veliki broj molekulskih deskriptora (preko 5000) [19]. Autori knjige "Recent advances in QSAR studies" [20] su definisali matematičke uslove koje molekulski deskriptor treba da ispuni, a to su:

- nezavisnost u odnosu na označavanje i numerisanje atoma;
- nezavisnost u odnosu na rotaciju i translaciju;
- jednostavna definicija i
- vrednosti u odgovarajućem opsegu za proučavani skup molekula.

Bez obzira na to kako su dobijeni i kakva je njihova matematička formulacija, većina molekulskih deskriptora je povezana. Na primer, promene u topologiji molekula mogu se povezati sa promenama u njegovoj geometriji. Zbog toga je poželjno proveriti korelaciju između dva deskriptora, odnosno, proveriti da li dva deskriptora nose iste informacije o molekulu.

1.2 Topološki molekularni deskriptori

Molekularni deskriptori zasnovani na topologiji molekula predstavljaju optimalan kompromis između količine korisnih informacija koje nose o strukturi jedinjenja i kompleksnosti njihovog izračunavanja. Ovakvi deskriptori se nazivaju topološki molekularni deskriptori ili grafovske invarijante [21,22]. Zbog jednostavnosti njihovog računanja i relativno jednostavne identifikacije veze struktura–osobina/aktivnost, oni zauzimaju značajno mesto u konstruisanju matematičkih predikcionih modela [23–28]. Takva vrsta modeliranja omogućila je hemičarima predviđanje fizičko–hemijskih osobina, toksičnosti i biološke aktivnosti jedinjenja koja još uvek nisu sintetisana [29,30]. Razumevanje veze između strukture i osobina/aktivnosti jedinjenja uticalo je na povećanje popularnosti topoloških molekularnih deskriptora. Da bi bolje upoznali topološke molekularne deskriptore, neophodno je da se prvo upoznamo sa teorijom grafova i njenom primenom u hemiji.

1.2.1 Hemijska teorija grafova

Teorija grafova [31,32] je pronašla značajnu primenu u hemiji kroz topološke molekularne deskriptore, formirajući pri tome čitavu novu naučnu disciplinu, hemijsku teoriju grafova [33–37]. Naime, u hemijskoj teoriji grafova molekul je predstavljen pomoću odgovarajućeg molekularnog grafa G u kojem čvorovi predstavljaju atome, a grane između njih predstavljaju hemijske veze. Ukupan broj čvorova u grafu se uobičajeno obeležava sa n , a broj grana sa m . Graf koji odgovara strukturalnoj formuli naziva se molekularni graf odgovarajućeg molekula. Primer molekularnog grafa, izvedenog iz molekula acetilsalicilne kiseline, prikazan je na slici 1.2 (B). U okviru molekularnih grafova možemo razlikovati dva tipa grafova, tj. plerograme i kenograme. Plerogram se konstruiše tako što se svaki atom prikazuje čvorom, dok se kod kenograma čvorom prikazuju svi atomi osim vodonikovih. Na taj način kenogram prikazuje samo skelet jedinjenja, i ova vrsta grafova se skoro uvek koristi kada se prikazuje molekularni graf nekog jedinjenja. U okviru ove disertacije pod molekularnim grafom podrazumevaćemo kenogram tog molekula.

Sada kada smo uveli pojam grafa, neophodno je da se upoznamo sa još nekim pojmovima vezanim za sam graf. Neki od tih pojmova su stepen čvora, cikl, stablo i rastojanje čvorova.

Definicija 1.2.1 Neka je v čvor grafa G . Broj čvorova koji su susedni čvoru v naziva se stepen tog čvora. Označavamo ga sa $d(v)$.

Na osnovu definicije 1.2.1 možemo zaključiti da, kod molekularnog grafa organskog jedinjenja, najveći stepen čvora ne sme biti veći od 4.

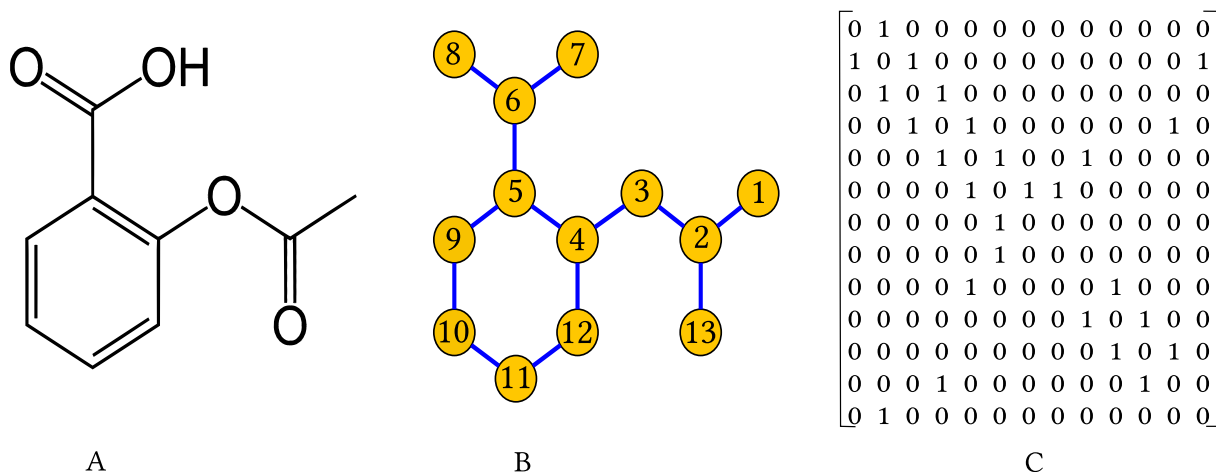
Definicija 1.2.2 Graf G , čiji su svi čvorovi stepena 2, naziva se cikl. Cikl sa n čvorova označava se sa C_n .

Definicija 1.2.3 Graf G bez ciklova naziva se stablo. Stablo sa n čvorova označava se sa T_n . Ukoliko stablo ne sadrži čvorove stepena većeg od 4, onda je to hemijsko stablo.

Definicija 1.2.4 Ako su čvorovi v_i i v_j grafa G povezani, tada je rastojanje od čvora v_i do čvora v_j , u oznaci $dist(v_i, v_j)$, jednako dužini najkraćeg puta između čvorova v_i i v_j .

Iz grafa G se može izvesti matrica susedstva $A(G)$. To je matrica tipa $n \times n$, koja nosi informacije o povezanosti čvorova, a koja je definisana na sledeći način:

$$[A(G)]_{i,j} = \begin{cases} 1 & \text{ako su čvorovi } v_i \text{ i } v_j \text{ povezani, } i \neq j; \\ 0 & \text{ako čvorovi } v_i \text{ i } v_j \text{ nisu povezani, } i \neq j; \\ 0 & \text{ako je } i = j. \end{cases} \quad (1.1)$$



Slika 1.2: A–Strukturna formula acetilsalicilne kiseline; B–Molekulski graf acetilsalicilne kiseline; C–Matrica susedstva izvedena iz grafa acetilsalicilne kiseline

Primer matrice susedstva, izvedene iz odgovarajućeg grafa, dat je na slici 1.2 (C). Pored ove matrice postoji još mnoštvo drugih matrica, kao npr. matrica rastojanja i Laplasova matrica [38]. Međutim, najširu i najznačajniju primenu je pronašla upravo matrica susedstva.

1.2.2 Sopstvene vrednosti matrice susedstva

Transformisanjem grafa u matricu susedstva omogućava se izračunavanje nekih korisnih veličina, koje nose važne informacije o grafu, tj. o molekulu. Tako iz matrice susedstva možemo dobiti karakteristični polinom, koji se može zapisati kao:

$$\phi(G, \lambda) = \det[\lambda I_n - A(G)] \quad (1.2)$$

gde je I_n jedinična matrica reda n , a $A(G)$ matrica susedstva. Karakteristični polinom se može zapisati i na sledeći način:

$$\phi(G, \lambda) = \sum_{k=0}^n a_k \lambda^{n-k} \quad (1.3)$$

odnosno, kao

$$\phi(G, \lambda) = a_0 \lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \dots + a_{n-1} \lambda + a_n \quad (1.4)$$

pri čemu je a_k njegov k -ti koeficijent, a $\lambda_1, \lambda_2, \dots, \lambda_n$ nule karakterističnog polinoma.

Definicija 1.2.5 Neka su $\lambda_1, \lambda_2, \dots, \lambda_n$ nule karakterističnog polinoma $\phi(G, \lambda)$ grafa G . Tada se $\lambda_1, \lambda_2, \dots, \lambda_n$ nazivaju sopstvene vrednosti grafa (tj. matrice susedstva), a sve sopstvene vrednosti zajedno čine spektar grafa, u oznaci $Sp(G)$.

Na primer, sopstvene vrednosti grafa prikazanog na slici 1.2 su:

$$\begin{aligned} \lambda_1 &= -2,307 \\ \lambda_2 &= -1,794 \\ \lambda_3 &= -1,583 \\ \lambda_4 &= -1,114 \\ \lambda_5 &= -0,845 \\ \lambda_6 &= 0 \end{aligned}$$

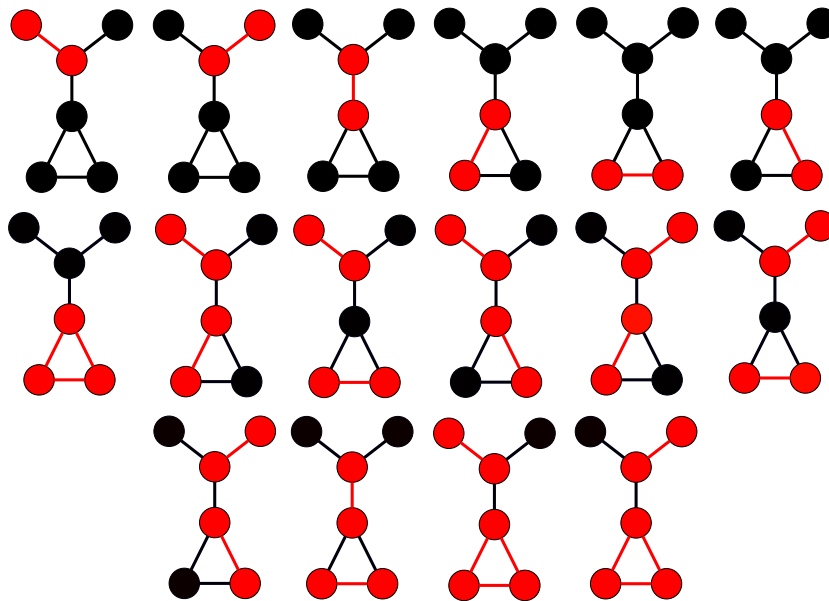
$$\begin{aligned}
\lambda_7 &= 0 \\
\lambda_8 &= 0 \\
\lambda_9 &= 0,845 \\
\lambda_{10} &= 1,114 \\
\lambda_{11} &= 1,583 \\
\lambda_{12} &= 1,794 \\
\lambda_{13} &= 2,307.
\end{aligned}$$

1.2.2.1 Saksova teorema

Kakva je veza između strukture grafa i njegovog karakterističnog polinoma? Odgovor na ovo pitanje dao je nemački matematičar Horst Saks [39]. Istorijski detalji vezani za Saksovu teoremu mogu se pronaći u radu [40]. Da bi predstavili Saksovu teoremu neophodno je da prvo razjasnimo šta su to put grafa i Saksovi grafovi.

Definicija 1.2.6 Stablo u kojem nijedan čvor nema stepen veći od 2 naziva se put. Put sa n čvorova označava se sa P_n .

Saksovi grafovi su grafovi čije su komponente P_2 i/ili C_3 i/ili C_4 i/ili C_5 itd. Odnosno, to su grafovi koji se sastoje od puta dužine 2 i/ili cikla. Saksova teorema zahteva da se u posmatranom grafu uoče svi Saksovi grafovi koji su u tom grafu sadržani. Konstrukcija svih Saksovih grafova kod iole većih grafova je veoma težak posao. Iz tog razloga je primena ove teoreme u hemiji ograničena [41]. Ilustracije radi, na slici 1.3 su prikazani svi Saksovi grafovi koji su sadržani u grafu izopropilciklopropana.



Slika 1.3: Saksovi grafovi izopropilciklopropana (označeni crvenom bojom)

Veza između strukture grafa i koeficijenata karakterističnog polinoma prikazana je u teoremi koja sledi.

Teorema 1.2.1 Ako karakteristični polinom grafa G ima oblik kao u formuli 1.3, onda je $a_0 = 1$, dok je za $k = 1, 2, \dots, n$,

$$a_k = \sum_S (-1)^{p(S)} 2^{c(S)} \quad (1.5)$$

gde $p(S)$ označava broj komponenti u Saksovom grafu S , a $c(S)$ broj ciklova. Sumiranje se vrši preko svih Saksovih grafova (S) sa k čvorova, sadržanih u grafu G .

Neposredna posledica Saksove teoreme je da za sve grafove važi:

$$a_1 = 0$$

$$a_2 = -m.$$

1.2.3 Spektralni momenti grafa

Iz sopstvenih vrednosti matrice susedstva mogu se izračunati veličine koje mogu biti od koristi u hemiji, a to su spektralni momenti. k -ti spektralni momenat grafa G , u oznaci $M_k(G)$, može se predstaviti kao [42]:

$$M_k(G) = \sum_{i=1}^n \lambda_i^k. \quad (1.6)$$

Spektralni momenti su uspešno primenjeni u hemiji čvrstog stanja [43–45] i u teorijskim ispitivanjima konjugovanih molekula [46–52]. Savremene primene spektralnih momenata odnose se na njihovu primenu na biološke sisteme, dizajn lekova i provodljivost molekula [53–55].

Odavno je poznato da je k -ti spektralni momenat jednak broju zatvorenih šetnji dužine k [42]. Na osnovu te činjenice su veze između spektralnih momenata za $k = 0, 2$ i 4 i nekih drugih grafovskih invarijanti već razjašnjene [56]:

$$M_0(T) = n \quad (1.7)$$

$$M_2(T) = 2 \cdot (n - 1) \quad (1.8)$$

$$M_4(T) = 2 \cdot (Zg_1(T) - n + 1) \quad (1.9)$$

gde je n broj čvorova stabla T , a Zg_1 je prvi zagrebački indeks.

1.2.4 Podela topoloških molekulskih deskriptora

Topološki molekulski deskriptori se mogu klasifikovati na osnovu parametara grafa koji se javljaju u njihovim definicijama. Na osnovu toga možemo izdvojiti osnovne grupe ovih topoloških indeksa:

- indeksi zasnovani na rastojanju čvorova;
- indeksi zasnovani na stepenu čvora i
- indeksi zasnovani na sopstvenim vrednostima neke od matrica pridruženih grafu.

Usled pojave sve većeg broja topoloških molekulskih deskriptora, od kojih neki nisu od hemijskog i matematičkog značaja, javila se potreba za uvođenjem određenih uslova koje bi indeks trebalo da ispuni kako bi mogao da se uvede. Značajnu listu uslova predložio je Milan Randić [57]. Ova lista se sastoji od sledećih uslova:

1. mogućnost neposredne strukturne interpretacije;
2. korelacija sa makar jednom fizičko–hemijском osobinom;
3. mogućnost razlikovanja izomera;
4. lokalna definisanost;
5. mogućnost uvođenja deskriptora višeg reda;
6. linearna nezavisnost;
7. jednostavnost;
8. nezasnovanost na fizičko–hemijским osobinama;
9. trivijalna nepovezanost sa drugim deskriptorima;
10. mogućnost efikasnog izračunavanja;
11. definisanost na osnovu razumljivih strukturnih detalja;
12. korektna zavisnost od veličine molekula i
13. postojanje višeg nivoa strukturne osetljivosti.

Iako se ova lista sastoji od razumnih uslova, većina topoloških indeksa ne ispunjava sve uslove. Nešto kraći spisak uslova predložio je Balaban [58].

1.2.4.1 Deskriptori zasnovani na rastojanju čvorova

Prvi topološki molekularni deskriptor je tzv. Vinerov indeks (W) koji je zasnovan na rastojanju čvorova [59]. Viner je želeo da modelira tačke ključanja alkana i bilo je neophodno da na neki način kvantifikuje molekularnu strukturu. Zbog toga je u svom modelu definisao parametar koji meri rastojanje između čvorova, koji je kasnije nazvan Vinerov indeks. On je definisan na sledeći način:

$$W(G) = \sum_{v_i, v_j \in V(G)} dist(v_i, v_j) \quad (1.10)$$

gde je $V(G)$ skup čvorova grafa G . Odnosno, ovaj indeks predstavlja zbir rastojanja između svih parova čvorova u grafu. Vinerov indeks se može izraziti i kao:

$$W(G) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij} \quad (1.11)$$

gde D predstavlja matricu rastojanja.

Vinerov indeks je stekao veliku popularnost i upotrebljen je za modeliranje različitih fizičko–hemijških osobina jedinjenja [60–63]. Uvođenjem Štajnerovog rastojanja (\bar{S}) [64] u jednačinu 1.10, dobija se generalizovani oblik Vinerovog indeksa [65]:

$$\bar{S}W_k(G) = \sum_{\substack{\bar{S} \subseteq V(G) \\ |\bar{S}|=k}} dist(\bar{S}). \quad (1.12)$$

Pored Vinerovog indeksa, postoji još mnoštvo topoloških molekularnih deskriptora koji su zasnovani na rastojanju čvorova. Među popularnijim su Balabanov indeks [66] i Hararijev indeks [67].

1.2.4.2 Deskriptori zasnovani na stepenu čvora

Veoma značajnu grupu topoloških molekulskih deskriptora čine deskriptori zasnovani na stepenu čvora [68, 69]. Ovoj grupi pripadaju čuveni zagrebački indeksi (prvi Zg_1 i drugi Zg_2). Oni su definisani na sledeći način:

$$Zg_1(G) = \sum_{v \in V(G)} d(v)^2 \quad (1.13)$$

$$Zg_2(G) = \sum_{v_i, v_j \in E(G)} [d(v_i) \cdot d(v_j)] \quad (1.14)$$

gde $E(G)$ predstavlja skup grana grafa G [70, 71]. Ubrzo po njihovom pojavljivanju, ovi indeksi su stekli veliku popularnost kako u hemijskim tako i u matematičkim istraživanjima [72].

Jedan od najpopularnijih topoloških molekulskih indeksa pripada baš ovoj grupi – Randićev indeks (R) [73–75]. On je definisan na sledeći način:

$$R(G) = \sum_{v_i, v_j \in E(G)} \frac{1}{\sqrt{d(v_i) \cdot d(v_j)}}. \quad (1.15)$$

U sledećem poglavlju biće predstavljeni topološki molekulski deskriptori zasnovani na sopstvenim vrednostima matrice susedstva.

1.3 Energija grafa, Estradin indeks i rezolventna energija

Hemijski smisao sopstvenih vrednosti matrice susedstva dat je u molekulsko-orbitalnoj teoriji Erika Hikel. Ovo je uticalo na popularnost topoloških molekulskih deskriptora koji u svojim definicijama koriste ove sopstvene vrednosti. Indeksi koji su zasnovani na sopstvenim vrednostima matrice susedstva su i energija grafa, Estradin indeks i rezolventna energija. Da bi pratili razvoj ovih indeksa, predstavice prvo čuvenu Hikelovu molekulsko-orbitalnu teoriju.

1.3.1 Hikelova molekulsko-orbitalna (HMO) teorija

Kvantno-mehanički opis nekog jedinjenja zahteva rešavanje Šredingerove jednačine koja ima oblik:

$$\hat{H}\Psi = \mathcal{E}\Psi$$

gde je \hat{H} Hamiltonov operator primenjen na talasnu funkciju Ψ , a \mathcal{E} je energija posmatranog sistema. Hamiltonov operator je poznat za svako jedinjenje, a zadatak se svodi na pronalaženje onih vrednosti \mathcal{E} i funkcija Ψ koje zadovoljavaju datu jednačinu. To zahteva obimna numerička izračunavanja, zbog čega je Hikel umesto Hamiltonovog operatora upotrebio aproksimaciju. Važno je naglasiti da je HMO teorija ograničena na konjugovane π -elektronske sisteme. U HMO teoriji [76–79] se umesto Hamiltonovog operatora koristi njegova matična reprezentacija, tj. kvadratna matrica H . Elementi te matrice su definisani na sledeći način:

$$[H]_{i,j} = \begin{cases} \alpha & \text{ako je } i = j; \\ \beta & \text{ako je } i \neq j \text{ i ako postoji hemijska veza između atoma } i \text{ i } j; \\ 0 & \text{ako je } i \neq j \text{ i ako atomi } i \text{ i } j \text{ nisu hemijski vezani.} \end{cases} \quad (1.16)$$

U kvantnoj hemiji se α naziva Kulonov integral, a β integral rezonancije i njihove numeričke vrednosti se određuju tako da se postigne što bolja saglasnost između izračunatih i eksperimentalno određenih osobina jedinjenja.

Pored matrice H u ovoj teoriji je od značaja i matrica preklapanja S , koja je u HMO teoriji definisana na sledeći način:

$$[S]_{i,j} = \begin{cases} 1 & \text{ako je } i = j; \\ 0 & \text{ako je } i \neq j. \end{cases} \quad (1.17)$$

Na osnovu ove definicije vidimo da se matrica preklapanja identifikuje sa jediničnom matricom. Kada su matrice H i S konstruisane, onda se konstruiše sekularna jednačina koja ima oblik:

$$\det(\mathcal{E}S - H) = 0 \quad (1.18)$$

koja se rešava po \mathcal{E} , pri čemu se dobija $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n$ rešenja koja predstavljaju energije molekulskih orbitala.

Veza između HMO Hamiltonovog operatora (H) i matrice susedstva molekuskog grafa data je sledećom relacijom [80]:

$$H = \alpha I_n + \beta A(G) \quad (1.19)$$

gde je I_n jedinična matrica. Kada jednačinu 1.19 primenimo na jednačinu 1.18 dobijamo:

$$\begin{aligned} \det(\mathcal{E}S - H) &= \det[\mathcal{E}I_n - (\alpha I_n + \beta A(G))] = \det[(\mathcal{E} - \alpha)I_n - \beta A(G)] \\ &= \beta^n \det\left(\frac{\mathcal{E} - \alpha}{\beta} I_n - A(G)\right) = \beta^n \det[\lambda I_n - A(G)] = \beta^n \phi(G, \lambda) \end{aligned}$$

gde smo uveli smenu

$$\lambda = \frac{\mathcal{E} - \alpha}{\beta}.$$

Vidimo da je uslov $\det(\mathcal{E}S - H) = 0$ ekvivalentan uslovu $\phi(G, \lambda) = 0$. Sa druge strane, rešenja jednačine $\phi(G, \lambda) = 0$ su upravo sopstvene vrednosti grafa G . Zaključujemo da između sopstvenih vrednosti molekulskog grafa G i π -elektronskih energetske nivoa postoji veza:

$$\lambda_i = \frac{\mathcal{E}_i - \alpha}{\beta}. \quad (1.20)$$

Ova veza je od velikog značaja jer pokazuje da se na osnovu sopstvenih vrednosti grafa može izračunati energija sistema na nivou HMO teorije.

1.3.2 Energija grafa

Ukupna π -elektronska energija se u HMO teoriji dobija kada se saberu energije pojedinačnih π -elektrona. Proučavajući ukupnu π -elektronsku energiju, Ivan Gutman je predložio generalizovani oblik HMO ukupne π -elektronske energije:

$$E(G) = \sum_{i=1}^n |\lambda_i| \quad (1.21)$$

gde je λ_i i -ta sopstvena vrednost grafa G . Ova veličina je nazvana energija grafa i uvedena je 1978. godine [81]. Neki od razloga za uvođenje ove veličine su sledeći:

- Energija grafa je definisana za sve grafove tj. nije ograničena samo na molekulske grafove;
- Energija grafa predstavlja simetričnu funkciju, a u matematici postoje razvijene teorije za proučavanje ovakvih funkcija;
- Energija grafa se može koristiti i kao molekulski deskriptor tj. nije ograničena samo na HMO teoriju.

Dokaz jednačine 1.21

U HMO teoriji π -elektronska energija se računa kao:

$$E_\pi = 2 \sum_+ \lambda_i$$

gde \sum_+ označava sumiranje preko pozitivnih sopstvenih vrednosti. Za sve grafove važi da je zbir pozitivnih i negativnih sopstvenih vrednosti jednak nuli, tj:

$$\sum_+ \lambda_i + \sum_- \lambda_i = 0$$

odnosno,

$$\sum_{+} \lambda_i = - \sum_{-} \lambda_i = \sum_{-} |\lambda_i|.$$

Takođe je

$$\sum_{+} \lambda_i = \sum_{+} |\lambda_i|.$$

Onda je:

$$\begin{aligned} 2 \sum_{+} \lambda_i &= \sum_{+} \lambda_i + \sum_{+} \lambda_i = \sum_{+} \lambda_i - \sum_{-} \lambda_i \\ &= \sum_{+} |\lambda_i| + \sum_{-} |\lambda_i| = \sum_{i=1}^n |\lambda_i|. \end{aligned}$$

Ovim je dokaz završen.

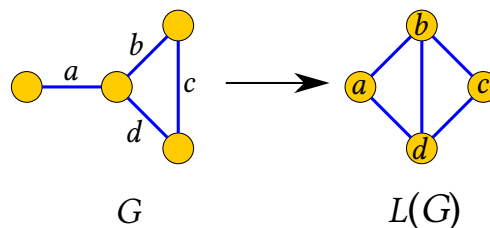
Energija grafa je danas veoma popularna, posebno u matematičkim istraživanjima [82–88]. Definicija 1.21 je poslužila kao kalup za uvođenje različitih grafovskih invarijanti zasnovanih na sopstvenim vrednostima. Takođe, neki od tih indeksa su zasnovani na sopstvenim vrednostima matrice različitih od matrice susedstva. To je dovelo do "poplave" različitih "energija" [89–91]. Energija grafa je odavno proširila primenu, pa se osim u hemiji i matematici, u novijoj literaturi spominje primena energije za obradu podataka [92], procenu kvaliteta [93], praćenje širenja kompjuterskih virusa [94] i prepoznavanje lica [95].

1.3.3 Estradin indeks

Pored energije grafa, koja predstavlja najpopularniji deskriptor zasnovan na sopstvenim vrednostima matrice susedstva, možemo slobodno reći da drugo mesto zauzima Estradin indeks (EE). Ovaj indeks se prvi put pojavljuje 2000. godine u radu Ernesta Estrade [96] i definisan je na sledeći način:

$$EE(G) = \sum_{i=1}^n e^{\lambda_i}. \quad (1.22)$$

Naime, Estrada je želeo da modelira uvijanje u proteinima, što je zahtevalo kvantifikovanje 3D molekulske strukture. On je uočio da se korišćenjem linijskih grafova [31] može izvesti veličina 1.22 koja u dobroj meri opisuje uvijanje proteina, a s druge strane je jednostavno definisana [97, 98]. Linijski graf, u oznaci $L(G)$, se dobija iz grafa G tako što grane iz grafa G postaju čvorovi u linijskom grafu $L(G)$. Primer konstrukcije linijskog grafa je prikazan na slici 1.4.



Slika 1.4: Konstrukcija linijskog grafa $L(G)$ iz grafa G

Estradin indeks se može izraziti preko spektralnih momenata [99]:

$$EE(G) = \sum_{k \geq 0} \frac{M_k(G)}{k!} \quad (1.23)$$

gde je M_k k -ti spektralni moment grafa G . Pored primene Estradinog indeksa za opisivanje uvijanja makromolekula, on se pokazao kao veoma korisna veličina za kvantifikovanje centralnosti mreže [100, 101]. Centralnost mreže (društvene, biološke, transportne, itd.) je bitna osobina bilo koje vrste mreže [102, 103] koja nam govori o različitoj važnosti čvorova unutar mreže.

Za kratko vreme je ovaj topološki molekularski deskriptor stekao veliku popularnost, posebno u hemijskoj i matematičkoj zajednici [99, 104–109] i na osnovu ovog indeksa uvedeni su slični novi indeksi [110–115].

1.3.4 Rezolventna energija

Još jedan topološki molekularski deskriptor koji je zasnovan na sopstvenim vrednostima matrice susedstva je rezolventna energija (ER). Ona je uvedena 2016. godine [116] od strane Ivana Gutmana i njegovih saradnika. Naziv "rezolventna" potiče od rezolventne matrice (R_M) [117], koja je definisana na sledeći način:

$$R_M = (zI_n - M)^{-1} \quad (1.24)$$

gde je I_n jedinična matrica reda n , a z je kompleksna promenljiva. Iz jednačine 1.24 se lako može uočiti da R_M postoji u svim slučajevima osim kada se vrednosti za promenljivu z poklapaju sa sopstvenim vrednostima matrice M .

Rezolventna matrica matrice susedstva je:

$$R_A = (zI_n - A)^{-1} \quad (1.25)$$

čije sopstvene vrednosti su:

$$\frac{1}{z - \lambda_i}, \quad i = 1, 2, 3, \dots, n. \quad (1.26)$$

Pošto je energija grafa definisana kao suma apsolutnih vrednosti sopstvenih vrednosti matrice susedstva, autori [116] su došli na ideju da definišu rezolventu energiju kao:

$$\sum_{i=1}^n \frac{1}{z - \lambda_i}.$$

Međutim, ona bi zavisila od promenljive z i ne bi bila definisana kada je $z = \lambda_i$. Zbog toga je umesto z u definiciju rezolventne energije ubačen broj čvorova n :

$$ER(G) = \sum_{i=1}^n \frac{1}{n - \lambda_i}. \quad (1.27)$$

Kao i Estradin indeks i rezolventna energija se može izraziti preko spektralnih momenata:

$$ER(G) = \frac{1}{n} \sum_{k=0}^{\infty} \frac{M_k(G)}{n^k}. \quad (1.28)$$

Ovo je nov indeks čije osobine i primenljivost u hemiji (eventualno i u drugim disciplinama) tek treba da se ispituju. Tako je rezolventna energija već predmet nekih matematičkih istraživanja [118–122].

1.4 Strukturna osetljivost molekulskih deskriptora

Jedna od najbitnijih osobina topološkog molekulskog deskriptora, navedenih u odeljku 1.2.4, jeste njegova strukturna osetljivost. Naime, deskriptor treba da postepeno menja svoju vrednost sa malim promenama u molekulskoj strukturi. Da bi se ovo desilo potrebno je da molekulski deskriptor ima visok nivo strukturne osetljivosti. Zato što je teško definisati šta je zapravo mala strukturna promena za različite skupove molekula, pojam strukturne osetljivosti postaje promenljiva kategorija koju je teško kvantifikovati.

Verovatno, iz ovog razloga strukturna osetljivost topoloških molekulskih deskriptora je veoma slabo proučavana. Tek su u radu Borisa Furtule i saradnika 2013. godine [123] prvi put definisane veličine za kvantifikovanje ove osobine. U ovom algoritmu je mala strukturna promena definisana kao strukturna razlika između dva molekulska grafa, sa istim brojem čvorova i grana, koji imaju minimalne vrednosti veličine koja se zove *graph edit distance* (GED) [124]. GED je zapravo broj osnovnih strukturnih promena koje se moraju izvršiti nad jednim grafom kako bi se dobio drugi graf. Tako da, ako dva grafa imaju isti broj čvorova i grana onda je minimalna vrednost GED jednaka 2.

Osnovna mana predloženog algoritma jeste upravo generisanje grafova koristeći GED, što je kompjuterski veoma zahtevan posao, posebno za veće grafove. Zbog toga je nedavno predložen novi algoritam za određivanje strukturne osetljivosti topološkog molekulskog deskriptora [125]. U ovoj unapređenoj metodi otklonjene su mane koje su postojale u prethodnom. Ovaj algoritam se sastoji od nekoliko zasebnih koraka.

ALGORITAM

1. Generisanje baze izomera i vrednosti topološkog molekulskog deskriptora, čija se strukturna osetljivost ispituje.
2. Za svaki par molekula iz baze izračunavanje Tanimoto indeksa zasnovanog na tzv. Morganovim kružnim fingerprintovima.
3. Za svaki molekul u bazi pravljenje grupe sličnih molekula na osnovu vrednosti Tanimoto indeksa.
4. Svaka grupa molekula podvrgnuta određenim numeričkim proračunima koji će zauzvrat dati vrednosti strukturne senzitivnosti (SS) i robustnosti (Abr) molekulskog deskriptora.
5. Na kraju, ukupna strukturna senzitivnost (\overline{SS}) i ukupna robustnost (\overline{Abr}) se dobijaju kao prosečne vrednosti SS i Abr izračunatih u koraku 4.

Ograničavanjem baze molekula na izomere sa istim brojem veza, tipova veza i atoma, u prvom koraku algoritma, eliminiše se uticaj ovih strukturnih parametara. Na ovaj način je

omogućena promena vrednosti topološkog indeksa samo na osnovu uticaja malih strukturnih razlika. Naravno, kada je potrebno ispitati uticaj vrste atoma na promenu vrednosti topološkog indeksa, onda ovakvo ograničenje treba zaobići. U drugom koraku ovog algoritma potrebno je izračunati Tanimoto indeks zasnovan na Morganovim kružnim fingerprintovima [126]. Naime, Tanimoto indeks (jednačina 1.29) pokazuje kolika je sličnost između dva molekula. Da bi se izračunao neophodno je da parametar na osnovu koga se određuje sličnost bude dat u binarnom obliku, gde bi 1 označavala prisustvo određene karakteristike kod oba molekula, a 0 odsustvo takvog podudaranja. Tada se Tanimoto indeks (T) za molekule A i B može zapisati u sledećem obliku:

$$T_{A,B} = \frac{c}{a + b - c} \quad (1.29)$$

gde a predstavlja broj bitova 1 u molekulu A, b broj bitova 1 u molekulu B, a c predstavlja broj bitova 1 zajedničkih za A i B. Tanimoto indeks u ovom slučaju može imati vrednost od 0 do 1. U okviru ovog algoritma se za računanje Tanimoto indeksa koriste Morganovi kružni fingerprintovi [127]. Sledeći korak algoritma podrazumeva formiranje podskupova sličnih molekula, za svaki molekul pojedinačno, na osnovu vrednosti Tanimoto indeksa. U koraku 4 se računaju strukturna senzitivnost (SS) i robustnost (Abr). Ove dve veličine su definisane sledećim izrazima:

$$SS(TI, m_r) = \sqrt{\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} (TI(m) - TI(m_r))^2}, \quad (1.30)$$

$$Abr(TI, m_r) = \max_{m \in \mathcal{M}} |(TI(m) - TI(m_r))|. \quad (1.31)$$

U jednačinama 1.30 i 1.31 m_r označava referenti molekul, a $|\mathcal{M}|$ broj molekula u podskupu. Iako na prvi pogled veličina SS u potpunosti određuje strukturnu senzitivnost topološkog indeksa, ona ne uzima u obzir postepene promene vrednosti indeksa. Zbog toga je pored strukturne senzitivnosti neophodno izračunati i njegovu robustnost (jednačina 1.31). Na kraju, ukupna senzitivnost (\overline{SS}) i ukupna robustnost (\overline{Abr}) se dobijaju kao prosečne vrednosti veličina izračunatih u koraku 4:

$$\overline{SS}(TI) = \frac{\sum_{m_r \in \mathcal{M}} SS(TI, m_r)}{|\mathcal{M}| \cdot \left(\max_{m_r \in \mathcal{M}} TI(m_r) - \min_{m_r \in \mathcal{M}} TI(m_r) \right)}, \quad (1.32)$$

$$\overline{Abr}(TI) = \frac{\sum_{m_r \in \mathcal{M}} Abr(TI, m_r)}{|\mathcal{M}| \cdot \left(\max_{m_r \in \mathcal{M}} TI(m_r) - \min_{m_r \in \mathcal{M}} TI(m_r) \right)}. \quad (1.33)$$

Na osnovu ovih definicija vidimo da je \overline{Abr} uvek veće od \overline{SS} . Topološki molekularni deskriptor sa visokom strukturnom osetljivošću bi trebalo da ima što veću vrednost \overline{SS} i što je moguće nižu vrednost \overline{Abr} .

1.5 Efekat cikla

1.5.1 Uticaj na energiju grafa

Odavno su hemičari uočili da su ciklični π -elektronski sistemi znatno stabilniji ili znatno nestabilniji u poređenju sa svojim acikličnim analogima. Parovi benzen–heksatrien (stabilizacija) i ciklobutadien–butadien (destabilizacija) su upravo pravi primeri. Još u tridesetim godinama XX veka Hikel je formulisao svoje $4n + 2$ pravilo, koje tvrdi da su monociklični konjugovani sistemi stabilniji ako poseduju $4n + 2$ (tj. 2, 6, 10, 14, ...) π -elektrona, a nestabilni ako je broj π -elektrona $4n$ (tj. 4, 8, 12, ...). U šezdesetim godinama se pokazalo da se ovde radi o jednom energetskom efektu [128].

Zahvaljujući srećnoj okolnosti da ukupna π -elektronska energija zavisi od strukturnih detalja koji odgovaraju Saksovim grafovima, i da neki Saksovi grafovi sadrže ciklove, postalo je moguće izraziti energetski efekat svakog pojedinačnog cikla C sadržanog u molekulskom grafu G [129–132]. Veza između ukupne π -elektronske energije i molekulske strukture data je sledećim izrazom:

$$E = \frac{1}{\pi} \int_{-\infty}^{+\infty} x^{-2} \ln |1 + \sum_S (-1)^{p(S)} 2^{c(S)} x^{n(S)}| dx \quad (1.34)$$

gde $p(S)$, $c(S)$ i $n(S)$ označava broj komponenti, broj ciklova i broj čvorova Saksovog grafa S , respektivno.

Ako označimo skup svih Saksovih grafova molekulskog grafa G sa $S = S(G)$, onda jednačinu 1.34 možemo pisati u skraćenom obliku kao:

$$E = f(S). \quad (1.35)$$

Označimo sa S_C podskup skupa S , koji čine oni Saksovi grafovi koji sadrže cikl C , i posmatrajmo izraz $f(S/S_C)$. Ovaj izraz sadrži efekte svih Saksovih grafova na E , osim efekata koji potiču od cikla C . Drugim rečima, $f(S/S_C)$ je jedna energiji–nalik veličina, nastala kao rezultat uticaja svih relevantnih strukturnih detalja iz posmatranog molekula, osim uticaja cikla C . Shodno tome, razlika

$$ef = f(S) - f(S/S_C) \quad (1.36)$$

se može interpretirati kao efekat cikla C na HMO ukupnu π -elektronsku energiju. Takođe, može se pokazati [133] da važi:

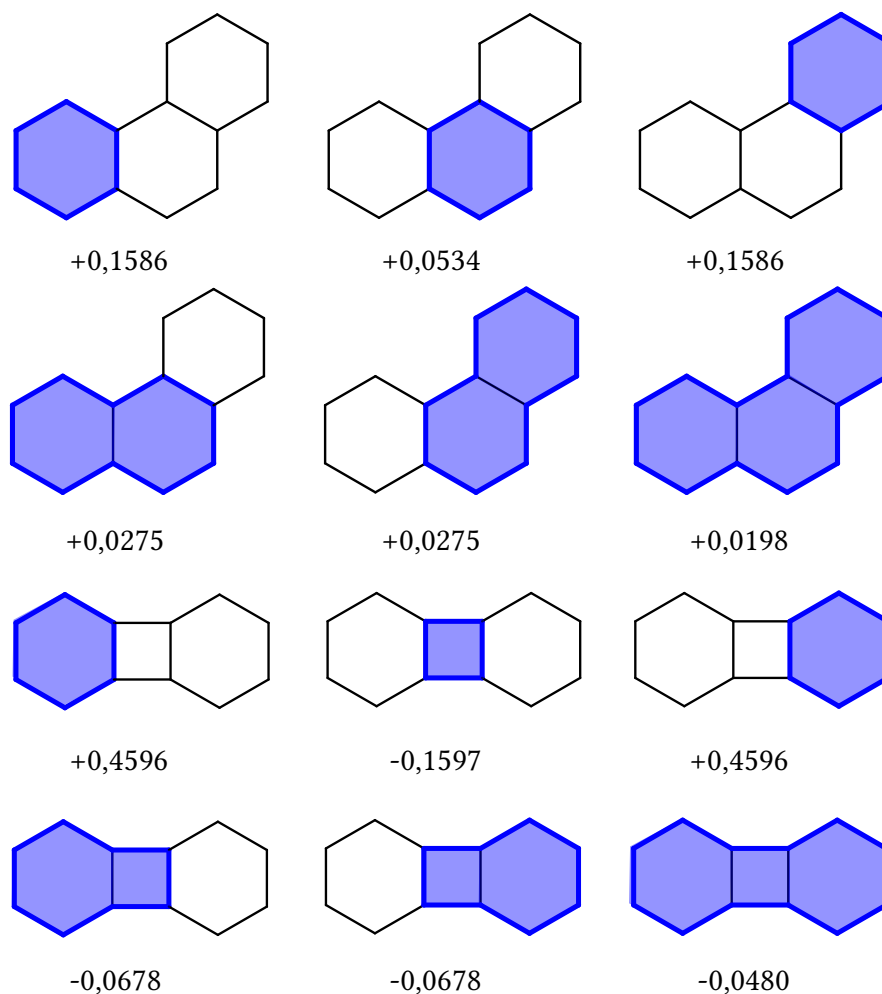
$$ef_E(G, C) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \ln \left| \frac{\phi(G, ix)}{\phi(G, ix) + 2\phi(G - C, ix)} \right| dx \quad (1.37)$$

gde je ϕ karakteristični polinom grafa G (videti poglavlje 1.2.2), a i je imaginarna jedinica ($i = \sqrt{-1}$).

U jednačini 1.37 $G - C$ označava podgraf dobijen izbacivanjem cikla C iz grafa G . Kada je $ef(G, C)$ pozitivan, cikl C stabilizuje molekul, a negativne ef -vrednosti ukazuju na destabilizaciju. Na slici 1.5 su prikazani energetski efekti za dva tipična policiklična konjugovana sistema.

Primeri prikazani na slici 1.5 ilustruju neke od osnovnih osobina ciklične konjugacije:

- Ne samo prstenovi, nego i veći ciklovi doprinose energiji.
- Energetski efekti obično opadaju sa porastom veličine cikla, ali nisu zanemarljivi čak i kod većih ciklova.



Slika 1.5: Energetski efekti ciklova (obojeni plavom bojom) fenantrena i bifenilena, izraženi u jedinicama HMO integrala rezonancije β [132]

- Ciklovi iste veličine mogu imati značajno različite energetske efekte.
- U primerima prikazanim na slici 1.5, ciklovi veličine 6, 10 i 14 imaju stabilizirajući, a ciklovi veličine 4, 8 i 12 destabilizirajući efekat. To je u punoj saglasnosti sa Hikelovim $4n + 2$ pravilom.
- Međutim, nasuprot onome što bi hemičari očekivali na osnovu svoje intuicije, Hikelovo pravilo nije zadovoljeno u opštem slučaju.
 - U svim alternantnim policikličnim konjugovanim ugljovodonicima, ciklovi veličine 4, 8, 12, 16, ... uvek imaju negativne ef -vrednosti i tako uvek destabilizuju odgovarajući molekul. (Ovo je upravo jedna polovina Hikelovog pravila.)
 - U većini slučajeva ciklovi veličine 6, 10, 14, 18, ... imaju stabilizujući efekat. Međutim, postoje izuzeci. Naime, postoje alternantni policiklični ugljovodonici u kojima neki od $(4n + 2)$ -članih ciklova imaju destabilizujući efekat i na taj način narušavaju Hikelovo pravilo.

1.5.2 Uticaj na Estradin indeks

Karakteristični polinom se može predstaviti kao:

$$\phi(G, \lambda) = \phi(C, \lambda) + \phi(G/C, \lambda) \quad (1.38)$$

gde je $\phi(C, \lambda)$ polinom cikla, a $\phi(G/C, \lambda)$ polinom grafa iz koga smo izbacili cikl. Ako važi sledeća jednakost:

$$\begin{aligned} \phi(C, \lambda) &= \sum_{s \in G|C} (-1)^{p(s')+1} \cdot 2^{c(s')+1} \cdot \lambda^{n-n(s')-|C|} \\ &= -2 \cdot \sum_{s \in G|C} (-1)^{p(s')} \cdot 2^{c(s')} \cdot \lambda^{n-n(s')-|C|} \\ &= -2\phi(G - C, \lambda) \end{aligned}$$

onda se polinom $\phi(G/C, \lambda)$ može izračunati kao:

$$\phi(G/C, \lambda) = \phi(G, \lambda) + 2\phi(G - C, \lambda). \quad (1.39)$$

Estradin indeks kod kompleksnih sopstvenih vrednosti se može izračunati pomoću formule:

$$EE(G) = \sum_{k=1}^n \cos b_k \cdot e^{a_k} \quad (1.40)$$

gde su a_k i b_k realni i kompleksni deo sopstvene vrednosti $z_k = a_k + b_k \cdot i$. Efekat cikla na Estradin indeks se onda može izračunati kao:

$$ef_{EE}(G, C) = EE[\phi(G, \lambda)] - EE[\phi(G/C, \lambda)]. \quad (1.41)$$

1.5.3 Uticaj na rezolventnu energiju

U prethodnom tekstu je prikazano kako se može izračunati uticaj cikla na vrednost energije grafa i Estradinog indeksa. Na sličan način može se izračunati uticaj cikla na vrednost rezolventne energije upotrebom Saksovih grafova. Da bi došli do tog izraza neophodno je da izvršimo neke pripreme. Rezolventna energija može se izračunati kao [116]:

$$ER(G) = \frac{\phi'(G, n)}{\phi(G, n)} \quad (1.42)$$

gde je $\phi'(G, n)$ prvi izvod polinoma $\phi(G, n)$.

Efekat cikla na rezolventnu energiju se može predstaviti kao:

$$ef_{ER}(G, C) = ER(f(\phi(G, \lambda))) - ER(f(\phi(G/C, \lambda))) \quad (1.43)$$

što nas dovodi do:

$$ef_{ER}(G, C) = ER(f(\phi(G, \lambda))) - ER(f(\phi(G, \lambda) + 2\phi(G - C, \lambda))). \quad (1.44)$$

Ako primenimo jednačinu 1.42 na jednačinu 1.44 dobijamo konačni izraz za računanje energetskog efekta cikla kod rezolventne energije:

$$ef_{ER}(G, C) = \frac{\phi'(G, n)}{\phi(G, n)} - \frac{[\phi(G, n) + 2\phi(G - C, n)]'}{\phi(G, n) + 2\phi(G - C, n)}. \quad (1.45)$$

1.6 Predviđanje fizičko–hemijskih osobina jedinjenja

Bez obzira na činjenicu da se broj eksperimentalnih podataka svakodnevno uvećava, broj sintetisanih i kompjuterski dizajniranih jedinjenja je mnogo veći. Mnoga jedinjenja mogu biti teorijski ispitivana na različitim nivoima, a da nisu još uvek sintetisana. Zbog toga je predviđanje njihovih fizičko–hemijskih osobina od velikog značaja [134–138]. Da bi se ovo omogućilo u praksi, neophodno je da postoje metode za predviđanje fizičko–hemijskih osobina molekula. Generalno, predviđanje fizičko–hemijskih osobina jedinjenja se može predstaviti kao:

$$\text{osobina molekula} = f(\text{molekulska struktura}).$$

Funkcija f može imati jednostavan oblik, ali u većini slučajeva f je komplikovana i njen oblik direktno zavisi od fizičko–hemijske osobine koja se modelira. Takođe, poteškoće u predviđanju fizičko–hemijskih osobina molekula mogu se javiti i prilikom odabira načina predstavljanja molekulske strukture. Velike razlike u vrednostima izračunatih molekulskih deskriptora mogu se javiti kod tautomera i konformera, što može loše uticati na razvijeni predikcioni model.

1.6.1 Metode za predviđanje osobina

U ovom odeljku biće predstavljene najčešće metode za predviđanje fizičko–hemijskih osobina jedinjenja. U te metode spadaju [138]:

- Predviđanje fizičko–hemijske osobine na osnovu druge fizičko–hemijske osobine
Ovaj pristup se zasniva na činjenici da su mnoge fizičko–hemijske osobine međusobno povezane. Tako na primer, jedinjenja sa većom tačkom topljenja su često slabije rastvorljiva. Ovakve veze među osobinama se mogu dobiti teorijskim ili eksperimentalnim putem.
- Predviđanje fizičko–hemijskih osobina na osnovu teorijskih proračuna
Najznačajnije metode iz ove grupe su one koje su bazirane na kvantno–mehaničkim proračunima. Zbog kompleksnosti proračuna, *ab initio* modeliranje celih sistema nije moguće; zbog toga se ove metode fokusiraju na modeliranje najvažnijih interakcija.
- Aditivne šeme za predviđanje fizičko–hemijskih osobina
Molekul se sastoji od atoma koji su povezani hemijskim vezama. Ideja da se fizičko–hemijska osobina molekula može raspodeliti preko atoma, tj. da se može izračunati preko doprinosa svih atoma u molekulu se pokazala kao ispravna kod nekih osobina jedinjenja. Na taj način, stvorene su nove metode za predviđanje fizičko–hemijskih osobina tzv. aditivne šeme.

- Šeme zasnovane na doprinosu atoma

Aditivne šeme koje se baziraju na doprinosu atoma se mogu generalno predstaviti kao:

$$\text{Osobina} = \sum_i^N a_i n_i$$

gde je a_i doprinos atoma i , a n_i je broj takvih atoma u molekulu.

- Šeme zasnovane na doprinosu funkcionalnih grupa

Ova metoda je veoma slična prethodnoj, ali se sada prati doprinos svake funkcionalne grupe u okviru molekula, a ne doprinos pojedinačnih atoma.

- Metode zasnovane na linernoj zavisnosti slobodne energije

Osnovna pretpostavka kod ovih metoda jeste da je uticaj određenog strukturnog detalja na promenu slobodne energije hemijskog procesa konstantan za seriju jedinjenja. Osobina, koja linearno zavisi od promene slobodne energije, onda može biti izračunata na osnovu osobine osnovnog jedinjenja u toj seriji.

- QSPR (engl. Quantitative Structure–Property Relationship) metoda

Ova metoda se zasniva na određivanju veze između strukturnih detalja i određene osobine, i njenom kvantifikovanju. Na ovaj način je moguće predvideti vrednost neke fizičko–hemijske osobine kod strukturno sličnih jedinjenja. Sličan pristup se može primeniti i za modeliranje aktivnosti jedinjenja, i tada se primenjuje QSPR metoda.

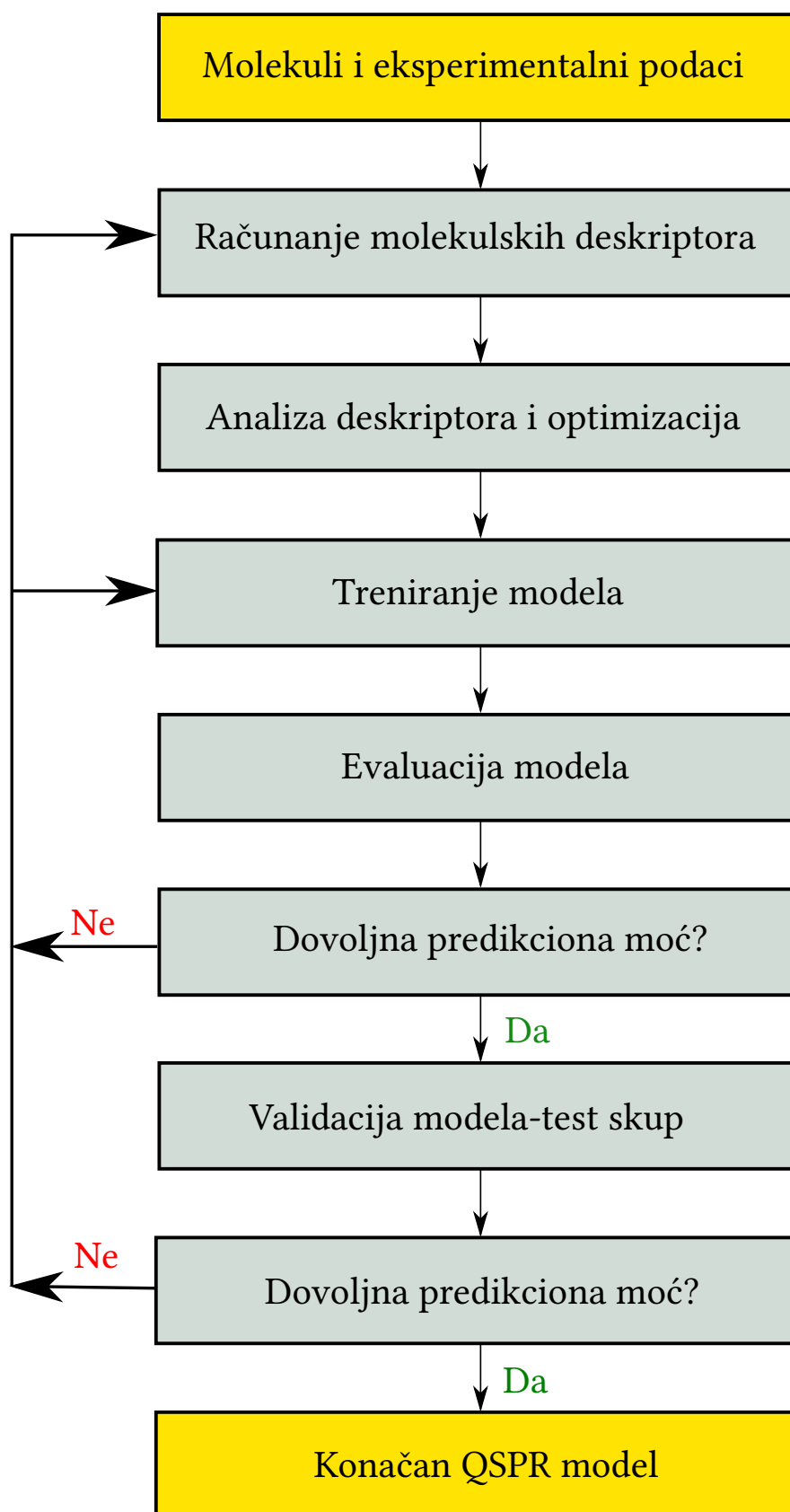
1.6.2 QSPR

QSPR metoda ima za cilj da kvantifikuje vezu između molekulske strukture i date osobine jedinjenja [17, 139]. Naravno, ovakav pristup se bazira na pretpostavci da je osobina koja se modelira određena strukturom datog molekula i da strukturno slična jedinjenja imaju slične osobine. Rad objavljen u časopisu američkog hemijskog društva (JACS) [140] se označava kao prvo istraživanje u kome je uspostavljen jedan QSPR model. Danas, pored osobina jedinjenja ovakav tip ispitivanja se primenjuje i za modeliranje biološke (fiziološke ili neke druge) aktivnosti, pa dobijamo QSAR modele, ili na toksičnost pa imamo QSTR model. Ovakav tip modeliranja je veoma značajan jer omogućava da se na osnovu osobina poznatih jedinjenja izvede matematički model koji će da predvidi osobine jedinjenja koja još uvek nisu sintetisana.

Razvijanje QSPR modela se odvija u nekoliko zasebnih koraka. Na slici 1.6 je šematski prikazano kako nastaje jedan QSPR model. Prvi korak predstavlja formiranje baze podataka koja se sastoji od molekula i njihovih eksperimentalnih podataka. Izvori iz kojih se eksperimentalni podaci sakupljaju su raznovrsni. Oni mogu biti prikupljeni samostalno u laboratoriji, mogu poticati iz objavljenih naučnih radova ili mogu biti prikupljeni iz online hemijskih baza podataka (npr. PubChem [141, 142] i NIST Chemistry Webbook [143, 144]). Naravno, valjanost eksperimentalnih podataka ima veliki uticaj na kvalitet i tačnost modela koji se razvija.

Sledeći korak predstavlja kvantifikovanje molekulske strukture. Ovo se vrši uz upotrebu molekulskih deskriptora. U okviru ove disertacije, ovaj korak će biti obavljen upotrebom topoloških molekulskih deskriptora, odnosno, upotrebom energije grafa, Estradinog indeksa i rezolventne energije. Analiza deskriptora i optimizacija je korak u kome se izabrani molekulski deskriptori ispituju na kvalitet informacija koje nose o strukturi i međusobnu povezanost. Kada je ovaj korak završen prikupljeni podaci se dele na dva skupa: na trening skup i na test skup. Podaci se uglavnom dele tako da 80% podataka bude u trening skupu, a 20% da pripada test skupu.

Treniranje modela je korak u kome neka od izabranih statističkih metoda (npr. višestruka linearna regresija, veštačke neuralne mreže i dr.) pokušava da uspostavi vezu između strukture i eksperimentalnih podataka tj. osobine molekula. Ovaj proces "učenja" modela se vrši na trening skupu podataka. Kada je ovaj korak gotov, matematička veza između strukture i date osobine je uspostavljena. Sledeći korak predstavlja ispitivanje datog modela na trening skupu. U ovom koraku se računaju različite statističke veličine koje opisuju razvijeni model. Ukoliko



Slika 1.6: Generalna šema konstrukcije QSPR modela [145]

model ne prođe ovu fazu ispitivanja ponovo se vrši treniranje modela, tj. traži se nova funkcija ili se pak vrši odabir novih molekulskih deskriptora. Posle ovoga, testiranje modela se vrši na test skupu podataka. To su podaci koji nisu korišćeni za razvijanje modela. Ovo je poslednji korak u konstruisanju QSPR modela i ukoliko na osnovu izračunatih statističkih veličina model pokaže dobru predikcionu moć, razvijeni model je konačni QSPR model.

1.6.2.1 Višestruka linearna regresija

Linearna regresija se može upotrebiti za modeliranje linearne zavisnosti između dve promenljive x i y . Veza između njih se može opisati pomoću jednačine prave koja ima oblik:

$$y = ax + b$$

gde a predstavlja nagib, a b odsečak prave. Cilj linearne regresije jeste da pronade one vrednosti za odsečak i nagib koje će dati najbolje predviđanje zavisne promenljive y pomoću nezavisne promenljive x . Kada imamo n nezavisno promenljivih x_i pomoću kojih želimo da modeliramo promenljivu y , tada se koristi višestruka linearna regresija. Ona ima sledeći oblik:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n. \quad (1.46)$$

Ukoliko jednačinu 1.46 primenimo na QSPR modeliranje, promenljiva y bi bila ispitivana osobina jedinjenja, x_i upotrebljeni molekulski deskriptor, a a_i parametar fitovanja. Prednost višestruke linearne regresije u odnosu na druge algoritme koji se koriste u QSPR (QSAR i QSTR) modeliranju je ta što nam koeficijent a_i , koji stoji uz molekulski deskriptor, direktno govori o značaju tog deskriptora za predviđanje osobine y . Drugim rečima, što a_i ima veću vrednost to je veći doprinos tog deskriptora u predviđanju y , i obrnuto. Zbog ovoga je višestruka linearna regresija jedna od najpopularnijih metoda za modeliranje osobina jedinjenja. Cilj višestruke linearne regresije jeste da pronade one vrednosti za a_i koje će dati najbolje predviđanje promenljive y pomoću n promenljivih x_i . Ovo se može postići na različite načine. Jedan od načina određivanja parametara fitovanja jeste metodom najmanjih kvadrata. U okviru ove disertacije biće korišćena ova metoda koja je implementirana u Python modulu za mašinsko učenje scikit-learn [146]. Neke od primena višestruke linearne regresije u QSPR/QSAR modeliranju se mogu pronaći u radovima [147–150].

1.6.2.2 Valdov test

Valdov test (W) predstavlja statistički test koji se može upotrebiti za procenjivanje značaja promenljivih u predikcionom modelu [151–153]. Na osnovu rezultata ovog testa može se statistički opravdati uvođenje ili odbacivanje neke promenljive iz datog modela. Često se u QSPR/QSAR modelima mogu naći promenljive koje ne doprinose objašnjenju varijanse podataka, pa je zbog toga neophodno sprovesti ovaj test kako bi bili sigurni da naš model sadrži samo korisne promenljive. Na ovaj način se mogu ukloniti beskorisne promenljive, bez uticaja na naš model. Valdova statistička formula ima sledeći oblik:

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\text{var}(\hat{\theta})} \quad (1.47)$$

gde je $\hat{\theta}$ procenjena, a θ_0 pretpostavljena vrednost.

1.6.2.3 Validacija modela

Validacija (evaluacija) je ključni korak u dobijanju i primeni predikcionih modela [154, 155]. Ona omogućava da se potvrdi pouzdanost razvijenog modela prateći svaki korak u konstrukciji samog modela kao što su kvalitet ulaznih podataka, testiranje na drugom skupu podataka i domen primene. Iako su QSPR modeli u primeni već dugi niz godina, i dalje postoje nedoumice oko toga na koji način okarakterisati kvalitet dobijenog modela. Ključni korak pri konstruisanju modela jeste prikupljanje podataka i njihova podela u trening i test skup. Trening skup predstavlja skup podataka koji se koristi za generisanje modela. U idealnim slučajevima on se sastoji od ogromnog broja molekula visoke raznolikosti. Test skup predstavlja podatke koji se koriste za određivanje predikcione greške odabranog modela. Cilj validacije jeste da obezbedi statistički pouzdan model koji će se sastojati od promenljivih koje stvarno nose neophodne informacije, a ne model koji je dobijen slučajno numeričkim putem [155–157].

U zavisnosti od toga na kojim podacima se dati model testira, validacija se može podeliti na internu i eksternu. Ukoliko se testiranje vrši na podacima koji su upotrebljeni za dobijanje modela (trening skup) to je interna validacija, a ukoliko se koriste drugi podaci (test skup) to je onda eksterna validacija.

Validacija QSPR/QSAR modela je veoma važna sa aspekta razumevanja pouzdanosti modela za predviđanje osobina/aktivnosti. Da bi se neki model u potpunosti okarakterisao neophodno je primeniti metode interne i eksterne validacije. Mali broj modela biva podvrgnut svim validacionim metodama. Najčešće se ispitivanje modela vrši računanjem nekih od statističkih veličina.

1.6.2.4 Osnovne validacione veličine

U ovom odeljku će biti predstavljene veličine na osnovu kojih se može okarakterisati predikciona moć konstruisanog QSPR modela, bilo u internoj ili u eksternoj validaciji.

Koeficijent determinacije i poboljšani koeficijent determinacije

R^2 označava koeficijent determinacije, koji predstavlja osnovnu statističku veličinu koja se računa prilikom validacije modela. Ovo je mera objašnjene varijanse pomoću nezavisno promenljivih od kojih je model izgrađen. Koeficijent determinacije je definisan na sledeći način:

$$R^2 = 1 - \frac{\sum(y_{eksp.} - y_{izr.})^2}{\sum(y_{eksp.} - \bar{y})^2} \quad (1.48)$$

gde \bar{y} označava srednju vrednost eksperimentalnih podataka, $y_{eksp.}$ eksperimentalnu vrednost, a $y_{izr.}$ izračunatu vrednost pomoću datog modela. Obično se izražava u procentima i npr. $R^2 = 99,9\%$ za neki model pokazuje da taj model odlično opisuje varijansu u datim podacima. Nizak koeficijent determinacije pokazuje da dati model ne sadrži potrebne informacije neophodne za opis varijanse. R^2 se upotrebljava i pri internoj i pri eksternoj validaciji modela. Loša strana R^2 je ta da se uvek povećava kako raste broj promenljivih u modelu [158]. To zapravo znači da u naš model mogu biti uključene promenljive koje ne nose nikakve nove informacije o podacima. Zbog toga je uz koeficijent determinacije neophodno izračunati i poboljšani koeficijent determinacije \bar{R}^2 . On je definisan na sledeći način:

$$\bar{R}^2 = 1 - \left[\frac{(1 - R^2) \cdot (N - 1)}{N - k - 1} \right] \quad (1.49)$$

U ovoj jednačini N predstavlja broj podataka, a k broj promenljivih u datom modelu. Na osnovu same definicije vidimo da je \bar{R}^2 uvek manje od R^2 . Razlika između ova dva koeficijenta je u tome što se koeficijent determinacije povećava sa povećanjem broja promenljivih, dok se poboljšani koeficijent determinacije povećava samo sa povećanjem broja korisnih promenljivih. To znači da ako promenljive nisu od koristi, odnosno ne doprinose objašnjenju varijanse, \bar{R}^2 će se smanjiti. Zbog toga je korisno pratiti vrednosti za oba koeficijenta i bilo bi dobro da su njihove vrednosti približne.

Prosečna apsolutna greška, prosečna relativna greška i koren kvadrata srednje greške

Statističke veličine za procenu modela zasnovane na greškama su prosečna apsolutna greška (AAE), prosečna relativna greška (ARE) i koren kvadrata srednje greške ($RMSE$). One su definisane na sledeći način:

$$AAE = \frac{1}{N} \sum |y_{eksp.} - y_{izr.}|, \quad (1.50)$$

$$ARE = \frac{1}{N} \sum \frac{|y_{eksp.} - y_{izr.}|}{|y_{eksp.}|}, \quad (1.51)$$

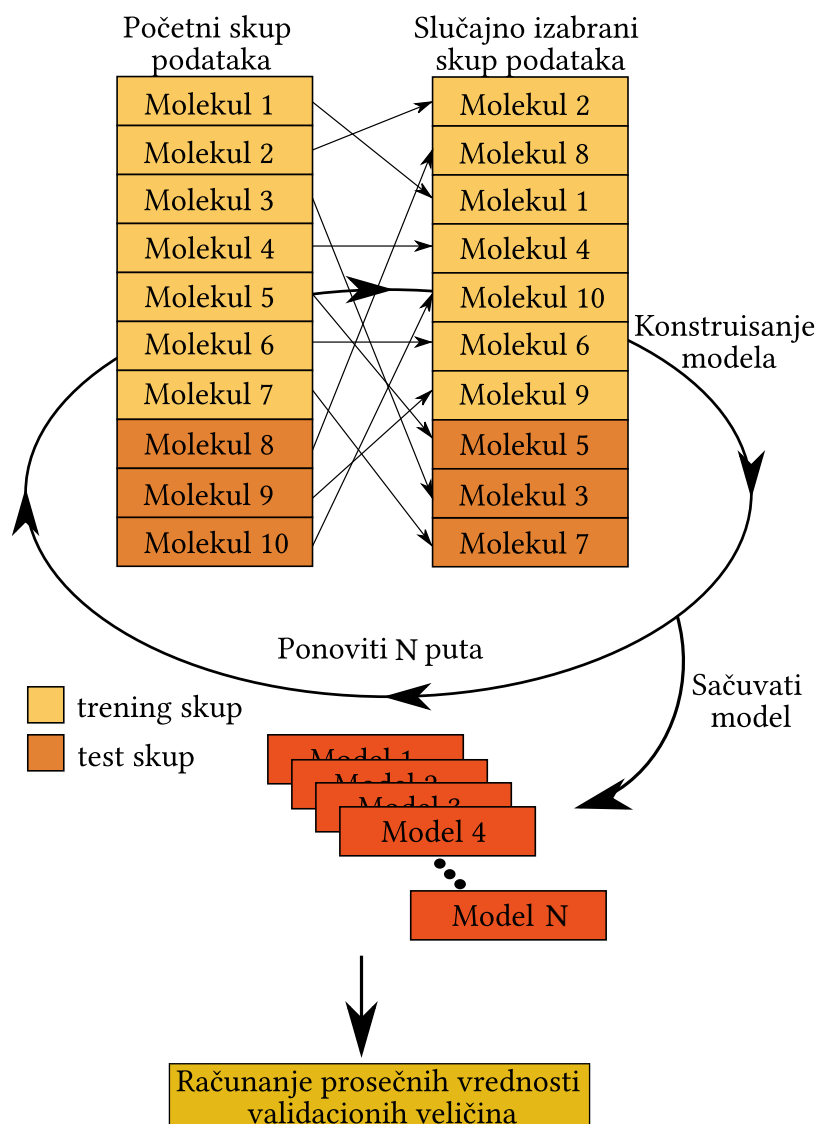
$$RSME = \sqrt{\frac{1}{N} \sum (y_{eksp.} - y_{izr.})^2}. \quad (1.52)$$

U ovim jednačinama N označava ukupan broj podataka, $y_{eksp.}$ eksperimentalnu vrednost, a $y_{izr.}$ izračunatu vrednost pomoću datog modela. Jasno je da sve veličine daju meru kolika je greška u predikciji u odnosu na ukupan broj podataka. Međutim, pretpostavlja se da je informacija koju nosi $RSME$ komplikovanija od one koju nose AAE i ARE . Zato što se kvadriraju greške u izrazu za $RSME$ može se promeniti disperzija tih vrednosti, odnosno, neće svaka greška imati isti doprinos. Na primer, kvadriranjem veće greške dobiće se veći broj koji značajnije doprinosi $RSME$. Kod apsolutne greške ovo nije slučaj i zato se ona smatra jednostavnim i dobrim pokazateljem predikcione moći modela. Ove greške se često izražavaju u procentima. Pri prikazivanju greške predikcije, poželjno je izračunati i standardnu devijaciju s . Tokom prakse uočilo se sledeće:

- potencijalno dobar model:
 $AAE \leq 0,1 \cdot \text{veličina trening skupa}$ i $AAE + 3 \cdot s \leq 0,2 \cdot \text{veličina trening skupa}$
- potencijalno loš model:
 $AAE \geq 0,15 \cdot \text{veličina trening skupa}$ i $AAE + 3 \cdot s \geq 0,25 \cdot \text{veličina trening skupa}$.

Butstrap tehnika

Butstrap tehnika (engl. bootstrap technique) je statistička metoda validacije predikcionog modela [155, 156, 159–161]. Zasniva se na ideji da postojeći uzorci predstavljaju populaciju i da se na osnovu njih može izvršiti statističko uzorkovanje. Ova metoda je posebno korisna kod relativno malog broja podataka. Procedura se sastoji u tome da se početni skup podataka, koji se sastoji od trening i test skupa, perturbira i da se slučajnim izborom generiše novi skup podataka iste veličine kao i početni. Na ovaj način dolazi do promene sastava trening i test skupa. Kada je generisanje novog skupa završeno, prelazi se na sledeći korak, a to je konstrukcija matematičkog predikcionog modela. Novi model se čuva i potom se ispituje. Naime, računaju se validacione veličine (koeficijent determinacije, prosečna apsolutna greška,



Slika 1.7: Šematski prikaz validacije matematičkog predikcionog modela pomoću butstrap tehnike

koren kvadrata srednje greške, i dr.). Postupak generisanja nove baze podataka i novog modela je potrebno ponoviti više puta. Broj ponavljanja zavisi od broja podataka, jer kod ogromnog broja podataka ova tehnika može biti kompjuterski veoma zahtevna. Na kraju je potrebno izračunati prosečne vrednosti računatih validacionih veličina i uporediti ih sa vrednostima dobijenim pomoću početnog skupa podataka. Šematski prikaz butstrap tehnike je prikazan na slici 1.7. Uspešnost butstrap tehnike u velikoj meri zavisi od izbora algoritma za slučajno biranje podataka.

2

CILJEVI DISERTACIJE

U okviru ove doktorske disertacije postavljeni su sledeći ciljevi:

- ◆ Određivanje veza između topoloških molekulskih deskriptora zasnovanih na sopstvenim vrednostima matrice susedstva;
- ◆ Konstruisanje modela za međusobno modeliranje deskriptora na osnovu prethodno identifikovanih veza;
- ◆ Određivanje degenerativnosti energije grafa, Estradinog indeksa i rezolventne energije;
- ◆ Ispitivanje strukturne osetljivosti ovih deskriptora na određenim grupama molekula;
- ◆ Ispitivanje uticaja cikla u strukturi molekula na vrednost energije grafa, Estradinog indeksa i rezolventne energije;
- ◆ Na kraju, određivanje hemijske primenjivosti energije grafa, Estradinog indeksa i rezolventne energije kroz modeliranje fizičko-hemijskih osobina jedinjenja.



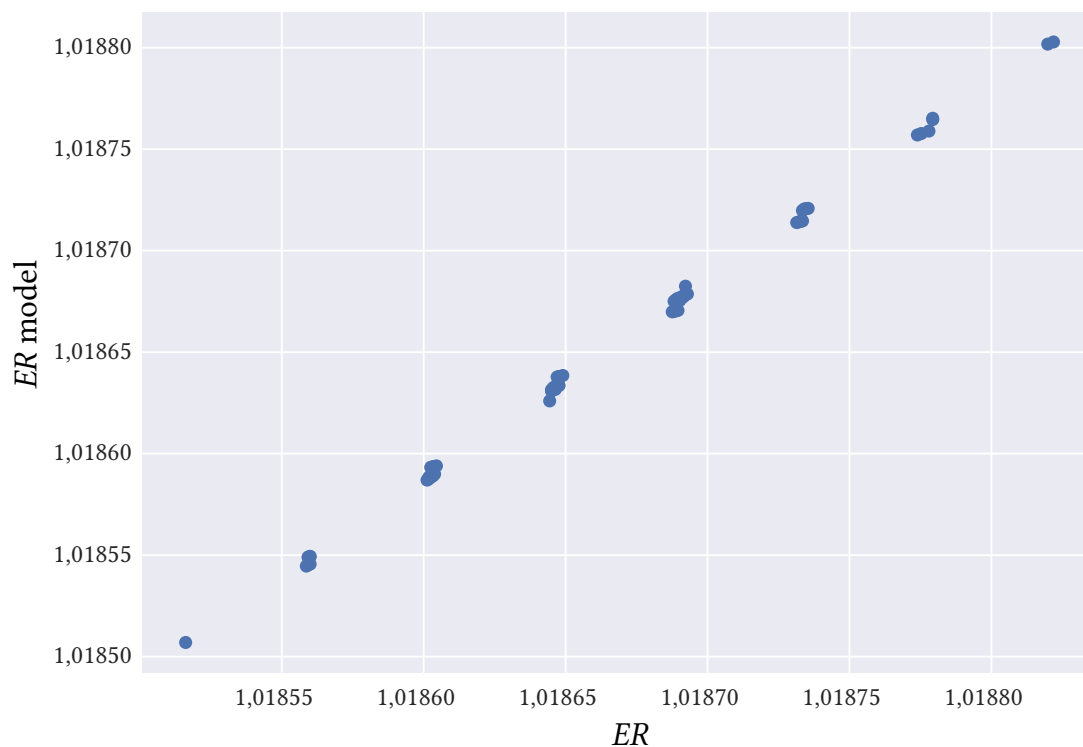
3

REZULTATI I DISKUSIJA

modul [146], koji smo implementirali u naš Python program. Da bi ispitali dati model, računata je rezolventna energija pomoću ovog modela za hemijska stabla od 6 do 20 čvorova. Rezultati ovog testiranja su predstavljeni u tabeli 3.1 i na slici 3.2.

Tabela 3.1: Vrednosti koeficijenata iz jednačine 3.1 za hemijska stabla sa $n = 6 - 20$. R i ARE predstavljaju koeficijent korelacije i prosečnu relativnu grešku, respektivno

n	A	B	C	D	R	ARE
6	$-2,21 \cdot 10^{-5}$	$3,19 \cdot 10^{-4}$	$-2,13 \cdot 10^{-5}$	1,044	1,00000	$8,43 \cdot 10^{-8}$
7	$6,90 \cdot 10^{-5}$	$1,47 \cdot 10^{-4}$	$2,11 \cdot 10^{-5}$	1,033	0,99999	$1,15 \cdot 10^{-6}$
8	$2,80 \cdot 10^{-5}$	$7,24 \cdot 10^{-5}$	$7,48 \cdot 10^{-6}$	1,026	0,99996	$1,29 \cdot 10^{-6}$
9	$1,55 \cdot 10^{-5}$	$3,91 \cdot 10^{-5}$	$4,56 \cdot 10^{-6}$	1,021	0,99998	$6,39 \cdot 10^{-7}$
10	$7,46 \cdot 10^{-6}$	$2,25 \cdot 10^{-5}$	$2,05 \cdot 10^{-6}$	1,018	0,99997	$4,12 \cdot 10^{-7}$
11	$4,01 \cdot 10^{-6}$	$1,37 \cdot 10^{-5}$	$1,15 \cdot 10^{-6}$	1,015	0,99998	$2,29 \cdot 10^{-7}$
12	$2,12 \cdot 10^{-6}$	$8,74 \cdot 10^{-6}$	$5,71 \cdot 10^{-7}$	1,013	0,99998	$1,42 \cdot 10^{-7}$
13	$1,25 \cdot 10^{-6}$	$5,79 \cdot 10^{-6}$	$3,48 \cdot 10^{-7}$	1,011	0,99998	$8,71 \cdot 10^{-8}$
14	$7,29 \cdot 10^{-7}$	$3,96 \cdot 10^{-6}$	$1,94 \cdot 10^{-7}$	1,009	0,99999	$5,62 \cdot 10^{-8}$
15	$4,61 \cdot 10^{-7}$	$2,78 \cdot 10^{-6}$	$1,25 \cdot 10^{-7}$	1,008	0,99999	$3,63 \cdot 10^{-8}$
16	$2,89 \cdot 10^{-7}$	$2,00 \cdot 10^{-6}$	$7,65 \cdot 10^{-8}$	1,007	0,99999	$2,45 \cdot 10^{-8}$
17	$1,92 \cdot 10^{-7}$	$1,47 \cdot 10^{-6}$	$5,11 \cdot 10^{-8}$	1,006	0,99999	$1,67 \cdot 10^{-8}$
18	$1,28 \cdot 10^{-7}$	$1,10 \cdot 10^{-6}$	$3,35 \cdot 10^{-8}$	1,006	0,99999	$1,17 \cdot 10^{-8}$
19	$8,83 \cdot 10^{-8}$	$8,36 \cdot 10^{-7}$	$2,32 \cdot 10^{-8}$	1,005	1,00000	$8,35 \cdot 10^{-9}$
20	$6,14 \cdot 10^{-8}$	$6,45 \cdot 10^{-7}$	$1,60 \cdot 10^{-8}$	1,005	1,00000	$6,07 \cdot 10^{-9}$



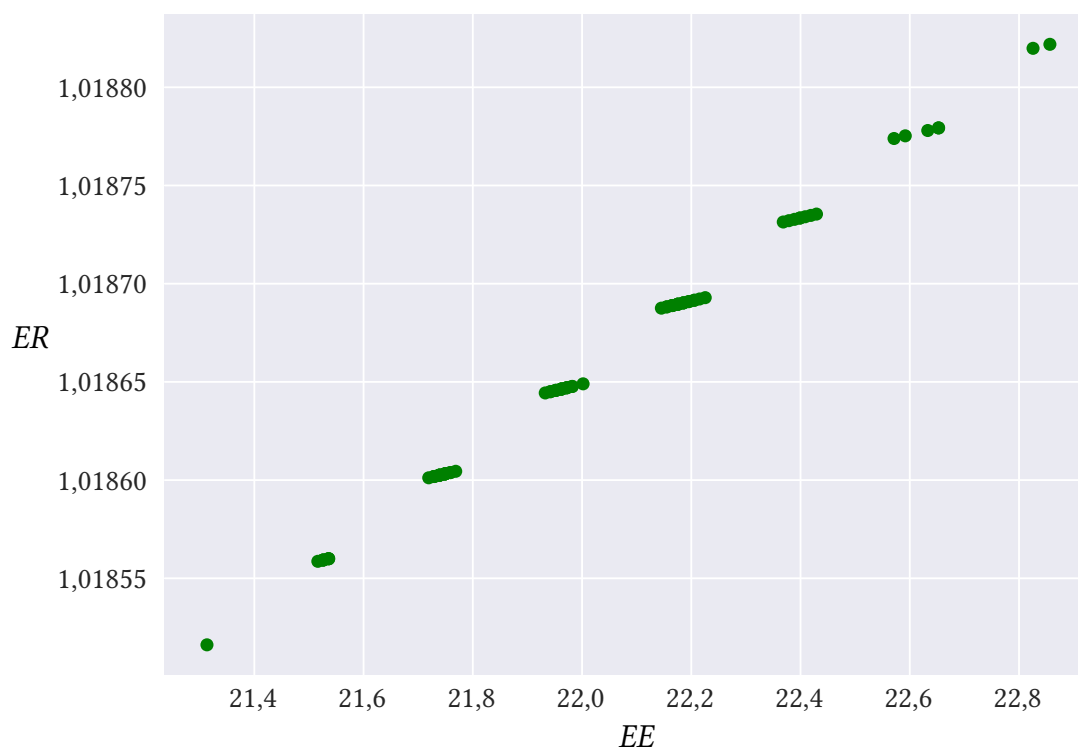
Slika 3.2: Korelacija između rezolventne energije i modela zasnovanog na jednačini 3.1 za hemijska stabla sa $n = 10$

Vrednosti za koeficijent korelacije (R) i prosečne relativne greške (ARE) pokazuju odlično

slaganje između rezolventne energije i rezolventne energije izračunate pomoću jednačine 3.1. To znači da prvi zagrebački indeks i broj nula u spektru skoro u potpunosti objašnjavaju zavisnost između energije grafa i rezolventne energije u slučaju hemijskih stabala.

3.1.2 Veza između Estradinog indeksa i rezolventne energije

Ilustrativni primer korelacije između Estradinog indeksa i rezolventne energije prikazan je na slici 3.3. Iako je korelacija između ova dva deskriptora dobra, lako se može primetiti da su vrednosti grupisane. Direktnim proveravanjem utvrdili smo da su vrednosti grupisane na osnovu prvog zagrebačkog indeksa. Naime, tačke u okviru iste grupe imaju istu vrednost prvog zagrebačkog indeksa.



Slika 3.3: Korelacija između Estradinog indeksa i rezolventne energije na primeru hemijskih stabala sa $n = 10$

Značajnost prvog zagrebačkog indeksa u korelaciji između Estradinog indeksa i rezolventne energije je matematički obrađena u tekstu koji sledi. Kao što smo ranije spomenuli Estradin indeks i rezolventna energija se mogu predstaviti pomoću spektralnih momenata koristeći Tejlorov niz [116]:

$$ER(G) = \sum_{k \geq 0} \frac{M_k(G)}{n^{k+1}}. \quad (3.2)$$

Koristeći jednačine 1.23 i 3.2, sledeća jednakost može biti uspostavljena:

$$ER(G) = \frac{EE(G)}{n} + \sum_{k \geq 0} \frac{k! - n^k}{k! \cdot n^{k+1}} M_k(G). \quad (3.3)$$

Kod bipartitnih grafova, kakva su hemijska stabla, neparni spektralni momenti su jednaki 0. Koristeći jednačinu 3.3, $ER(T)$ se može aproksimirati koristeći $EE(T)$ i nekoliko prvih

spektralnih momenata:

$$ER(T) \approx \frac{\alpha}{n} EE(T) + \frac{1-\alpha}{n} M_0(T) + \frac{2-\alpha \cdot n^2}{2n^3} M_2(T) + \frac{24-\alpha \cdot n^4}{24n^5} M_4(T) \quad (3.4)$$

gde je α odgovarajući parametar.

Uključivanjem jednačina 1.7, 1.8 i 1.9 u jednačinu 3.4, aproksimativna formula koja povezuje Estradin indeks i rezolventnu energiju dobija konačni oblik:

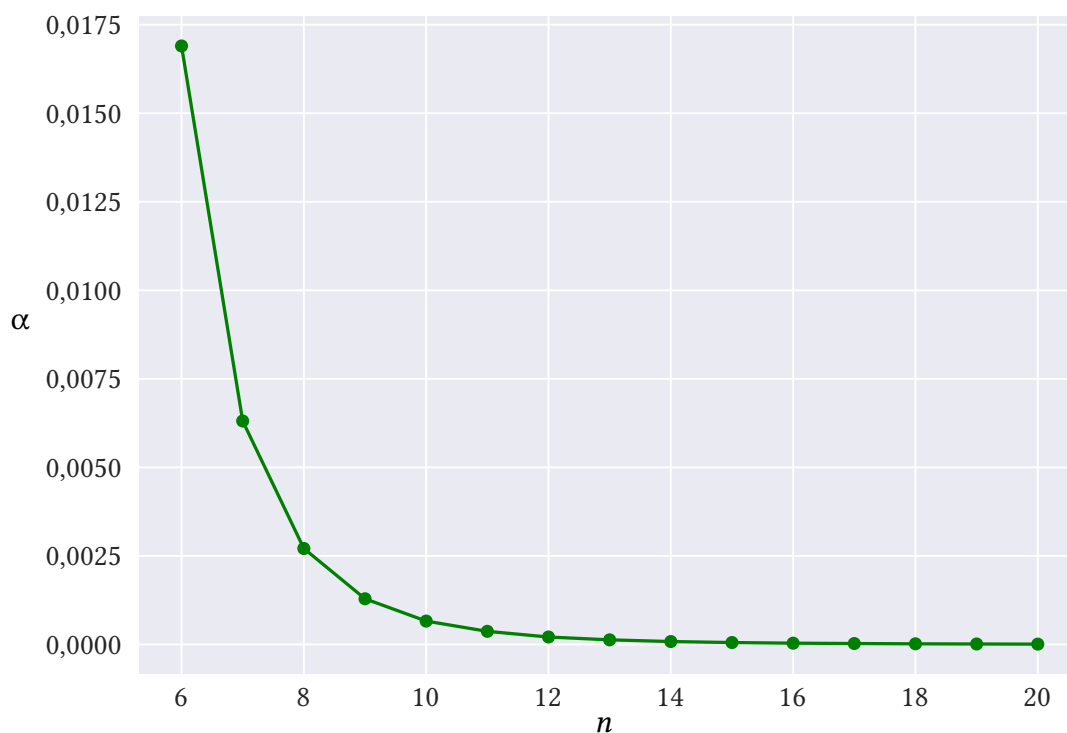
$$ER(T) \approx \frac{\alpha}{n} EE(T) - \frac{\alpha \cdot n^4 - 24}{12n^5} Zg_1(T) - \frac{11 \cdot \alpha \cdot n^4 - 24n^2 + 24}{12n^5} (n-1) - \alpha + 1. \quad (3.5)$$

U cilju određivanja parametra α koji se pojavljuje u aproksimativnoj formuli 3.5 napravljen je poseban kompjuterski program. Ovaj program je napisan u Pitonu i izračunate su vrednosti parametra α za hemijska stabla od 6 do 20 čvorova. Rezultati su predstavljeni u sledećoj tabeli.

Tabela 3.2: Vrednosti parametra α iz jednačine 3.5

n	α	n	α
6	0,0169	14	$8,2 \cdot 10^{-5}$
7	0,00631	15	$5,4 \cdot 10^{-5}$
8	0,00271	16	$3,6 \cdot 10^{-5}$
9	0,00129	17	$2,6 \cdot 10^{-5}$
10	0,00066	18	$1,8 \cdot 10^{-5}$
11	0,00037	19	$1,3 \cdot 10^{-5}$
12	0,00021	20	$9,0 \cdot 10^{-6}$
13	0,00013		

Kada se upotrebe vrednosti parametra α , predstavljene u tabeli 3.2, koeficijent korelacije iznosi 1 zaokružen na sedam decimala.



Slika 3.4: Zavisnost parametra α od broja čvorova kod hemijskih stabala

Na slici 3.4 je prikazana zavisnost parametra α od broja čvorova kod hemijskih stabala. Iako je ova zavisnost kompleksna, jasno je da je α određeno brojem čvorova. Ovo ukazuje da kod izomernih hemijskih stabala, gde je $n = const.$, na vezu EE i ER utiče samo Zg_1 .

Kako bi dalje ispitali vezu između Estradinog indeksa i rezolventne energije odlučili smo da ispitamo njihovu međusobnu zavisnost kod benzenoidnih ugljovodnika. Ova grupa molekula se može podeliti na katakondenzovane i perikondenzovane benzenoidne ugljovodonike. Osnovna razlika između ove dve grupe jedinjenja jeste ta što perikondenzovani molekuli sadrže unutrašnje čvorove (čiji je broj označen sa n_i). Grafovi izvedeni iz benzenoidnih ugljovodnika sadrže samo čvorove stepena 2 i 3. Broj čvorova stepena 2 (n_{d2}) i broj čvorova stepena 3 (n_{d3}) može se dobiti na osnovu sledećih izraza [164]:

$$n_{d2} = 2h + 4 - n_i \quad (3.6)$$

$$n_{d3} = 2h - 2 \quad (3.7)$$

gde h označava broj heksagona u datom molekulu (grafu). U grupi izomernih benzenoidnih ugljovodnika su h i n_i konstante, a kako Zg_1 zavisi od stepena čvorova to dovodi do toga da je $Zg_1 = const.$ Na osnovu jednačine 3.5 očekuje se linearna zavisnost između rezolventne energije i Estradinog indeksa, odnosno, očekuje se da kod izomernih benzenoidnih ugljovodnika ER zavisi samo od EE . Kako ova dva indeksa variraju u različitim opsezima, modifikovana normalizacija vrednosti ovih indeksa je izvršena na osnovu sledećeg izraza:

$$TI'_i = \frac{TI_i - TI_{max}}{s(TI)}$$

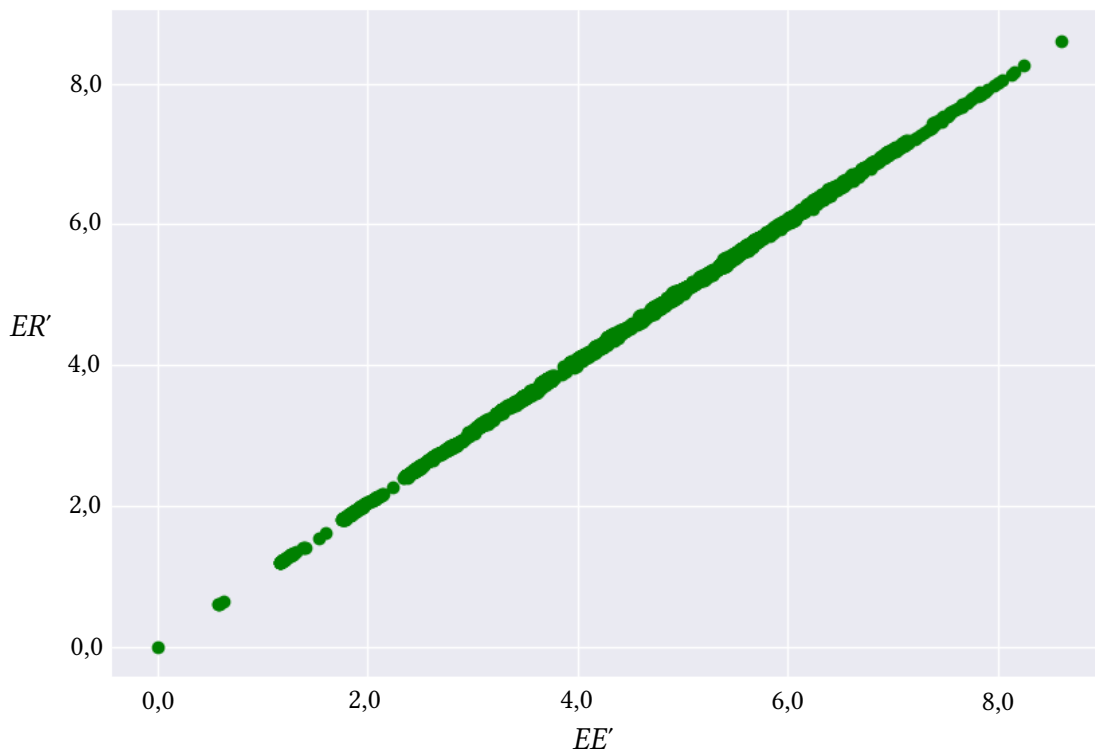
gde $s(TI)$ označava standardnu devijaciju.

Koristeći kompjuterski program CaGe [165], konstruisane su serije benzenoidnih ugljovodnika. Kod katakondenzovanih molekula je $h = 5 - 10$, dok je kod perikondenzovanih $h = 8 - 10$. U okviru jedne grupe, kod perikondenzovanih molekula, broj unutrašnjih čvorova nije menjan. Zato što postoji tek nekoliko izomera sa brojem heksagona manjim od 8, ispitivanja na takvim perikondenzovanim molekulima nisu vršena. Rezultati su predstavljeni u tabeli 3.3.

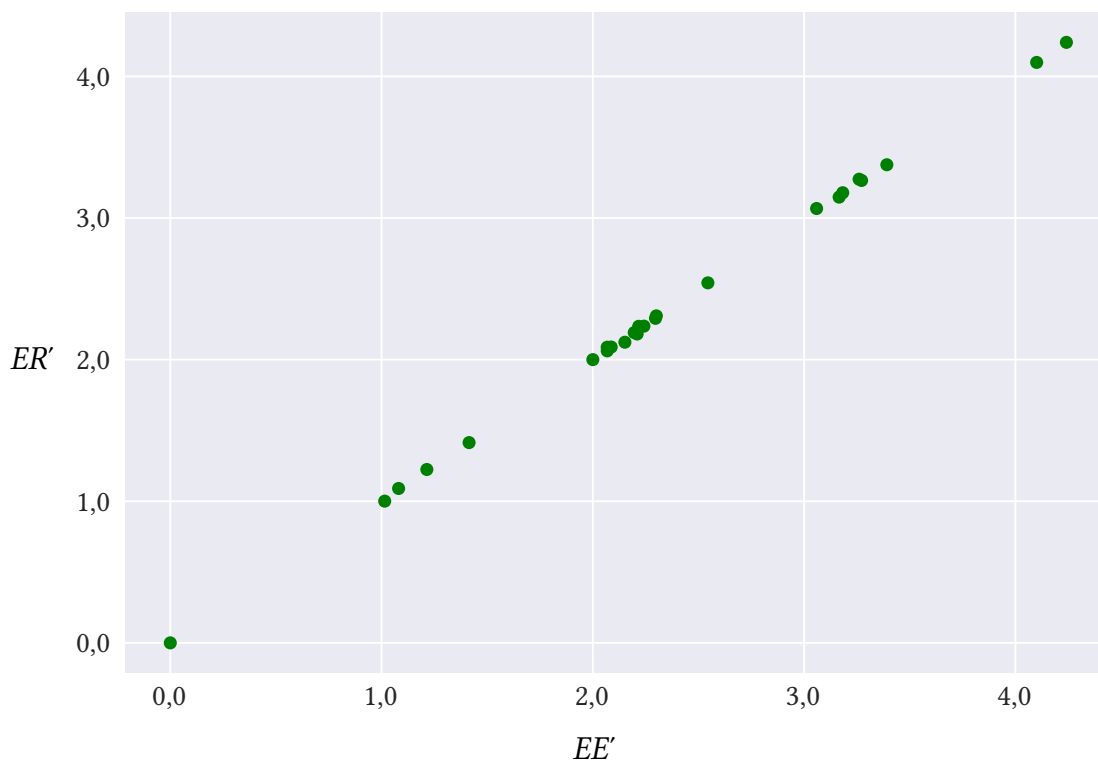
Tabela 3.3: Koeficijenti korelacije (R) između Estradinog indeksa i rezolventne energije kod benzenoidnih ugljovodnika

	h	broj izomera	R	
perikondenzovani	5	12	0,99990	
	6	36	0,99978	
	7	118	0,99981	
	8	411	0,99986	
	9	1489	1,00000	
	10	5572	0,99987	
	h	broj izomera	n_i	R
katakondenzovani	8	8	6	0,99994
	9	12	7	0,99997
	10	25	8	0,99992

Na slikama 3.5 i 3.6 su prikazane linearne korelacije između normalizovanog Estradinog indeksa (EE') i normalizovane rezolventne energije (ER') kod katakondenzovanih i perikondenzovanih benzenoidnih ugljovodnika.



Slika 3.5: Korelacija između normalizovanog Estradinog indeksa i normalizovane rezolventne energije u grupi koja se sastoji od 5572 katakondenzovana benzenoidna ugljovodonika



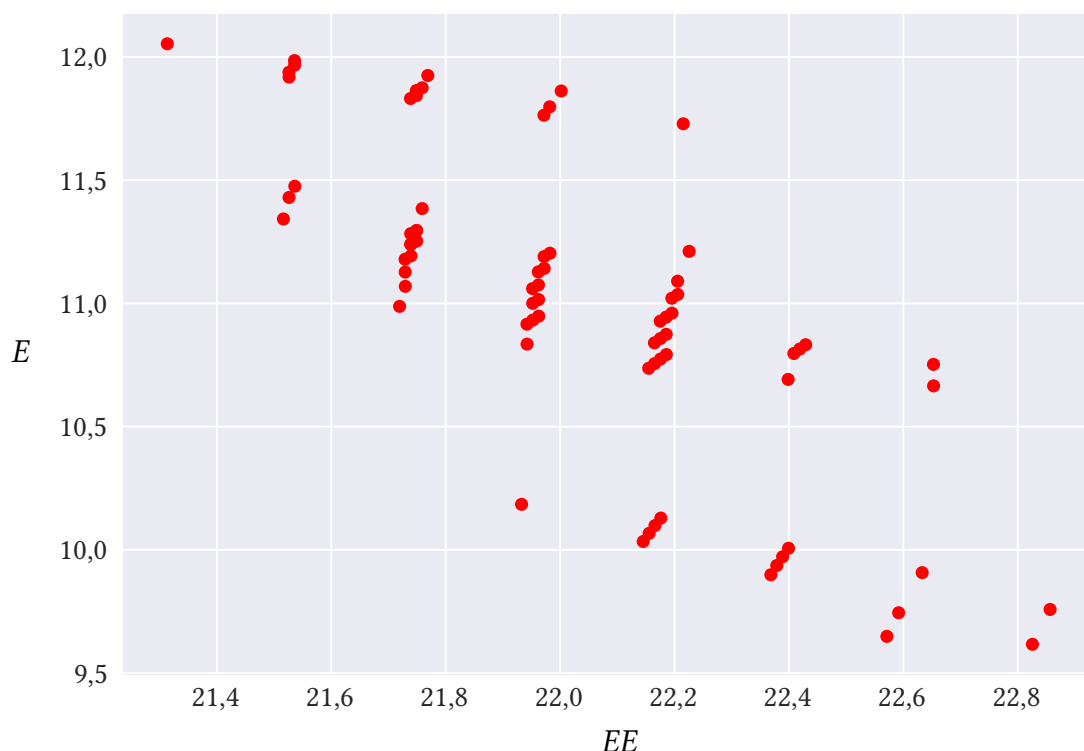
Slika 3.6: Korelacija između normalizovanog Estradinog indeksa i normalizovane rezolventne energije u grupi koja se sastoji od 25 perikondenzovana benzenoidna ugljovodonika

Koeficijenti korelacije prikazani u tabeli 3.3, kao i slike 3.5 i 3.6, pokazuju da između

Estradinog indeksa i rezolventne energije postoji odlična linearna korelacija.

3.1.3 Veza između energije grafa i Estradinog indeksa

Od uvođenja Estradinog indeksa otpočela su intenzivna istraživanja ovog deskriptora. Takođe, određeno je nekoliko granica za ovaj indeks, koje su izražene preko energije grafa [108, 109]. Međutim, do sada nije ispitivana veza između energije grafa i Estradinog indeksa. Na slici 3.7 je prikazana korelacija između ova dva indeksa na primeru hemijskih stabala sa 10 čvorova. Aproksimativne formule 3.1 i 3.5 sugerišu da prvi zagrebački indeks i broj nula utiču na vezu između energije grafa i Estradinog indeksa.



Slika 3.7: Korelacija između Estradinog indeksa i energije grafa na primeru hemijskih stabala sa $n = 10$

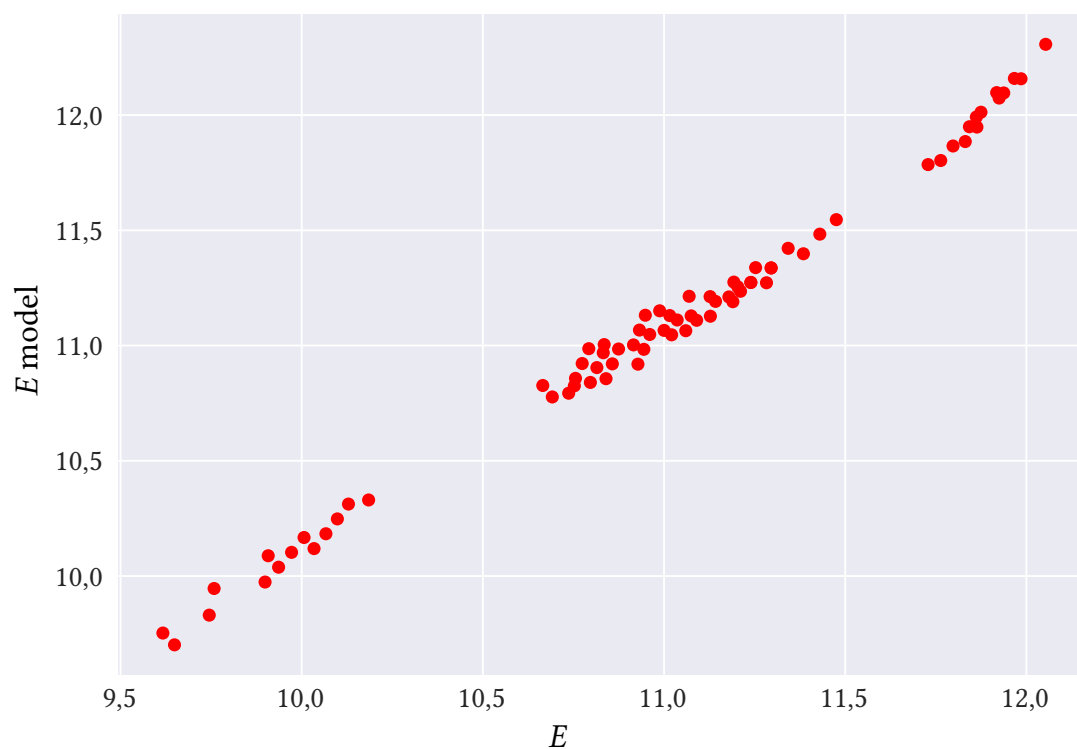
Na osnovu toga, pretpostavili smo da se energija grafa može modelirati pomoću sledeće formule:

$$E(T) \approx A \cdot EE(T) + B \cdot Zg_1(T) + C \cdot n_0(T) + D. \quad (3.8)$$

Model 3.8 je podvrgnut testiranju, i u tu svrhu konstruisan je Piton program i rezultati su predstavljeni u tabeli 3.4. Prikazani statistički podaci pokazuju da model 3.8 objašnjava 98% varijanse podataka. Vrednosti za prosečne relativne greške su veoma male. Ilustrativni primer korelacije između energije grafa i energije izračunate pomoću modela dat je na slici 3.8. Bez obzira na veoma veliki procenat objašnjene varijanse pomoću modela, jasno je da pored prvog zagrebačkog indeksa i broja nula u spektru, još neki parametar utiče na vezu između energije grafa i Estradinog indeksa.

Tabela 3.4: Vrednosti koeficijenata iz jednačine 3.8 za hemijska stabla sa $n = 6 - 20$

n	A	B	C	D	R	ARE
6	-215,6	22,85	-1,384	2224,4	0,99995	0,07%
7	12,08	-1,388	-0,314	-135,9	0,99936	0,23%
8	7,78	-0,933	-0,31	-96,5	0,99671	0,47%
9	7,48	-0,903	-0,318	-104,3	0,99746	0,36%
10	6,26	-0,771	-0,307	-94,9	0,99563	0,45%
11	5,72	-0,713	-0,322	-94,3	0,99543	0,40%
12	5,53	-0,695	-0,309	-99,1	0,99427	0,42%
13	5,21	-0,659	-0,319	-100,1	0,99423	0,41%
14	5,11	-0,65	-0,31	-105,6	0,99339	0,41%
15	4,93	-0,63	-0,315	-108,5	0,99337	0,39%
16	4,86	-0,623	-0,31	-114	0,99278	0,39%
17	4,74	-0,609	-0,312	-117,6	0,9927	0,38%
18	4,69	-0,604	-0,309	-122,9	0,99229	0,37%
19	4,6	-0,595	-0,31	-127	0,99219	0,36%
20	4,56	-0,59	-309	-132,2	0,99189	0,36%



Slika 3.8: Korelacija između energije grafa i modela zasnovanog na jednačini 3.8 za hemijska stabla sa $n = 10$

3.2 Degenerativnost molekulskih deskriptora zasnovanih na sopstvenim vrednostima i r -ekvienergetska hemijska stabla

Degenerativnost molekulskog deskriptora je jedna od važnijih osobina deskriptora. U ovom poglavlju biće predstavljeni rezultati koji se odnose na degenerativnost energije grafa, Estradinog indeksa i rezolventne energije, a koji su publikovani u [166].

3.2.1 Degenerativnost

Jedna od glavnih odlika korisnog topološkog molekulskog deskriptora jeste njegova sposobnost razlikovanja izomera [167]. Iako postoji mnoštvo deskriptora, nijedan od njih nije uspeo da u potpunosti ispuni ovaj zahtev. Zbog toga je značajno rangirati molekulske deskriptore prema njihovoj sposobnosti razlikovanja izomera tj. prema njihovoj senzitivnosti (S). Da bi to bilo moguće neophodno je da postoji način kvantitativnog određivanja senzitivnosti.

Iako je ranije bilo više pokušaja kvantitativnog određivanja S , danas je u upotrebi veličina koju je 1996. godine uvela Konstantinova [168]. Ona ju je nazvala senzitivnost molekulskog deskriptora. Definisana je na sledeći način:

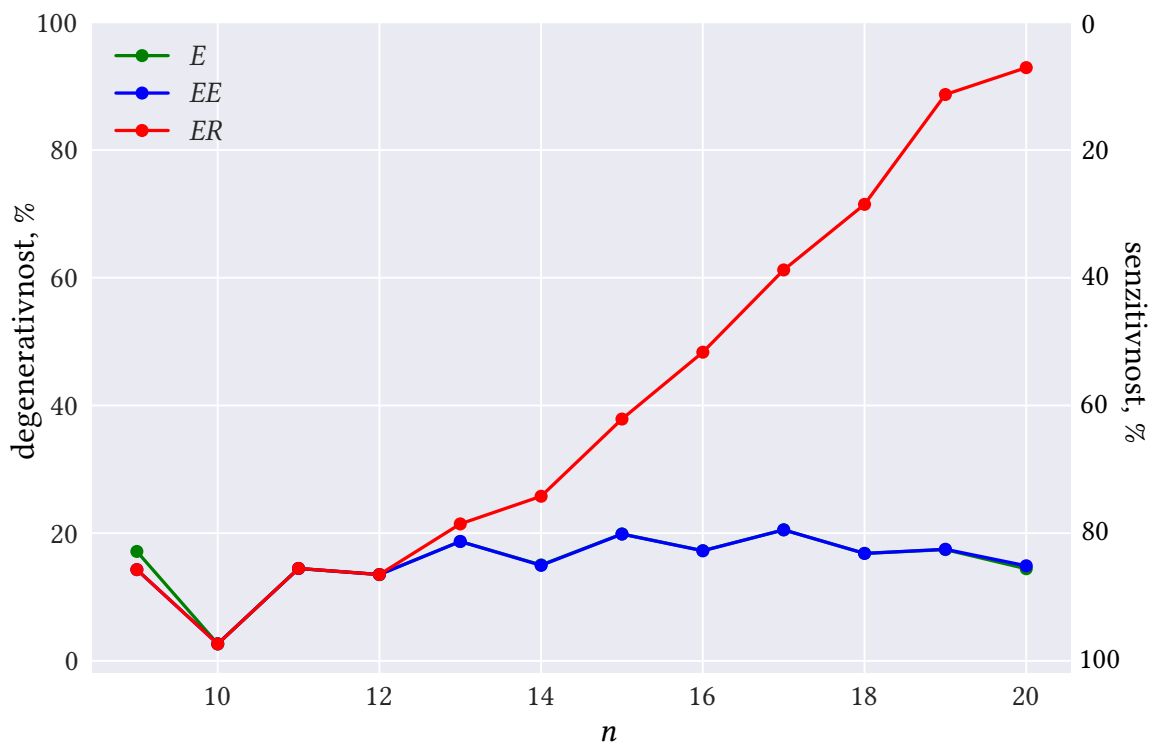
$$S = \frac{N - N_{TI}}{N} \quad (3.9)$$

gde je N ukupan broj izomera, a N_{TI} je broj izomera koje topološki indeks TI ne razlikuje. Vrednost za S se nalazi između 0 i 1. Zbog toga se često S predstavlja u procentima, ili još češće, predstavlja se procenat degenerativnosti D . Procenat degenerativnosti topološkog deskriptora se računa kao: $D = (1 - S) \cdot 100\%$. Poželjno je da deskriptor ima što veću vrednost za S , odnosno, što manju za D .

U tekstu koji sledi biće predstavljena diskriminativna moć energije grafa, Estradinog indeksa i rezolventne energije kod hemijskih stabala sa $n = 9 - 20$. Za dva hemijska stabla A i B možemo reći da imaju istu vrednost topološkog deskriptora ako je $|TI(A) - TI(B)| < 10^{-13}$. Ovaj prag je uzet zbog limita preciznosti uobičajenih proračuna u Python programima, a to je 10^{-15} . Rezultati ovog istraživanja su predstavljeni u tabeli 3.5 i na slici 3.9.

Tabela 3.5: Procenat degenerativnosti energije grafa, Estradinog indeksa i rezolventne energije

n	broj hemijskih stabala	$D(E)$	$D(EE)$	$D(ER)$
9	35	17,14%	14,29%	14,29%
10	75	2,67%	2,67%	2,65%
11	159	14,47%	14,47%	14,48%
12	355	13,52%	13,52%	13,52%
13	802	18,70%	18,70%	21,45%
14	1858	15,02%	14,96%	25,78%
15	4347	19,88%	19,85%	37,87%
16	10359	17,26%	17,24%	48,34%
17	24894	20,52%	20,52%	61,22%
18	60523	16,83%	16,83%	71,49%
19	148284	17,43%	17,48%	88,69%
20	366319	14,43%	14,89%	92,90%



Slika 3.9: Promena procenta degenerativnosti/senzitivnosti energije grafa, Estradinog indeksa i rezolventne energije kod hemijskih stabala sa porastom broja čvorova

Podaci iz tabele i sa slike pokazuju da je procenat degenerativnosti za sva tri deskriptora približno isti do hemijskih stabala sa 13 čvorova. Nakon toga energija grafa i Estradin indeks pokazuju sličan trend promene degenerativnosti sa porastom broja čvorova. Kod oba indeksa maksimalan procenat degenerativnosti iznosi oko 20%. Takođe, primetan je blagi porast diskriminativne moći ovih deskriptora kod hemijskih stabala sa velikim brojem čvorova. Sa druge strane, kod rezolventne energije primećuje se nagli porast degenerativnosti kod hemijskih stabala sa većim brojem čvorova. Iznenadujuće je da senzitivnost rezolventne energije kod hemijskih stabala sa 20 čvorova iznosi tek oko 7%. Dobijeni rezultati za rezolventnu energiju ukazuju da postoji mnoštvo hemijskih stabala sa istom vrednošću ER . Međutim, na osnovu definicije rezolventne energije (jednačina 1.27) očekivao bi se veći procenat senzitivnosti. Ovo će biti diskutovano u tekstu koji sledi.

3.2.2 r -Ekvienergetska hemijska stabla

Topološki molekularni deskriptor zasnovan na sopstvenim vrednostima ne može u potpunosti da razlikuje molekule u većini klasa izomera. Razlog ove pojave može se pripisati postojanju izospektralnih molekularnih grafova [42]. Njihovo postojanje je poznato već dugo vremena, ali otkriće da postoje neizospektralni grafovi koji imaju istu vrednost energije grafa bilo je iznenadujuće. Takvi grafovi su nazvani ekvienergetski. Pretraživanje stabala koja imaju istu energiju [169], otkrilo je da postoje i grafovi koji imaju vrlo sličnu energiju. Oni su nazvani skoro-ekvienergetski.

Rezultati ispitivanja degenerativnosti rezolventne energije ukazuju da kako raste veličina molekula, većina izomera pripada grupi molekula koji imaju istu vrednost rezolventne energije. Ovakva tendencija rezolventne energije je prilično iznenadujuća. Hemijska stabla koja imaju istu ER nazvaćemo r -ekvienergetska hemijska stabla.

Kako bi ispitali ovu sumnju, sproveli smo sistematsko pretraživanje među grafovima koji imaju između 9 i 19 čvorova tražeći r -ekvienergetska hemijska stabla. Hemijska stabla A i B su r -ekvienergetska ako je $|ER(A) - ER(B)| < 10^{-20}$. Razlog zbog koga je prag povećan na 10^{-20} je zato što su prvih pet brojeva rezolventne energije isti za sve n -izomere. Takođe, upotrebom većeg praga eliminišu se skoro- r -ekvienergetski molekuli. Kako bi izračunali rezolventnu energiju sa ovakvom preciznošću razvijen je Piton program sa implementiranim *mpmath* modulom [170]. Ovaj modul se koristi kod proračuna koji zahtevaju visoku preciznost. Rezultati su prikupljeni i prezentovani u sledećoj tabeli.

Tabela 3.6: Familije hemijskih stabala čije se rezolventne energije razlikuju za manje od 10^{-20} . # *familija* označava broj familija date veličine i tipa. *Tip* $a + b + \dots$ označava familiju koja se sastoji od a izospektralnih stabala, b izospektralna stabla nisu izospektralna sa stablima a , i tako dalje

n	# <i>familija</i>	veličina <i>familije</i>	<i>tip</i>
9	5	2	2+0
10	2	2	2+0
11	19	2	2+0
	2	3	3+0
12	38	2	2+0
	5	3	3+0
13	98	2	2+0
	26	3	3+0
14	232	2	2+0
	20	3	3+0
	2	4	4+0
15	597	2	2+0
	99	3	3+0
	16	4	4+0
	5	5	5+0
16	1245	2	2+0
	207	3	3+0
	34	4	4+0
17	3144	2	2+0
	645	3	3+0
	144	4	4+0
	40	5	5+0
	14	6	6+0
	2	7	7+0
18	6605	2	2+0
	1147	3	3+0
	236	4	4+0
	49	5	5+0
	11	6	6+0
	2	7	7+0
	1	8	8+0
	497	2	1+1
	199	3	2+1

nastavak na sledećoj strani

Tabela 3.6 – nastavak tabele

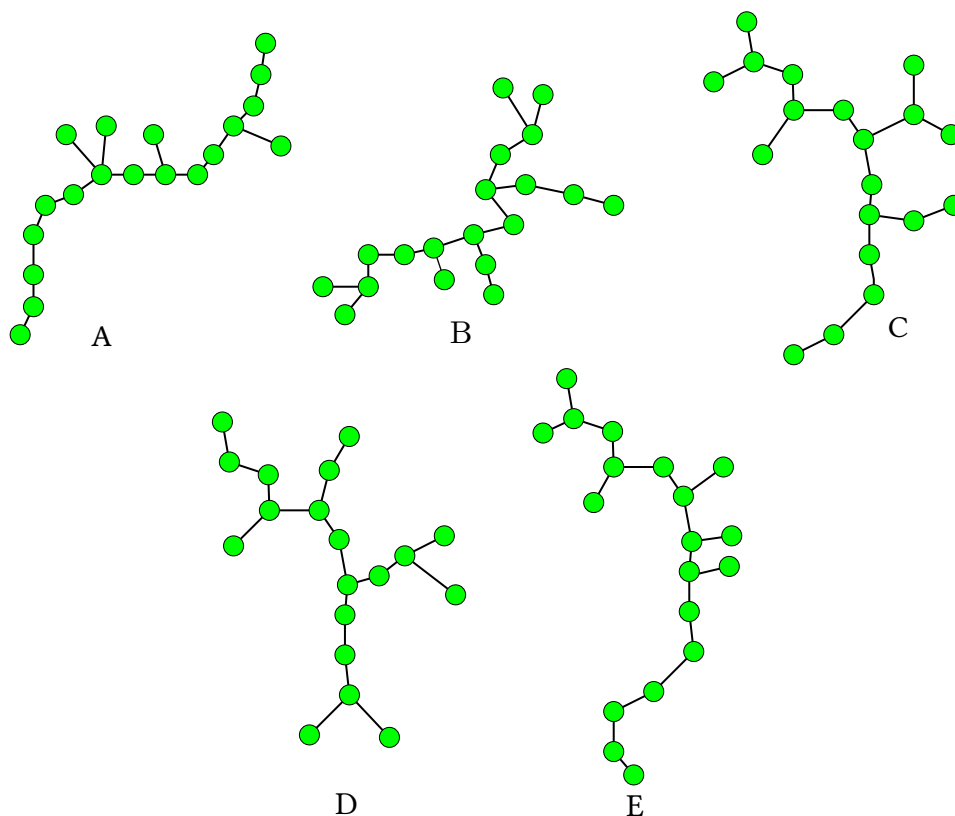
<i>n</i>	# familija	veličina familije	tip
	19	3	1+1+1
	24	4	3+1
	20	4	2+2
	9	4	2+1+1
	5	5	4+1
	6	5	3+2
	4	5	3+1+1
	1	5	2+2+1
	2	6	4+2
	1	6	3+2+1
	1	8	4+4
19	15819	2	2+0
	2601	3	3+0
	633	4	4+0
	193	5	5+0
	58	6	6+0
	14	7	7+0
	5	8	8+0
	2	9	9+0
	2680	2	1+1
	1080	3	2+1
	176	3	1+1+1
	241	4	3+1
	145	4	2+2
	129	4	2+1+1
	15	4	1+1+1+1
	56	5	4+1
	57	5	3+2
	25	5	3+1+1
	39	5	2+2+1
	15	5	2+1+1+1
	2	5	1+1+1+1+1
	18	6	5+1
	8	6	4+2
	5	6	3+3
	2	6	4+1+1
	14	6	3+2+1
	3	6	2+2+2
	10	6	2+2+1+1
	1	6	2+1+1+1+1
	1	7	6+1
	7	7	5+2
	8	7	4+3
	1	7	5+1+1
	2	7	4+2+1

nastavak na sledećoj strani

Tabela 3.6 – nastavak tabele

n	# familija	veličina familije	tip
	1	7	3+3+1
	2	7	3+2+2
	1	7	2+2+2+1
	1	7	2+2+1+1+1
	1	8	6+2
	1	8	5+3
	1	8	3+2+2+1
	1	9	5+4
	1	9	5+2+2
	1	10	7+3
	1	11	4+4+3

Evidentno je da za hemijska stabla do 17 čvorova postoje samo izospektralna hemijska stabla koja su r -ekvienergetska. Kod hemijskih stabala sa 18 čvorova postoji mnogo neizospektralnih grafova kod kojih se rezolventna energija razlikuje za manje od 10^{-20} . Broj takvih stabala je još veći kod grafova sa 19 čvorova. Pet takvih neizospektralnih hemijskih stabala čija je $ER = 1,0052992142869941494$, su prikazana na slici 3.10.



Slika 3.10: Neizospektralna hemijska stabla ($n = 19$) koja imaju istu vrednost rezolventne energije do 20 decimalnih mesta

Spektri hemijskih stabala prikazanih na slici 3.10 su:

$$Sp(A) = \{-2, 2404; -2, 0064; -1, 7806; -1, 6064; -1, 2162; -1, 1216; -0, 5564; -0, 3969; 0; 0; 0; 0, 3969; 0, 5564; 1, 1216; 1, 2162; 1, 6064; 1, 7806; 2, 0064; 2, 2404\},$$

$$Sp(B) = \{-2, 2406; -2, 0043; -1, 7907; -1, 5877; -1, 2656; -1, 0705; -0, 5966; -0, 3625; 0; 0; 0; 0, 3625; 0, 5966; 1, 0705; 1, 2656; 1, 5877; 1, 7907; 2, 0043; 2, 2406\},$$

$$Sp(C) = \{-2, 2412; -1, 9953; -1, 8236; -1, 4820; -1, 4142; -0, 9268; -0, 7357; -0, 2713; 0; 0; 0; 0, 2713; 0, 7357; 0, 9268; 1, 4142; 1, 4820; 1, 8236; 1, 9953; 2, 2412\},$$

$$Sp(D) = \{-2, 2409; -2; -1, 8082; -1, 5473; -1, 3331; -1; -0, 6631; -0, 3125; 0; 0; 0; 0, 3125; 0, 6631; 1; 1, 3331; 1, 5473; 1, 8082; 2; 2, 2409\},$$

$$Sp(E) = \{-2, 2407; -2, 0022; -1, 7998; -1, 5684; -1, 3008; -1, 0336; -0, 6307; -0; 3357; 0; 0; 0; 0, 3357; 0, 6307; 1, 0336; 1, 3008; 1, 5684; 1, 7998; 2, 0022; 2, 2407\}.$$

Otkriće neizospektralnih hemijskih stabala sa istom rezolventnom energijom bilo bi veoma značajno. Zbog toga smo izračunali ER sa većom preciznošću za hemijska stabla sa 18 i 19 čvorova. Za grafove sa 18 čvorova pronađeno je da su sva r -ekvienergetska stabla izospektralna, računato na 23 decimale. Ista je situacija kod grafova sa 19 čvorova za računatu rezolventnu energiju na 24 decimale.

3.3 Strukturna osetljivost molekulskih deskriptora zasnovanih na sopstvenim vrednostima

Strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije ispitivana je na benzenoidnim ugljovodonicima, koristeći metodologiju predstavljenu u odeljku 1.4. Ova jedinjenja su poznati zagađivači i predmet su različitih studija [171–174]. Rezultati koji će biti predstavljeni u ovom odeljku su publikovani u radu [175].

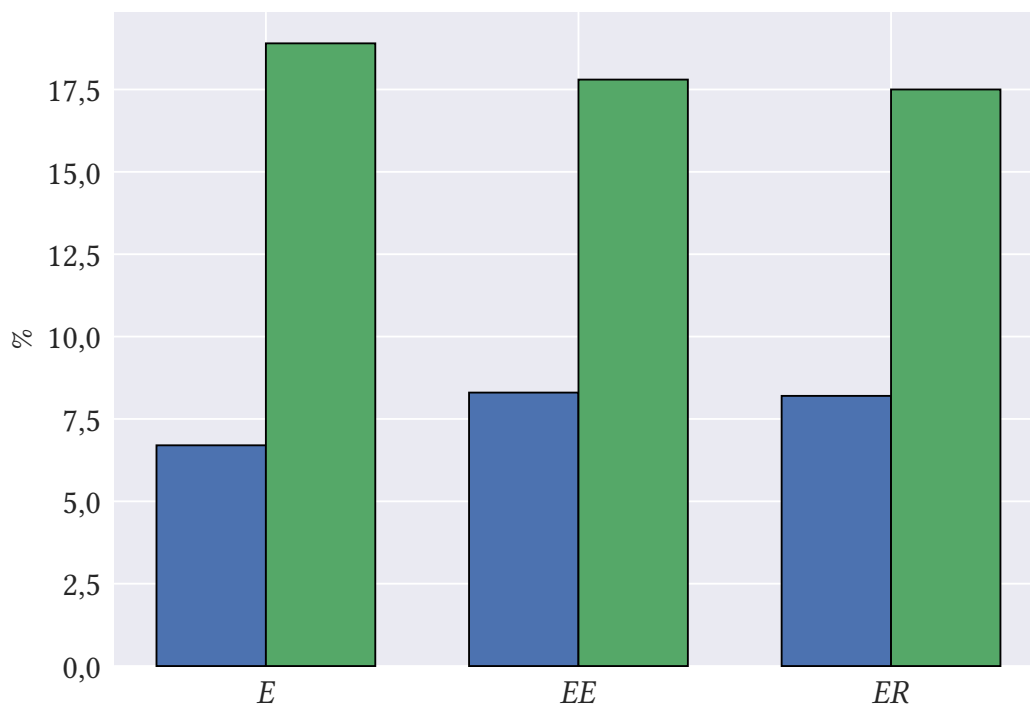
3.3.1 Katakondenzovani i perikondenzovani benzenoidni ugljovodoni

Prvi zadatak sastojao se u tome da ispitamo da li ima razlike u strukturnoj osetljivosti E , EE i ER u odnosu na katakondenzovane i perikondenzovane benzenoidne ugljovodonike. Zbog toga, konstruisane su po tri serije katakondenzovanih i perikondenzovanih izomera, gde je $h=8, 9$ i 10 , koristeći kompjuterski program CaGe [165]. Ovi skupovi se sastoje od 411, 1489 i 5572 katakondenzovana izomera i 8, 12 i 25 perikondenzovana izomera. Razlog za veliku razliku u broju katakondenzovanih i perikondenzovanih jeste u tome što je kod perikondenzovanih broj unutrašnjih čvorova ograničen na 6, 7 i 8. Ograničavanjem broja unutrašnjih čvorova ograničava se broj izomera. Vrednosti za \overline{SS} i \overline{Abr} energije grafa, Estradinog indeksa i rezolventne energije date su u tabeli 3.7.

Tabela 3.7: Strukturna osetljivost E , EE i ER kod katakondenzovanih i perikondenzovanih benzenoidnih ugljovodonika

TI		$h=8$	$h=9$	$h=10$
Katakondenzovani molekuli				
E	\overline{SS}	8,0%	7,0%	6,7%
	\overline{Abr}	17,0%	17,1%	18,9%
EE	\overline{SS}	9,0%	8,6%	8,3%
	\overline{Abr}	16,2%	18,5%	17,8%
ER	\overline{SS}	9,0%	8,5%	8,2%
	\overline{Abr}	15,8%	18,1%	17,5%
Perikondenzovani molekuli				
E	\overline{SS}	25,4%	14,0%	24,9%
	\overline{Abr}	30,2%	18,5%	37,7%
EE	\overline{SS}	10,3%	0,5%	7,9%
	\overline{Abr}	12,9%	0,6%	15,2%
ER	\overline{SS}	9,9%	0%	7,7%
	\overline{Abr}	12,5%	0%	15,0%

Rezultati prezentovani u tabeli 3.7, u slučaju katakondenzovanih molekula, pokazuju da Estradin indeks i rezolventna energija ispoljavaju sličnu strukturnu senzitivnost. Naime, njihove \overline{SS} vrednosti se razlikuju za manje od 0,1%. Energija grafa pokazuje nešto slabiju strukturnu senzitivnost u odnosu na EE i ER . Kod sve tri grupe katakondenzovanih izomera rezolventna energija ima najmanje $|\overline{SS} - \overline{Abr}|$ vrednosti. Ovaj rezultat ukazuje na to da, kod ove grupe molekula, rezolventna energija pokazuje najbolje performanse. Takođe, evidentno je da strukturna osetljivost sva tri topološka deskriptora opada sa povećanjem broja izomera. Ovaj trend

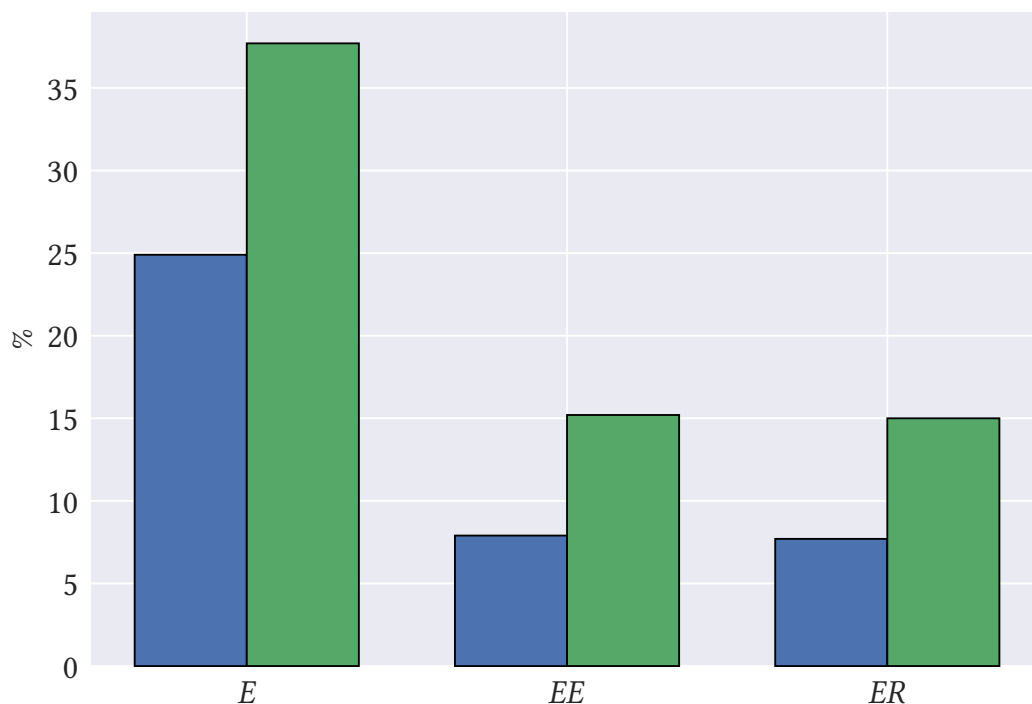


Slika 3.11: Strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije kod katakondenzovanih molekula sa 10 heksagona. Plavi stubići predstavljaju \overline{SS} , dok zeleni stubići predstavljaju \overline{Abr}

je praćen povećanjem njihovih \overline{Abr} vrednosti. Na slici 3.11 je prikazana strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije kod katakondenzovanih molekula sa 10 heksagona. U ovoj grupi rezolventna energija ima najmanje vrednosti za $|\overline{SS} - \overline{Abr}|$, dok energija grafa ima najveće.

Kod perikondenzovanih molekula je sasvim suprotna situacija. Naime, kod ovih molekula najveći nivo strukturne osetljivosti pokazuje energija grafa ($\overline{SS}_{max}=25,4\%$), dok Estradin indeks i rezolventna energija pokazuju znatno niži nivo osetljivosti. Ovo pogoršanje senzitivnosti ide do te mere da kod molekula sa $h=9$ rezolventna energija ne oseća strukturne detalje ($\overline{SS}=0$), dok Estradin indeks ispoljava ekstremno nizak nivo osetljivosti ($\overline{SS}<1\%$). Iako E pokazuje veću strukturnu osetljivost, njene $|\overline{SS} - \overline{Abr}|$ vrednosti su veće u odnosu na te vrednosti Estradinog indeksa i rezolventne energije. Kao i u prethodnom slučaju, sa porastom broja izomera strukturna osetljivost sva tri deskriptora opada. Na slici 3.12 je prikazana strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije kod perikondenzovanih molekula sa 10 heksagona.

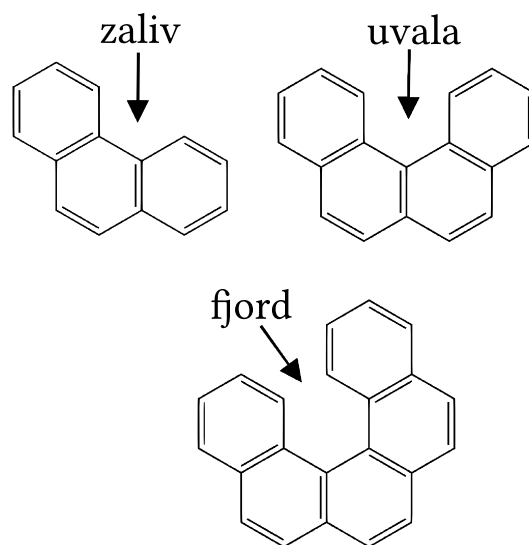
Slika 3.12 oslikava superiornost E u odnosu na EE i ER kod perikondenzovanih molekula. Poređenjem strukturnih osetljivosti tri topološka molekulska deskriptora kod oba tipa molekula, očigledno je da Estradin indeks i rezolventna energija pokazuju vrlo slično ponašanje. Njihove \overline{SS} vrednosti ne prelaze 11%. Strukturna osetljivost energije grafa je značajno veća kod perikondenzovanih benzenoidnih ugljovodonika. Zanimljiv rezultat je dobijen za sva tri deskriptora kod perikondenzovanih molekula. Naime, vrednosti za strukturnu osetljivost su značajno manje kod molekula sa 9 heksagona. Za rezolventnu energiju \overline{SS} vrednost je jednaka nuli.



Slika 3.12: Strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije kod perikondenzovanih molekula sa 10 heksagona. Plavi stubići predstavljaju \overline{SS} , dok zeleni stubići predstavljaju \overline{Abr}

3.3.2 Broj zaliva, uvala i fjordova

Kako bi dalje ispitali strukturnu osetljivost energije grafa, Estradinog indeksa i rezolventne energije konstruisane su tri serije katakondenzovanih molekula. U ovim serijama broj heksagona je konstantan i iznosi 20. U prvoj seriji povećavan je broj zaliva (**B**), u drugoj broj uvala (**C**), a u trećoj broj fjordova (**F**). Zaliv, uvala i fjord su strukturni detalji katakondenzovanih i perikondenzovanih molekula koji određuju oblik molekula. Prikazani su na slici 3.13.

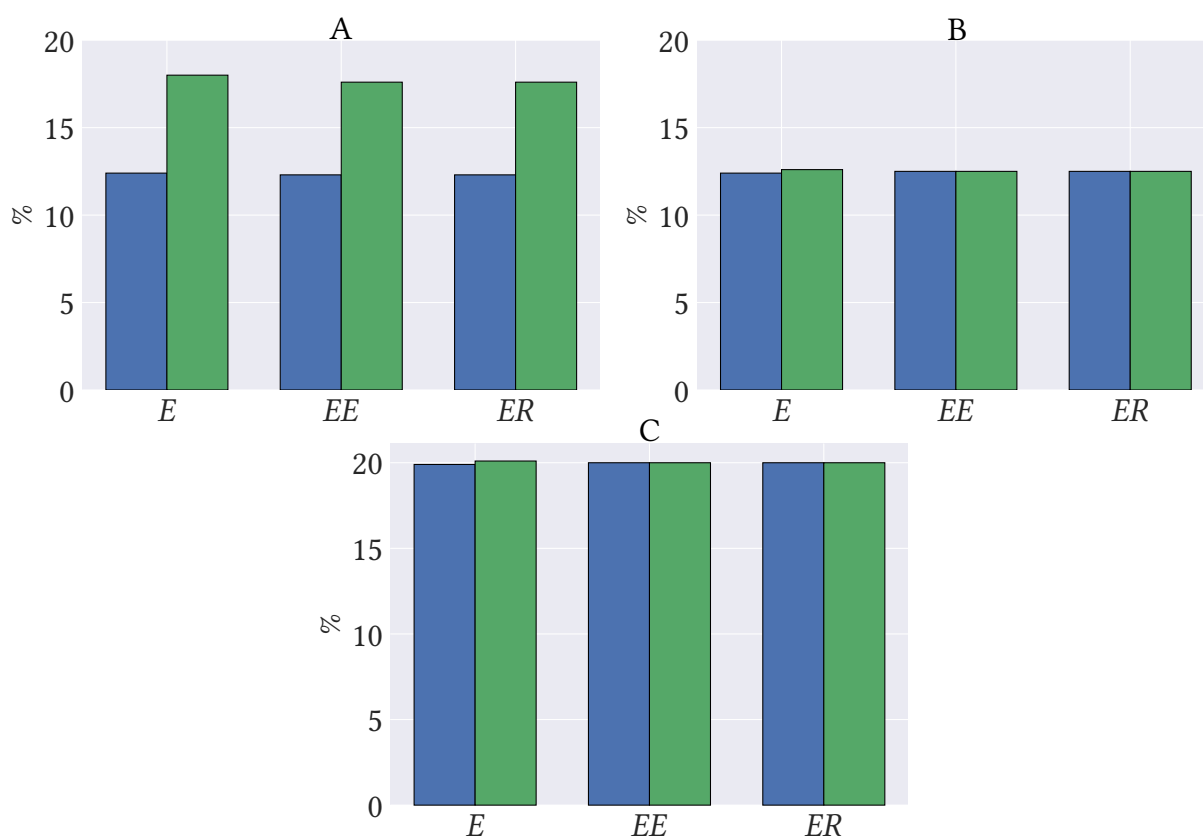


Slika 3.13: Strukturni detalji katakondenzovanih benzenoidnih ugljovodonika

Molekulski grafovi u ovim serijama su konstruisani tako da svaka veza pripada samo jednom zalivu, jednoj uvali i jednom fjordov. Vrednosti za **B**, **C** i **F** se kreću u opsegu od 1 do 18, 9 i 6, respektivno. Rezultati testiranja strukturne osetljivosti *E*, *EE* i *ER* na ovim molekulima su predstavljeni u tabeli 3.8 i na slici 3.14.

Tabela 3.8: Strukturna senzitivnost *E*, *EE* i *ER* kod katakondenzovanih molekula sa različitim brojem zaliva, uvala i fjordova

<i>TI</i>		B	C	F
<i>E</i>	\overline{SS}	12,4%	12,4%	19,9%
	\overline{Abr}	18,0%	12,6%	20,1%
<i>EE</i>	\overline{SS}	12,3%	12,5%	20,0%
	\overline{Abr}	17,6%	12,5%	20,0%
<i>ER</i>	\overline{SS}	12,3%	12,5%	20,0%
	\overline{Abr}	17,6%	12,5%	20,0%



Slika 3.14: Strukturna senzitivnost *E*, *EE* i *ER* kod katakondenzovanih molekula sa različitim brojem zaliva (A), uvala (B) i fjordova (C). Plavi stubići predstavljaju \overline{SS} , dok zeleni stubići predstavljaju \overline{Abr}

Rezultati predstavljeni u tabeli 3.8 pokazuju da Estradin indeks i rezolventna energija ispoljavaju isti nivo strukturne osetljivosti. Energija grafa pokazuje zanemarljivu razliku od 0.1%. Ovakvo ponašanje Estradinog indeksa i rezolventne energije može se pripisati dobroj korelaciji između ova dva deskriptora [162, 163]. Sva tri molekulska deskriptora pokazuju najveći nivo senzitivnosti kod molekula sa različitim brojem fjordova.

\overline{Abr} vrednosti iz tabele 3.8 su veoma blizu ili su jednake \overline{SS} vrednostima. Ovakav rezultat ukazuje na to da bi energija grafa, Estradin indeks i rezolventna energija trebalo postepeno da menjaju svoju vrednost sa promenom broja fjordova. Ovo se posebno očekuje kod **C** i **F** serija molekula, gde su \overline{Abr} vrednosti EE i ER jednake njihovim \overline{SS} vrednostima. Što se tiče energije grafa, njene $|\overline{SS} - \overline{Abr}|$ vrednosti su zanemarljivo male i iznose 0,2%.

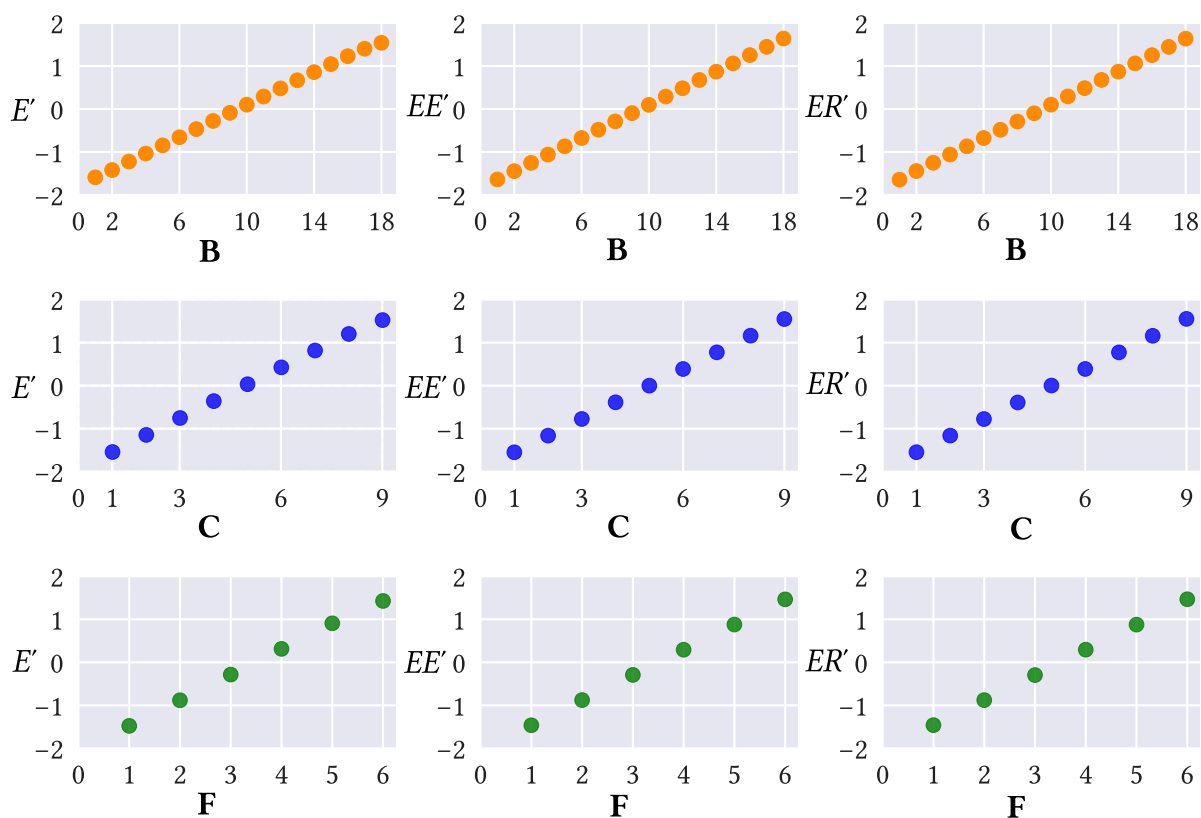
Kako bi dobili bolji uvid u ovaj rezultat, korelirali smo izračunate vrednosti energije grafa, Estradinog indeksa i rezolventne energije sa **B**, **C** i **F**. Pošto vrednosti ovih deskriptora variraju u različitim opsezima, standardizacija njihovih vrednosti izvršena je upotrebom sledeće formule:

$$TI'_i = \frac{TI_i - \overline{TI}}{s(TI)}. \quad (3.10)$$

U ovoj jednačini \overline{TI} predstavlja srednju vrednost indeksa, dok $s(TI)$ označava standardnu devijaciju. Na ovaj način dobili smo standardizovanu energiju grafa (E'), standardizovani Estradin indeks (EE') i standardizovanu rezolventnu energiju (ER'). Rezultati su predstavljeni u tabeli 3.9 i na slici 3.15.

Tabela 3.9: Koeficijenti korelacije između standardizovane energije grafa, standardizovanog Estrading indeksa, i standardizovane rezolventne energije i **B**, **C** i **F**

TI'	B	C	F
E'	0,99869	0,99980	0,99976
EE'	0,99927	1,00000	1,00000
ER'	0,99925	1,00000	1,00000



Slika 3.15: Korelacija između E' , EE' i ER' i **B**, **C** i **F**

3.4 Efekat cikla na vrednost molekulskih deskriptora zasnovanih na sopstvenim vrednostima

Koristeći izraze predstavljene u poglavlju 1.5 moguće je izraziti uticaj cikla na vrednost energije grafa, Estradinog indeksa i rezolventne energije. U ovom odeljku će biti predstavljeni rezultati tih ispitivanja, a koji su publikovani u radu [176].

3.4.1 Uticaj veličine cikla

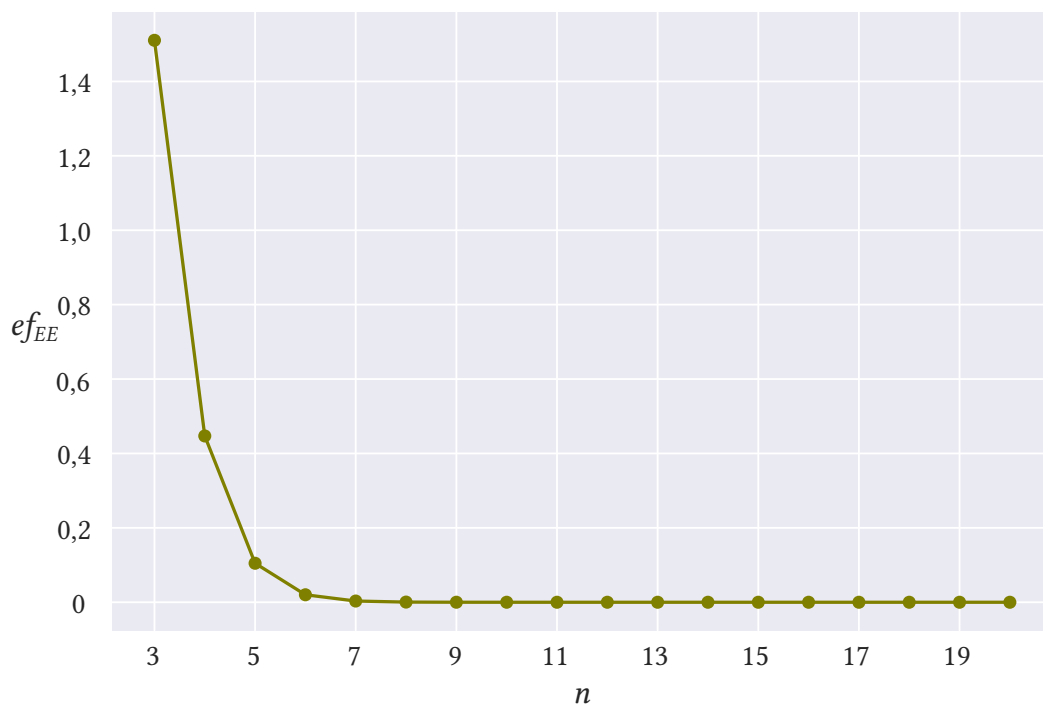
Kako bi ispitali uticaj veličine cikla na vrednosti ef_E , ef_{EE} i ef_{ER} osmišljen je sledeći eksperiment. Početni graf je cikl koji se sastoji od 20 čvorova. Na taj način je ukupan broj čvorova držan konstantnim. On sadrži još jednu vezu koja povezuje dva čvora. Ta veza formira još jedan cikl čija se veličina u svakoj iteraciji povećava za jedan čvor. Veličina novonastalog cikla se kreće od 3 do 20 čvorova.

Rezultati uticaja veličine ovakvog cikla na vrednosti ef_E , ef_{EE} i ef_{ER} su prikazani u tabeli 3.10.

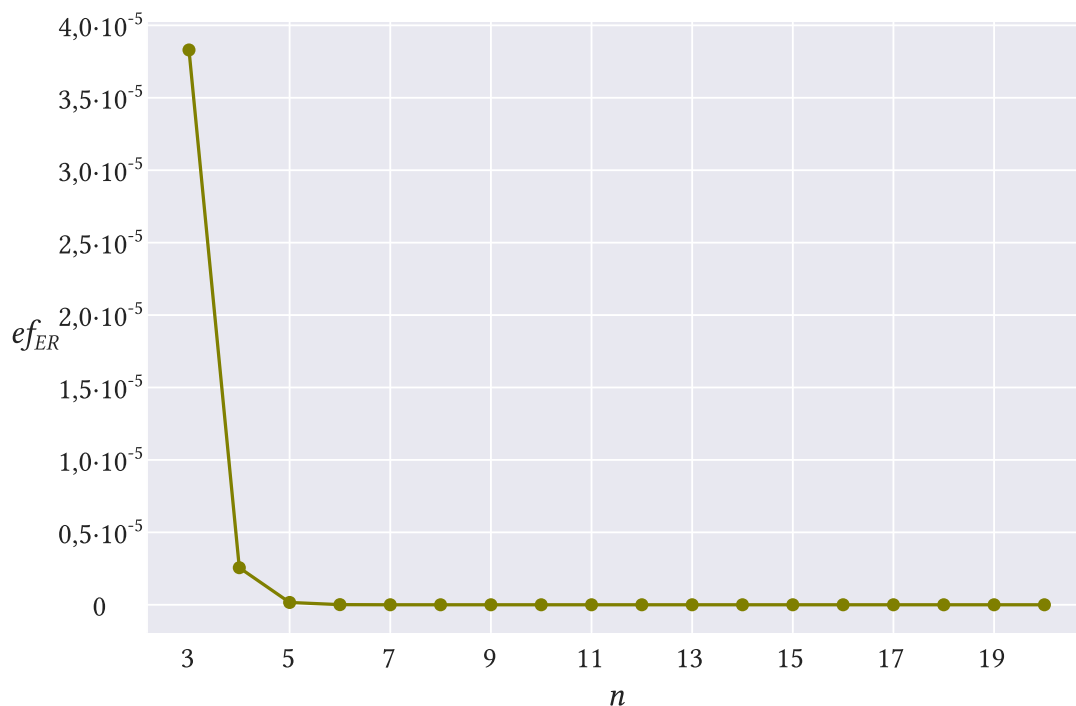
Tabela 3.10: Vrednosti efekta cikla na energiju grafa, Estradin indeks i rezolventnu energiju za ciklove veličine od 3 do 20 čvorova

n	ef_E	ef_{EE}	ef_{ER}
3	0,19701	1,51049	$3,83 \cdot 10^{-5}$
4	-0,41860	0,44700	$2,56 \cdot 10^{-6}$
5	0,12766	0,10483	$1,60 \cdot 10^{-7}$
6	0,26441	0,02012	$9,63 \cdot 10^{-9}$
7	0,09584	0,00326	$5,63 \cdot 10^{-10}$
8	-0,17942	0,00046	$3,22 \cdot 10^{-11}$
9	0,07505	$5,60 \cdot 10^{-5}$	$1,82 \cdot 10^{-12}$
10	0,19042	$6,14 \cdot 10^{-6}$	$1,01 \cdot 10^{-13}$
11	0,05920	$6,08 \cdot 10^{-7}$	$5,33 \cdot 10^{-15}$
12	-0,11716	$5,48 \cdot 10^{-8}$	$6,66 \cdot 10^{-16}$
13	0,04593	$4,53 \cdot 10^{-9}$	0
14	0,16718	$3,47 \cdot 10^{-10}$	0
15	0,03404	$2,48 \cdot 10^{-11}$	0
16	-0,09000	$1,48 \cdot 10^{-12}$	0
17	0,02277	$1,56 \cdot 10^{-13}$	0
18	0,16534	$-9,95 \cdot 10^{-14}$	0
19	-0,10270	$-4,26 \cdot 10^{-14}$	0
20	-0,23598	$-1,42 \cdot 10^{-14}$	0

Dobijeni podaci jasno pokazuju trend eksponencijalnog opadanja vrednosti ef_{EE} i ef_{ER} sa porastom veličine cikla. Takođe, može se zaključiti da je uticaj veličine prstena na ef_{EE} i ef_{ER} zanemarljiv kod ciklova čiji je broj čvorova veći od 6. Štaviše, zanimljivo je da se uticaj veličine prstena kod rezolventne energije potpuno gubi kod prstenova čija je veličina jednaka 13 i više čvorova. Ovaj rezultat je dobijen i kada je preciznost proračuna povećana na 50 decimala. Grafički prikaz dobijenih rezultata za ef_{EE} i ef_{ER} dat je na slikama 3.16 i 3.17.



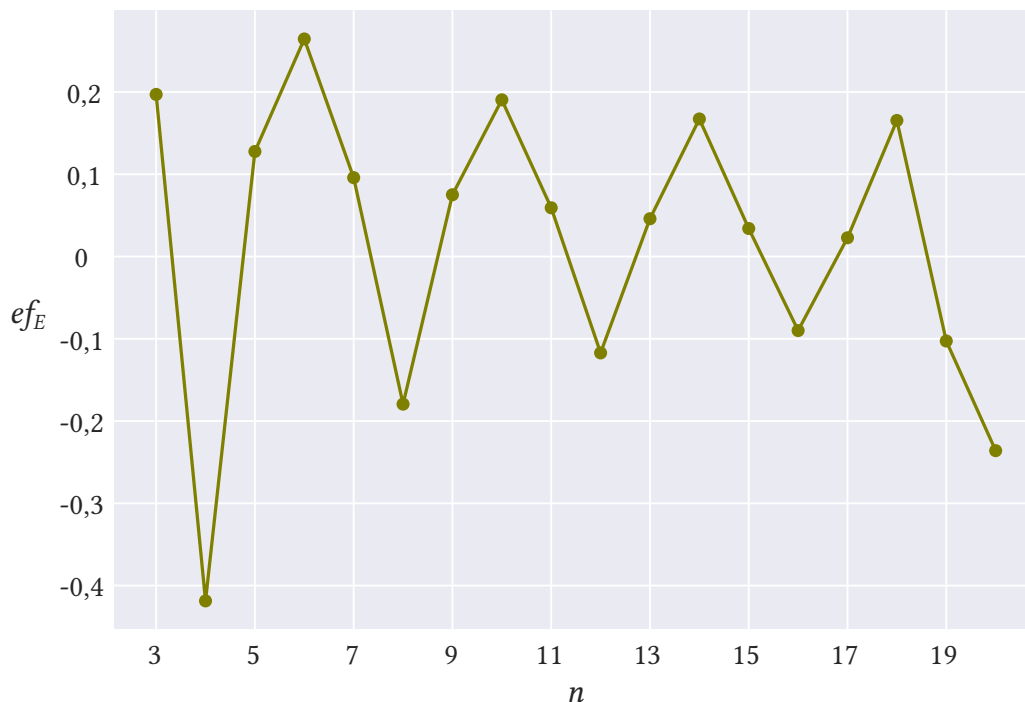
Slika 3.16: Grafik zavisnosti ef_{EE} od veličine cikla



Slika 3.17: Grafik zavisnosti ef_{ER} od veličine cikla

Uticaj veličine cikla na ef_E pokazuje drugačiji trend u odnosu na ef_{EE} i ef_{ER} . Naime, dobijaju se negativne vrednosti za svaki $4k$ -cikl i pozitivne vrednosti za $4k + 2$ -ciklove. Ovo je u saglasnosti sa proširenim Hikelovim pravilom o uticaju veličine prstena na stabilnost konjugovanog molekula. Takav fizičko-hemijski smisao se ne može primetiti kod ef -vrednosti

Estradinog indeksa i rezolventne energije. Zavisnost ef_E od veličine cikla prikazana je na slici 3.18.



Slika 3.18: Grafik zavisnosti ef_E od veličine cikla

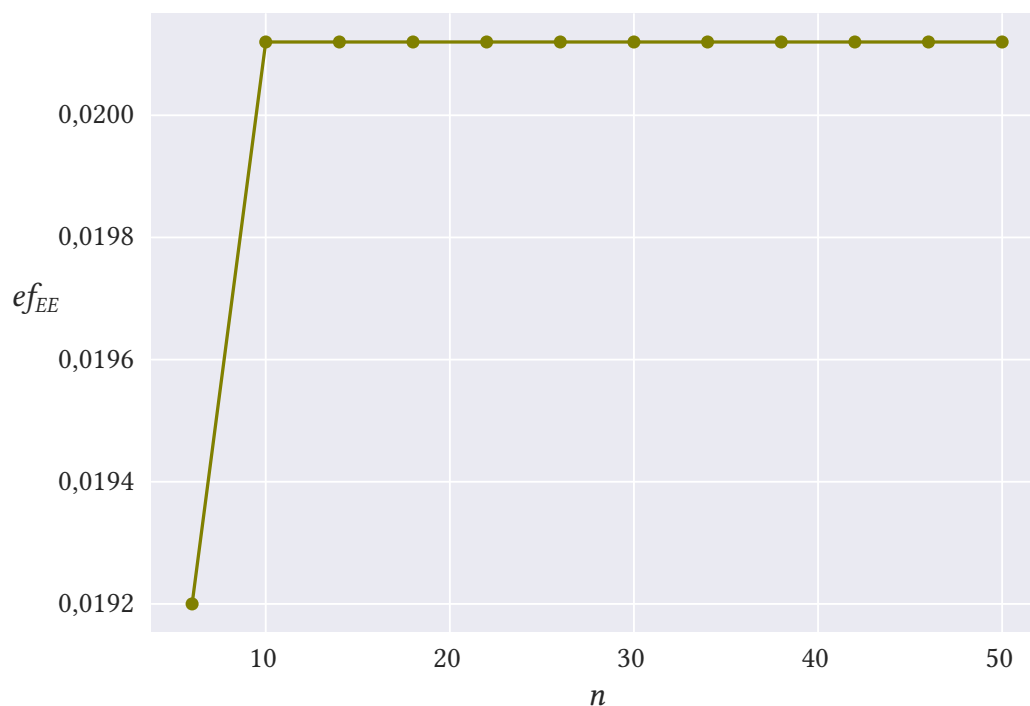
3.4.2 Uticaj veličine molekula

Uticaj veličine molekula na ef_E , ef_{EE} i ef_{ER} ispitan je na seriji molekula poliacena. Veličina početnog cikla (benzenovog prstena) je držana konstantnom dok je u svakoj iteraciji molekul rastao. Odnosno, ispitan je uticaj terminalnog šestočlanog prstena u seriji molekula od benzena ($n = 6$) do dodekacena ($n = 50$). Rezultati ovog ispitivanja su predstavljeni u tabeli 3.11.

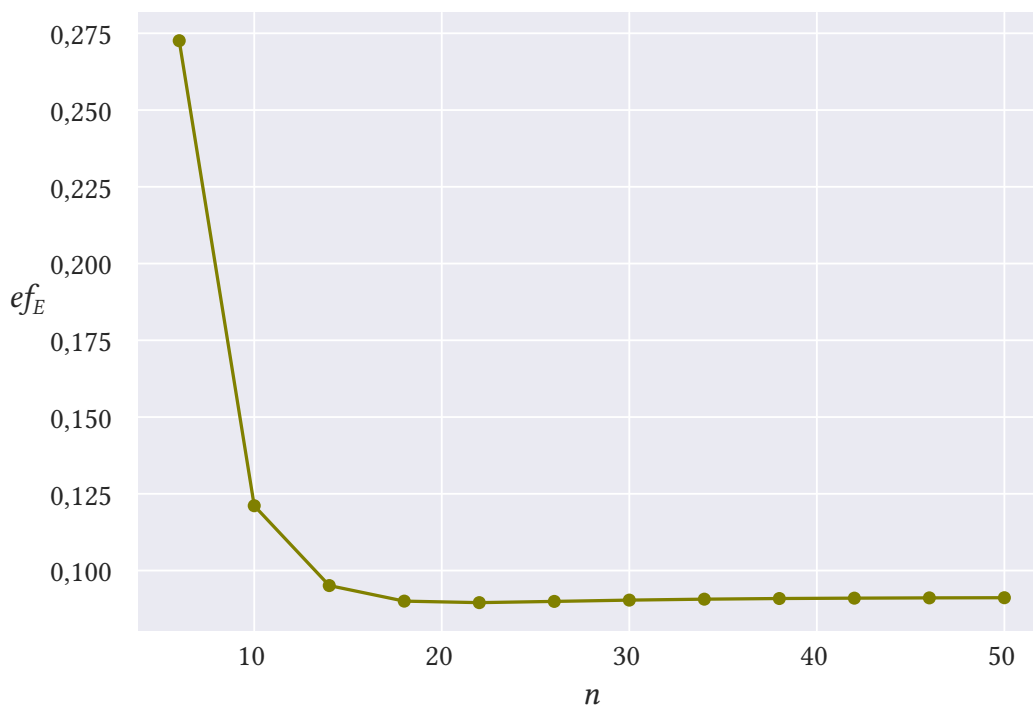
Tabela 3.11: Efekat terminalnog benzenovog prstena na vrednosti energije grafa, Estradinog indeksa i rezolventne energije u klasi poliacena do 50 ugljenikovih atoma

n	ef_E	ef_{EE}	ef_{ER}
6	0,27259	0,01920	$5,41 \cdot 10^{-5}$
10	0,12111	0,02012	$1,34 \cdot 10^{-6}$
14	0,09511	0,02012	$1,21 \cdot 10^{-7}$
18	0,09004	0,02012	$2,03 \cdot 10^{-8}$
22	0,08955	0,02012	$4,92 \cdot 10^{-9}$
26	0,08995	0,02012	$1,52 \cdot 10^{-9}$
30	0,09037	0,02012	$5,55 \cdot 10^{-10}$
34	0,09068	0,02012	$2,31 \cdot 10^{-10}$
38	0,09089	0,02012	$1,06 \cdot 10^{-10}$
42	0,09101	0,02012	$5,24 \cdot 10^{-11}$
46	0,0911	0,02012	$2,77 \cdot 10^{-11}$
50	0,09115	0,02012	$1,54 \cdot 10^{-11}$

Iz tabele 3.11 se može videti da efekat prstena na vrednost Estradinog indeksa praktično ne zavisi od veličine molekula. Odnosno, ovaj efekat raste kada se broj ugljenikovih atoma sa 6 poveća na 10 i posle toga se više ne menja.

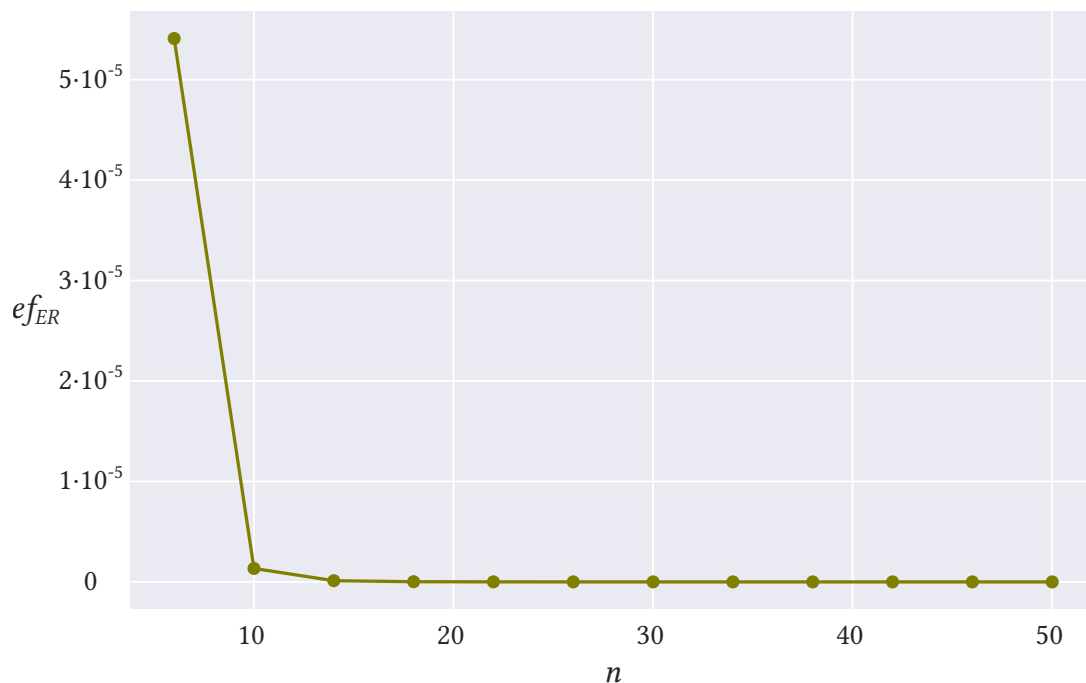


Slika 3.19: Zavisnost ef_{EE} od veličine molekula



Slika 3.20: Zavisnost ef_E od veličine molekula

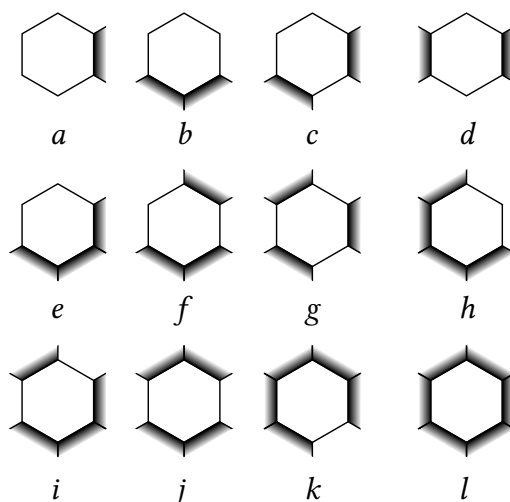
Što se tiče uticaja cikla na energiju grafa i rezolventnu energiju tu je situacija drugačija. Naime, vrednost ef_{ER} opada u svakoj iteraciji, dok ef_E dostiže minimum kod molekula sa 22 ugljenikova atoma, a potom blago raste. Grafički prikaz rezultata je dat na slikama 3.19, 3.20 i 3.21.



Slika 3.21: Zavisnost ef_{ER} od veličine molekula

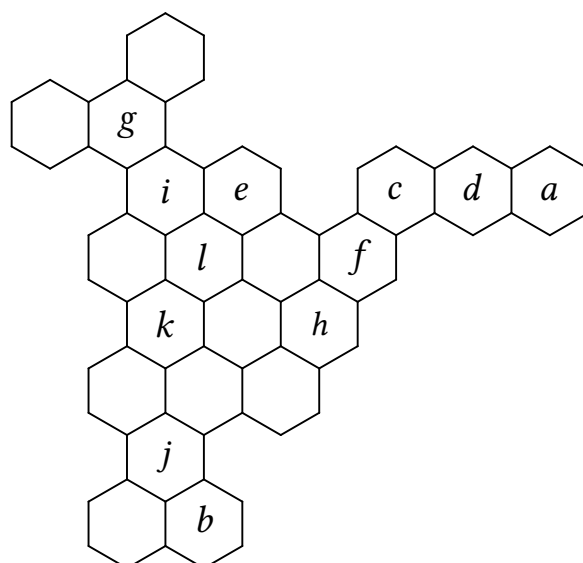
3.4.3 Uticaj tipa anelacije

Treći eksperiment je dizajniran tako da se ispita uticaj tipa anelacije prstena na vrednosti ef_E , ef_{EE} i ef_{ER} , ali u isto vreme da se veličina cikla i veličina molekula ne menjaju. Ispitivan je uticaj benzenovog prstena. Kod policikličnih aromatičnih jedinjenja postoji 12 načina anelacije benzenovog prstena i oni su prikazani na slici 3.22.



Slika 3.22: Tipovi anelacije benzenovog prstena

Ideja je da konstruišemo benzenoidni molekul koji sadrži sve tipove benzenoidnih prstenova, prikazanih na slici 3.22, za određivanje uticaja na vrednosti ef_E , ef_{EE} i ef_{ER} . Na slici 3.23 je prikazan ovakav molekul. Ciklovi čiji je uticaj ispitivan označeni su istim slovima kao na slici 3.22, i rezultati su predstavljeni u tabeli 3.12.



Slika 3.23: Benzenoidni ugljovodoni koji sadrži sve tipove prstenova prikazanih na slici 3.22

Tabela 3.12: Vrednosti ef_E , ef_{EE} i ef_{ER} za sve tipove anelacije benzenovog prstena. Vrednosti za ef_{ER} su pomnožene sa 10^{12}

anelacija	ef_E	ef_{EE}	$ef_{ER} \cdot 10^{12}$
<i>a</i>	0,107	0,02	1,199
<i>b</i>	0,103	0,02	1,199
<i>c</i>	0,043	0,021	1,201
<i>d</i>	0,08	0,021	1,201
<i>e</i>	0,091	0,021	1,201
<i>f</i>	0,067	0,022	1,201
<i>g</i>	0,034	0,022	1,201
<i>h</i>	0,044	0,022	1,201
<i>i</i>	0,043	0,022	1,201
<i>j</i>	0,028	0,022	1,201
<i>k</i>	0,032	0,022	1,201
<i>l</i>	0,032	0,022	1,201

Rezultati pokazuju da efekat prstena na vrednosti Estradinog indeksa i rezolventne energije ne zavisi od tipa anelacije. S druge strane, efekat na energiju grafa u velikoj meri zavisi od tipa anelacije benzenovog prstena. Ovo se može objasniti Klarovim strukturama ispitivanog benzenoidnog ugljovodnika. Naime, anelacije benzena koje sadrže više π -elektrona u Klarovim strukturama trebalo bi da imaju veći energetski efekat, i obrnuto.

3.5 Predikcioni potencijal molekulskih deskriptora zasnovanih na sopstvenim vrednostima

Važna karakteristika potencijalno korisnog molekuskog deskriptora jeste dobra korelacija sa nekom fizičko–hemijskom osobinom molekula. Dobra korelacija je znak da se taj deskriptor može koristiti za konstruisanje modela za modeliranje tih fizičko–hemijskih osobina. Zbog toga, u ovom odeljku biće predstavljeni rezultati koji se odnose na predviđanje tačke ključanja (TK), entalpije obrazovanja ($\Delta_f H^\circ$) i logaritma koeficijenta oktanol/voda ($\log P$) alkana, pomoću energije grafa, Estradinog indeksa i rezolventne energije. Ovi rezultati su publikovani u radu [177].

Tačka ključanja i entalpija obrazovanja predstavljaju dve često određivane osobine molekula. Naime, studije koje koriste druge topološke indekse za predviđanje tačke ključanja i entalpija obrazovanja alkana su već izvođene [178–180]. Koeficijent oktanol/voda je veoma bitan u savremenom dizajniranju lekova, i do danas ne postoji teorijski dobijena veličina koja bi mogla da ga uspešno zameni.

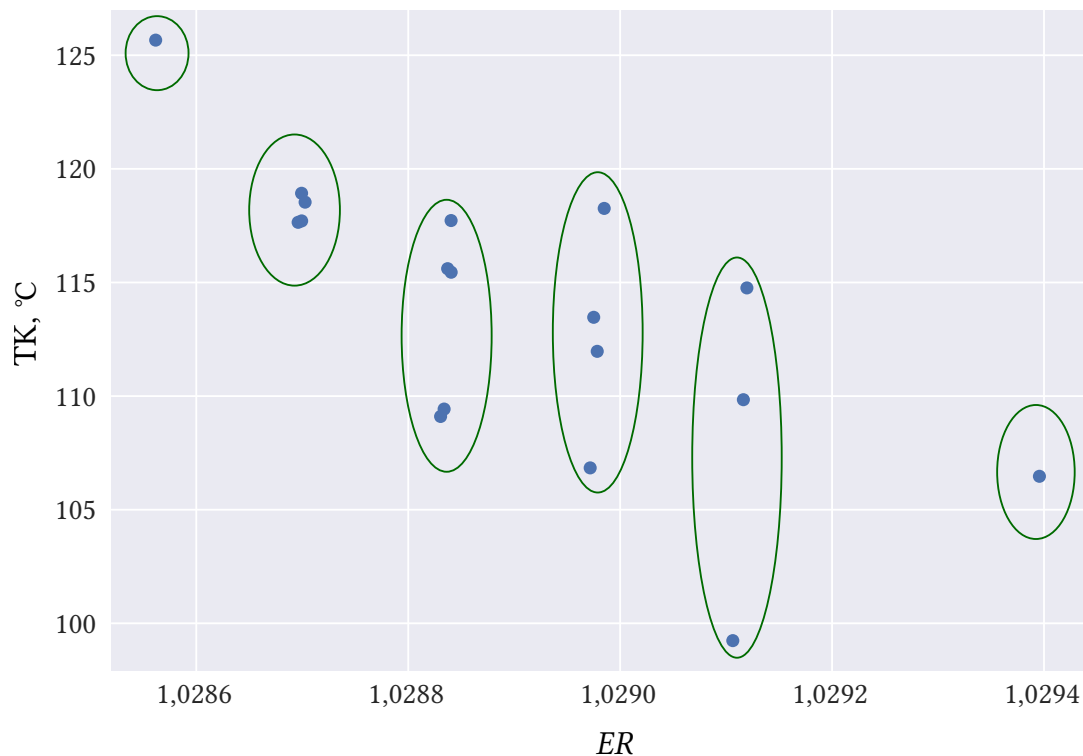
Za razvijanje linearnih modela za predviđanje ovih osobina koristili smo eksperimentalne podatke za izomere oktana. Ovi podaci su preuzeti sa internet stranice [181] i predstavljeni su u tabeli 3.13. Korišćenjem podataka za izomere izbegli smo uticaj veličine molekula, i fini strukturni detalji koji utiču na osobine su izraženi.

Tabela 3.13: Eksperimentalni podaci za izomere oktana

Molekul	TK, °C	$\Delta_f H^\circ$, kJ mol ⁻¹	$\log P$
2,2,4–trimetilpentan	99,2	-224,1	3,58
2,2,3,3–tetrametilbutan	106,5	-225,9	3,62
2,2,3–trimetilpentan	109,8	-220,1	3,58
2,2–dimetilheksan	106,8	-224,7	3,65
3,3–dimetilheksan	112,0	-220,1	3,65
2,3,3–trimetilpentan	114,8	-216,4	3,58
2,5–dimetilheksan	109,1	-222,6	3,54
2,4–dimetilheksan	109,4	-219,4	3,54
2–metilheptan	117,6	-215,5	3,61
4–metilheptan	117,7	-212,1	3,61
2,3–dimetilheksan	115,6	-213,9	3,54
2,3,4–trimetilpentan	113,5	-217,4	3,48
3–etil–3–metilpentan	118,3	-215,0	3,65
3,4–dimetilheksan	117,7	-213,0	3,54
3–metilheptan	118,9	-212,6	3,61
3–etilheksan	118,5	-210,9	3,61
3–etil–2–metilpentan	115,5	-211,2	3,54
<i>n</i> –oktan	125,7	-208,4	3,67

3.5.1 Tačka ključanja i entalpija obrazovanja

Izračunate vrednosti energije grafa, Estradinog indeksa i rezolventne energije oktana korelirali smo sa eksperimentalnim vrednostima za TK iz tabele 3.13. Kao primer, na slici 3.24 prikazana je korelacija između rezolventne energije i tački ključanja izomera oktana.



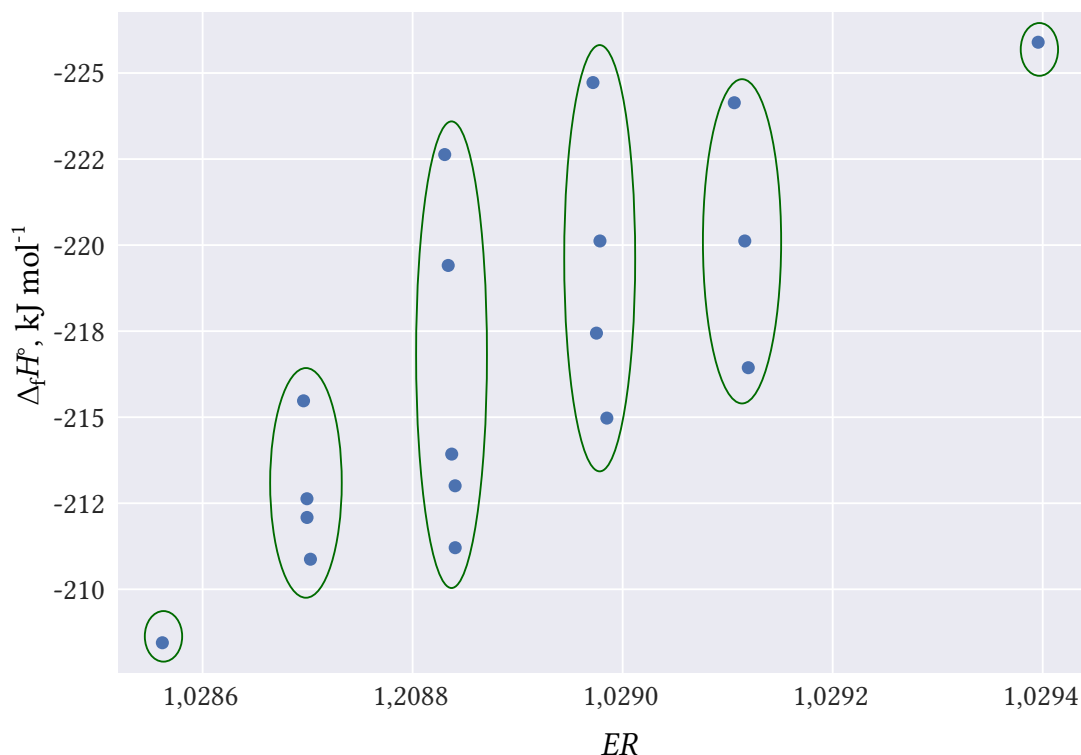
Slika 3.24: Korelacija između rezolventne energije i tački ključanja izomera oktana

Na osnovu slike 3.24 možemo zaključiti da su podaci grupisani (uokvireno zelenom bojom). Naime, vrednosti za rezolventnu energiju leže na nekoliko, skoro paralelnih linija. Direktnim proveravanjem utvrdili smo da su vrednosti za ER grupisane na osnovu prvog zagrebačkog indeksa. Odnosno, vrednosti na istoj liniji imaju istu vrednost za Zg_1 , koji grubo meri grananje u molekulu. Isto grupisanje je primećeno i kod energije grafa i kod Estradinog indeksa. Zbog toga, pretpostavili smo da je za modeliranje tački ključanja alkana uz pomoć deskriptora zasnovanih na sopstvenim vrednostima, neophodno uključiti Zg_1 kao drugi parametar.

Kao i kod tački ključanja, entalpije obrazovanja smo korelirali sa izračunatim vrednostima energije grafa, Estradinog indeksa i rezolventne energije alkana. Na slici 3.25 je, kao primer, prikazana korelacija između rezolventne energije i entalpija obrazovanja izomera oktana. Sa slike se vidi da su vrednosti rezolventne energije grupisane (uokvireno zelenom bojom). Kao u prethodnom slučaju, direktnim proveravanjem utvrdili smo da su vrednosti grupisane na osnovu prvog zagrebačkog indeksa, tj. tačke na istoj liniji imaju istu vrednost za Zg_1 . Ista situacija je uočena i kod energije grafa i kod Estradinog indeksa.

Na osnovu ovih rezultata, pretpostavili smo da je za uspešno predviđanje tački ključanja i entalpija obrazovanja alkana u odgovarajuće modele neophodno uključiti i prvi zagrebački indeks kao jedan od parametara. Takođe, na osnovu prethodno dobijenih rezultata [162] odlučili smo da i broj nula u spektru (n_0) stabala (alkana) uključimo u modele za predviđanje TK i $\Delta_f H^\circ$.

Inicijalna hipoteza je da se tačke ključanja i entalpije obrazovanja hemijskih stabala (T) mogu aproksimirati kao linearna kombinacija indeksa zasnovanog na sopstvenim vrednostima,



Slika 3.25: Korelacija između rezolventne energije i entalpija obrazovanja izomera oktana

prvog zagrebačkog indeksa i broja nula u spektru. Da bi testirali ovu pretpostavku konstruisali smo tri tipa modela (jednačine 3.11, 3.12 i 3.13) koji su zasnovani na višestrukoj linearnoj regresiji. U ove svrhe napravljen je Piton program sa implementiranim scikit-learn modulom [146]. Koeficijenti A , B , C i D su određeni metodom najmanjih kvadrata.

$$\text{TK}, \Delta_f H^\circ(T) \approx A \cdot \text{TI}(T) + B \quad (3.11)$$

$$\text{TK}, \Delta_f H^\circ(T) \approx A \cdot \text{TI}(T) + B \cdot Zg_1(T) + C \quad (3.12)$$

$$\text{TK}, \Delta_f H^\circ(T) \approx A \cdot \text{TI}(T) + B \cdot Zg_1(T) + C \cdot n_0(T) + D \quad (3.13)$$

Konstrukcija ovih modela, njihovo interno testiranje i međusobno upoređivanje obavljeno je koristeći podatke iz tabele 3.13. Očigledno je da se broj parametara u ovim modelima povećava. Sa svakim uvođenjem nove promenljive, konstruisani model je evaluiran pomoću Valdovog testa sa 95% nivoom poverenja. Ovaj test se izvodi kako bi se ispitao doprinos novog parametra u objašnjavanju varijanse. Rezultati ispitivanja valjanosti fitovanja dobijenih modela predstavljeni su u tabeli 3.14.

U ovoj tabeli R^2 , \bar{R}^2 , $RMSE$ i p -vrednost predstavljaju koeficijent determinacije, popravljene koeficijent determinacije, koren kvadrata srednje greške i p -vrednost za Valdov test.

Upoređujući rezultate Valdovog testa primećeno je da uvođenje prvog zagrebačkog indeksa u modelima zasnovanim na energiji grafa za predviđanje TK i $\Delta_f H^\circ$ ne dovodi do značajnog povećanja objašnjenja varijanse. Zbog toga, modeli zasnovani na jednačinama 3.12 i 3.13 u slučaju energije grafa su izostavljeni iz daljeg proučavanja. Suprotna situacija je kod modela zasnovanih na Estradinom indeksu i rezolventnoj energiji. Naime, kod ovih modela najbolje slaganje između eksperimentalnih i izračunatih tački ključanja se dobija kod modela zasnovanih na jednačini 3.13. Iako broj nula kod modela za predviđanje tačke ključanja predstavlja važan parametar, kod modela za predviđanje entalpije obrazovanja statistički podaci ne opravdavaju njegovo uvođenje. Najbolji modeli za predviđanje entalpije obrazovanja su oni koji su

Tabela 3.14: Rezultati ispitivanja valjanosti fitovanja modela zasnovanih na jednačinama 3.11, 3.12 i 3.13

Deskriptor	Model	R^2	\bar{R}^2	RMSE	p -vrednost
Tačka ključanja					
<i>E</i>	3.11	77,67%	76,27%	2,810	
	3.12	77,67%	74,70%	2,813	0,943
<i>EE</i>	3.11	48,23%	45,00%	4,281	
	3.12	91,36%	90,21%	1,745	$3 \cdot 10^{-7}$
	3.13	94,73%	93,61%	1,372	0,010
<i>ER</i>	3.11	49,96%	46,84%	4,205	
	3.12	91,35%	90,19%	1,745	$4 \cdot 10^{-7}$
	3.13	94,73%	93,59%	1,372	0,010
Entalpija obrazovanja					
<i>E</i>	3.11	71,51%	69,73%	2,716	
	3.12	72,87%	69,26%	2,650	0,399
<i>EE</i>	3.11	54,77%	51,95%	3,422	
	3.12	87,11%	85,39%	1,827	$2 \cdot 10^{-5}$
	3.13	87,11%	84,35%	1,827	0,933
<i>ER</i>	3.11	56,36%	53,64%	3,361	
	3.12	87,10%	85,38%	1,827	$3 \cdot 10^{-5}$
	3.13	87,11%	84,35%	1,827	0,932

zasnovani na jednačini 3.12. Zanimljivo je da modeli koji koriste Estradin indeks i modeli koji koriste rezolventnu energiju praktično pokazuju istu tačnost.

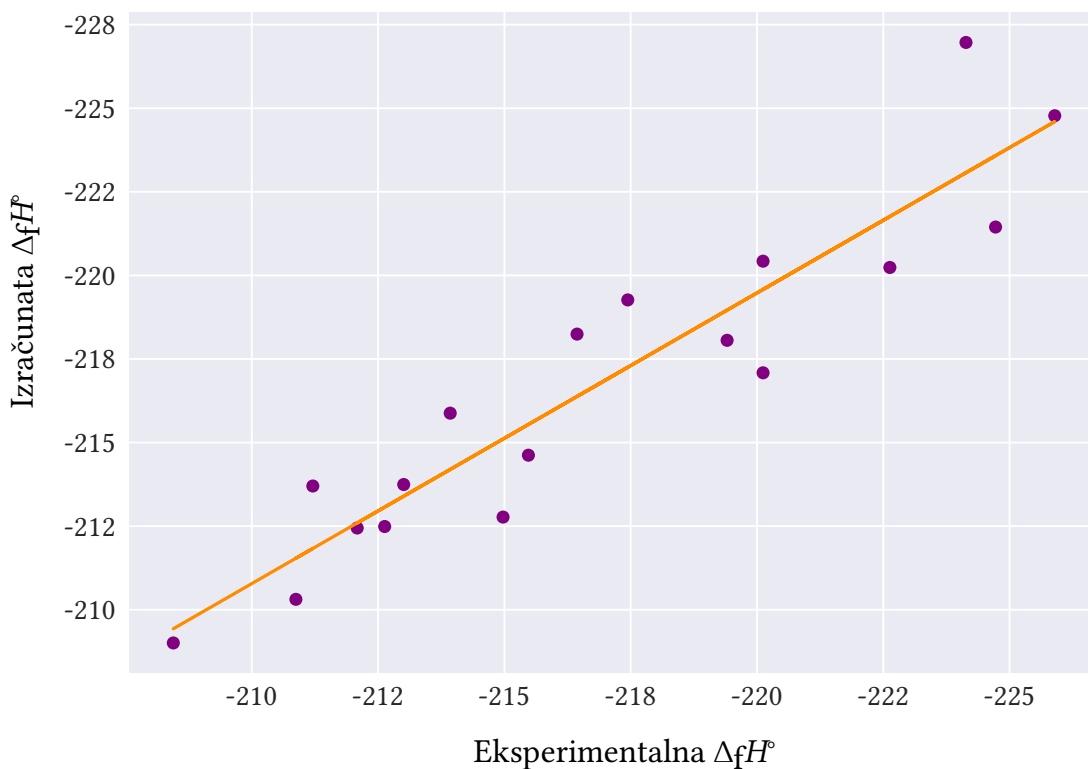
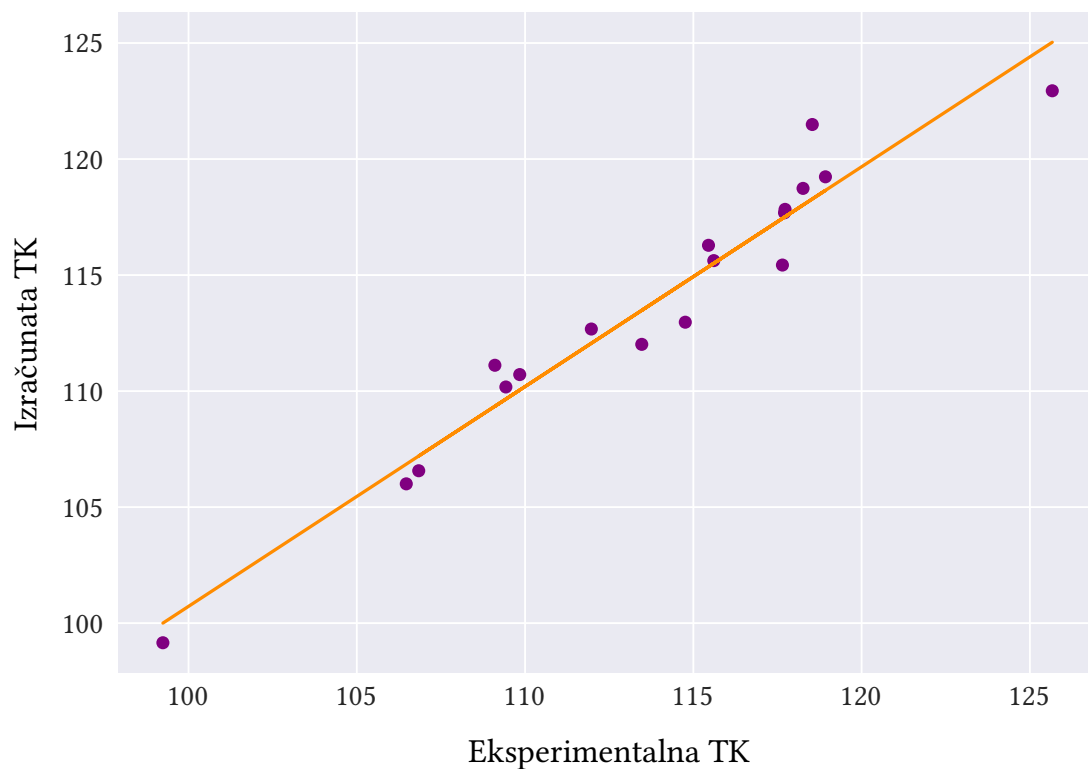
Na osnovu vrednosti koeficijenta determinacije i poboljšanog koeficijenta determinacije možemo zaključiti da najbolje slaganje pokazuju modeli zasnovani na Estradinom indeksu i rezolventnoj energiji.

U tabeli 3.15 su prikazane vrednosti koeficijenata koji se javljaju u najboljim predikcionim modelima.

Tabela 3.15: Vrednosti koeficijenata za najbolje predikcione modele

Osobina	Koeficijent	<i>E</i>	<i>EE</i>	<i>ER</i>
TK	<i>A</i>	8,40	229,65	678365,40
	<i>B</i>	41,16	-26,27	-48,56
	<i>C</i>		-1,59	-1,60
	<i>D</i>		-3041,51	-696355,18
$\Delta_f H^\circ$	<i>A</i>	6,99	221,81	655474,73
	<i>B</i>	-276,45	-25,33	-46,87
	<i>C</i>		-3266,77	-673186,57

Ilustracije radi, na slici 3.26 je prikazana korelacija između eksperimentalnih i izračunatih tački ključanja i entalpija obrazovanja pomoću modela zasnovanih na rezolventnoj energiji.



Slika 3.26: Korelacija između eksperimentalnih i izračunatih tački ključanja i entalpija obrazovanja izomera oktana

Kako bi imali bolji uvid u predikcionu moć naših modela, uporedili smo dobijene rezultate za TK i $\Delta_f H^\circ$ koristeći druge topološke molekulske deskriptore sa našim rezultatima. Naime,

modelirali smo tačke ključanja i entalpije obrazovanja oktana koristeći Vinerov [59], Balabanov [66] i Randićev indeks [73]. U jednačinama 3.11 i 3.12 umesto E , EE i ER korišćeni su ovi deskriptori. Sa vrednostima koeficijenta determinacije od 82% i 85% za modele zasnovane na jednačini 3.11 za predviđanje tačke ključanja i entalpije obrazovanja, Randićev indeks je dao najbolje rezultate. Sa uvođenjem prvog zagrebačkog indeksa, za tačke ključanja najbolje rezultate je dao Randićev indeks ($R^2 = 88\%$), dok je za entalpije obrazovanja to bio Balabanov indeks ($R^2 = 93\%$). Naši modeli zasnovani na jednačini 3.12 koristeći Estradin indeks i rezolventnu energiju su se pokazali boljim za predviđanje tačke ključanja ($R^2 \approx 92\%$), dok je za entalpiju obrazovanja Balabanov indeks pokazao malo bolju predikcionu moć.

Razvijeni modeli zasnovani na E , EE i ER su pokazali solidan predikcioni potencijal, ali da bi stvarno testirali konstruisane modele eksterna validacija je neophodna. U tu svrhu, odlučili smo se da testiramo model zasnovan na rezolventnoj energiji, jer je to nov indeks. Takođe, Estradin indeks i rezolventna energija su pokazali veoma dobar predikcioni potencijal na osnovu vrednosti koeficijenta determinacije.

Naš novi skup podataka, predstavljen u tabeli 3.16, sastoji se od 117 tački ključanja i 47 entalpija obrazovanja. Ovi podaci su dostupni u onlajn hemijskim bazama [141, 143]. U ovom skupu alkana broj ugljenikovih atoma se kreće od 2 do 10. Ovaj skup je podeljen na trening skup (80%), čiji su podaci iskorišćeni za fitovanje modela, i test skup (20%) koristeći butstrap tehniku. Podaci iz trening skupa su slučajnim izborom mešani sa podacima iz test skupa 1000 puta. Na taj način je posle svakog mešanja podataka formiran novi trening i test skup, što je rezultiralo formiranjem 1000 različitih linearnih modela.

Tabela 3.16: Eksperimentalni podaci korišćeni za butstrap ispitivanje predikcionih modela

Molekul	TK, °C	$\Delta_f H^\circ$, kJ mol ⁻¹
etan	-88,6	-84,0
propan	-42,1	-104,7
<i>n</i> -butan	-0,1	-125,6
izobutan	-11,2	-134,2
<i>n</i> -pentan	36,1	-146,8
izopentan	28,0	-153,7
<i>n</i> -heksan	68,8	-167,1
2,2-dimetilbutan	49,8	-185,6
3-metilpentan	63,3	-171,6
2-metilpentan	60,9	-174,3
2,3-dimetilbutan	58,1	-177,8
<i>n</i> -heptan	98,4	-187,8
2-metilheksan	90,1	-195,0
2,4-dimetilpentan	80,6	-202,1
2,2-dimetilpentan	79,2	-206,2
2,3-dimetilpentan	89,8	-199,2
3-metilheksan	91,9	-192,3
3,3-dimetilpentan	86,1	-201,5
2,2,3-trimetilbutan	81,0	-204,8
3-etilpentan	93,5	-189,7
2,2,4-trimetilpentan	99,2	-224,1
2,2,3,3-tetrametilbutan	106,5	-225,9
2,2,3-trimetilpentan	109,8	-220,1

nastavak na sledećoj strani

Tabela 3.16 – nastavak tabele

Molekul	TK, °C	$\Delta_f H^\circ$, kJ mol ⁻¹
2,2-dimetilheksan	106,8	-224,7
3,3-dimetilheksan	112,0	-220,1
2,3,3-trimetilpentan	114,8	-216,4
2,5-dimetilheksan	109,1	-222,6
2,4-dimetilheksan	109,4	-219,4
2-metilheptan	117,6	-215,5
4-metilheptan	117,7	-212,1
2,3-dimetilheksan	115,6	-213,9
2,3,4-trimetilpentan	113,5	-217,4
3-etil-3-metilpentan	118,3	-215,0
3,4-dimetilheksan	117,7	-213,0
3-metilheptan	118,9	-212,6
3-etilheksan	118,5	-210,9
3-etil-2-metilpentan	115,5	-211,2
<i>n</i> -oktan	125,7	-208,4
<i>n</i> -nonan	150,7	-228,3
2-metiloktan	143,0	
2,2,5-trimetilheksan	123,9	-253,1
3-metiloktan	143,9	
4-metiloktan	141,9	
2,4-dimetilheptan	132,9	
2,2-dimetilheptan	131,9	-246,1
2,2,4,4-tetrametilpentan	122,3	-241,5
3,3-dietilpentan	146,2	
2,3,5-trimetilheksan	130,9	
2,2,4-trimetilheksan	126,5	
2,3-dimetilheptan	139,9	
2,5-dimetilheptan	134,9	
2,6-dimetilheptan	134,9	
3,3-dimetilheptan	136,9	
2,2,3-trimetilheksan	134,3	
2,2,3,3-tetrametilpentan	140,3	
3-etilheptan	143,6	
2,2,3,4-tetrametilpentan	133,0	
2,4,4-trimetilheksan	130,9	
3,4-dimetilheptan	140,7	
4-etilheptan	141,2	
2,3,4-trimetilheksan	138,9	
3,5-dimetilheptan	136,1	
4,4-dimetilheptan	135,3	
2,3,3-trimetilheksan	137,8	
2,3,3,4-tetrametilpentan	139,9	
3,3,4-trimetilheksan	140,4	
3-etil-2,4-dimetilpentan	136,7	-227,4
3-etil-2,2-dimetilpentan	133,8	

nastavak na sledećoj strani

Tabela 3.16 – nastavak tabele

Molekul	TK, °C	$\Delta_f H^\circ$, kJ mol ⁻¹
4-etil-2-metilheksan	133,8	
3-etil-3-metilheksan	140,6	
3-etil-2-metilheksan	138,0	
3-etil-4-metilheksan	140,4	
3-etil-2,3-dimetilpentan	141,6	
<i>n</i> -dekan	174,1	-249,7
3-metilnonan	167,8	
2-metilnonan	166,9	-260,2
4-metilnonan	166,9	
5-metilnonan	165,2	-258,6
2,6-dimetiloktan	158,9	
2,3-dimetiloktan	164,3	
2,7-dimetiloktan	160,0	
3,3-dimetiloktan	161,2	
4-etiloktan	163,4	
2,4,6-trimetilheptan	145,9	
4,4-dimetiloktan	157,6	
2,2,4-trimetilheptan	146,0	
3,5-dimetiloktan	158,9	
2,4-dimetiloktan	155,9	
3,3,5-trimetilheptan	155,7	
2,5-dimetiloktan	156,9	
2,2,5,5-tetrametilheksan	137,6	-285,0
3,4-dietilheksan	157,6	
3,3,4,4-tetrametilheksan	170,1	
4-propilheptan	156,9	
4,5-dimetiloktan	163,0	
3-etil-3-metilheptan	163,9	
2,2,3,3-tetrametilheksan	158,1	
2,2,6-trimetilheptan	148,0	
2,3,6-trimetilheptan	137,5	
2,2,3,3,4-pentametilpentan	166,1	
2,5,5-trimetilheptan	152,8	
2,4,4-trimetilheptan	149,6	
4-etil-2,2-dimetilheksan	147,1	
2,3,4-trimetilheksan	155,9	
2,2,3,5-tetrametilheksan	148,9	
2,2,4,5-tetrametilheksan	147,9	
2,3,4,4-tetrametilheksan	161,6	
2,2,3,4,4-pentametilpentan	159,3	
2,2,3,4-tetrametilheksan	158,9	
2,3,4,5-tetrametilheksan	156,0	
2,3,3,4-tetrametilheksan	164,6	
2,3,3,5-tetrametilheksan	152,9	
3-etil-2,2,4-trimetilpentan	155,3	

nastavak na sledećoj strani

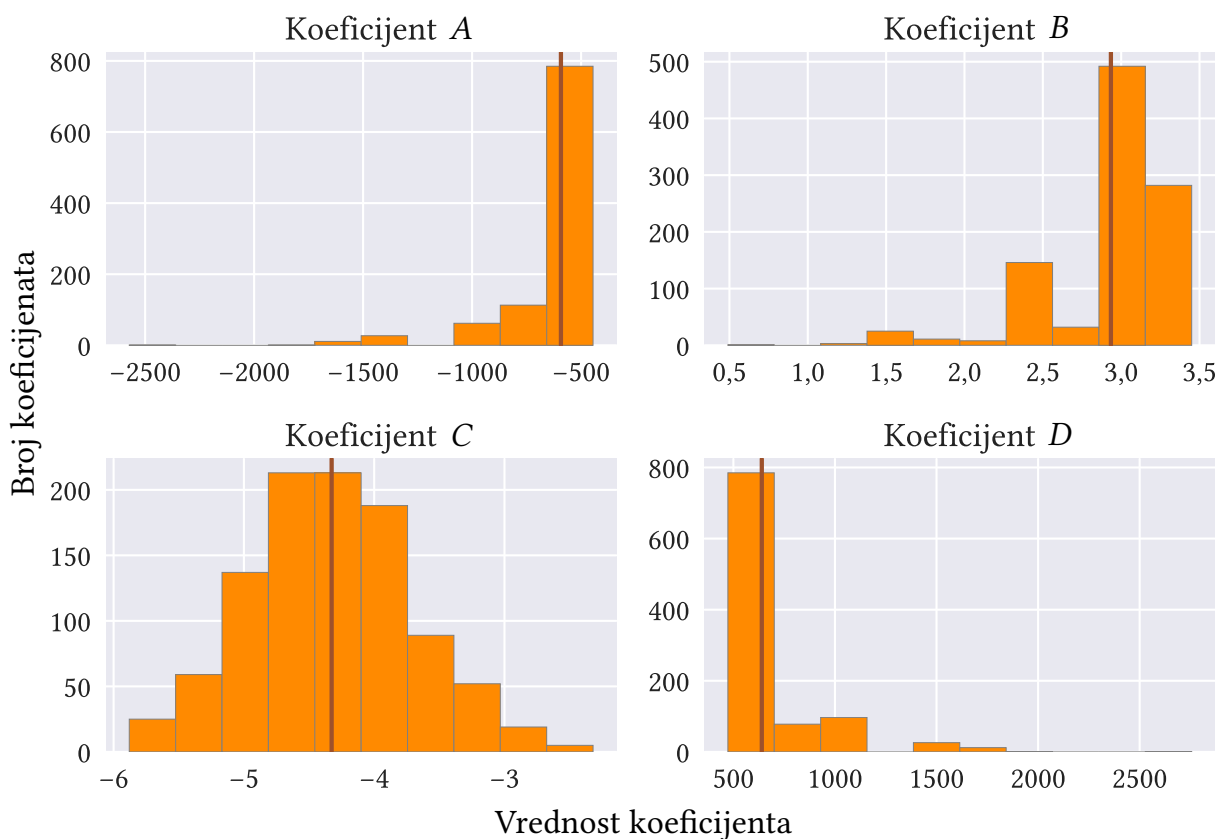
Tabela 3.16 – nastavak tabele

Molekul	TK, °C	$\Delta_f H^\circ$, kJ mol ⁻¹
3-izopropil-2,4-dimetilpentan	157,0	
3-etil-2,3,4-trimetilpentan	169,4	
3,3-dietilheksan	162,7	
2,3,3-trimetilheptan	160,1	

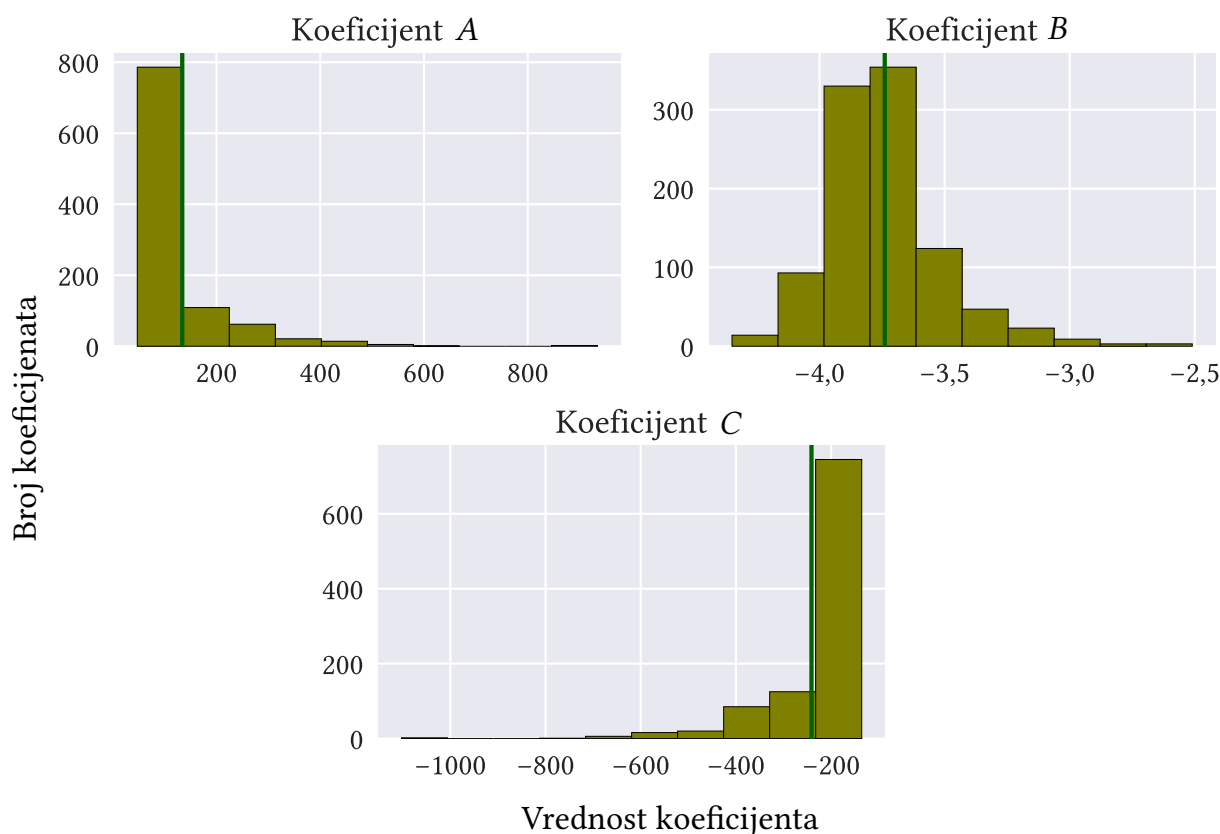
U tabeli 3.17 su date prosečne vrednosti koeficijenata fitovanja kao i prosečne vrednosti koeficijenata determinacije za 1000 predikcionih modela. Na slikama 3.27 i 3.28 su prikazane distribucije vrednosti koeficijenata koji se javljaju u modelima.

Tabela 3.17: Prosečne vrednosti koeficijenata fitovanja i koeficijenata determinacije predikcionih modela

Osobina	Koeficijent	R^2 , %
TK	A	-592 ± 20
	B	$2,93 \pm 0,02$
	C	$-4,33 \pm 0,04$
	D	640 ± 20
$\Delta_f H^\circ$	A	131 ± 5
	B	$-3,75 \pm 0,01$
	C	-238 ± 6
		Trening skup 88,7 Test skup 75,86
		Trening skup 95,18 Test skup 89,09



Slika 3.27: Distribucija vrednosti koeficijenata A, B, C i D iz modela 3.13 za tačku ključanja. Vertikalna linija označava srednju vrednost



Slika 3.28: Distribucija vrednosti koeficijenata A , B i C iz modela 3.12 za entalpiju obrazovanja. Vertikalna linija označava srednju vrednost

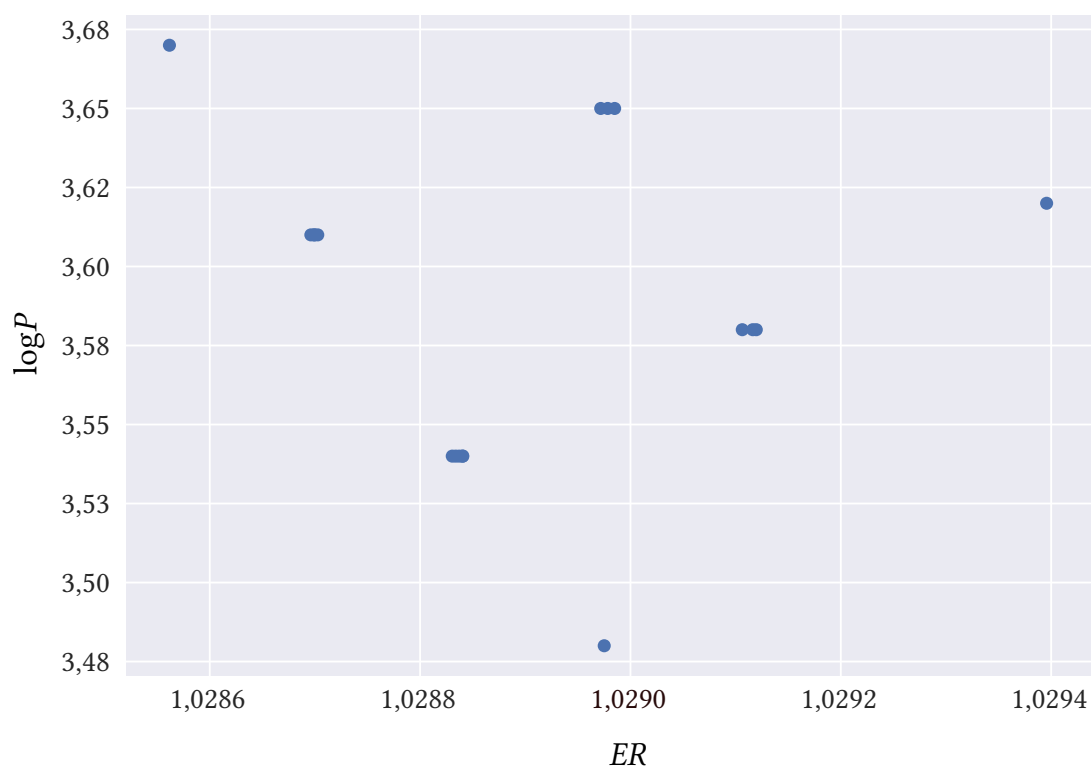
Procenat objašnjene varijanse od 75,86% i 89,09% u test skupovima pokazuju da su izračunate tačke ključanja i entalpije obrazovanja u dobrom slaganju sa eksperimentalnim vrednostima. Posebno dobro slaganje se primećuje kod modela za predviđanje entalpije obrazovanja. Ovi rezultati ukazuju da se molekularni deskriptori zasnovani na sopstvenim vrednostima (posebno rezolventna energija) mogu upotrebiti za modeliranje međumolekulskih interakcija i termodinamičkih osobina jedinjenja.

3.5.2 Koeficijent $\log P$

Predikciona moć molekularnih deskriptora zasnovanih na sopstvenim vrednostima testirana je i u slučaju predviđanja koeficijenta $\log P$. Kao i kod tačke ključanja i entalpije obrazovanja prvi korak u konstruisanju predikcionog modela jeste ispitivanje korelacije izračunatih vrednosti deskriptora sa eksperimentalnim vrednostima $\log P$. Na slici 3.29 je prikazana korelacija između rezolventne energije i eksperimentalnih vrednosti $\log P$.

Sa slike se jasno vidi da su tačke grupisane. Direktnim proveravanjem utvrdili smo da se ovo grupisanje vrši na osnovu broja čvorova stepena 1 (n_{d1}) tj. na osnovu broja metil grupa. Na osnovu ovoga pretpostavili smo da se koeficijent $\log P$ hemijskih stabala može aproksimirati kao linearna kombinacija molekularnog deskriptora zasnovanog na sopstvenim vrednostima i n_{d1} . Jednačina 3.14 predstavlja model za predviđanje koeficijenta $\log P$ koji je zasnovan na višestrukoj linearnoj regresiji:

$$\log P(T) \approx A \cdot TI(T) + B \cdot n_{d1}(T) + C \quad (3.14)$$



Slika 3.29: Korelacija između rezolventne energije i koeficijenta $\log P$ izomera oktana

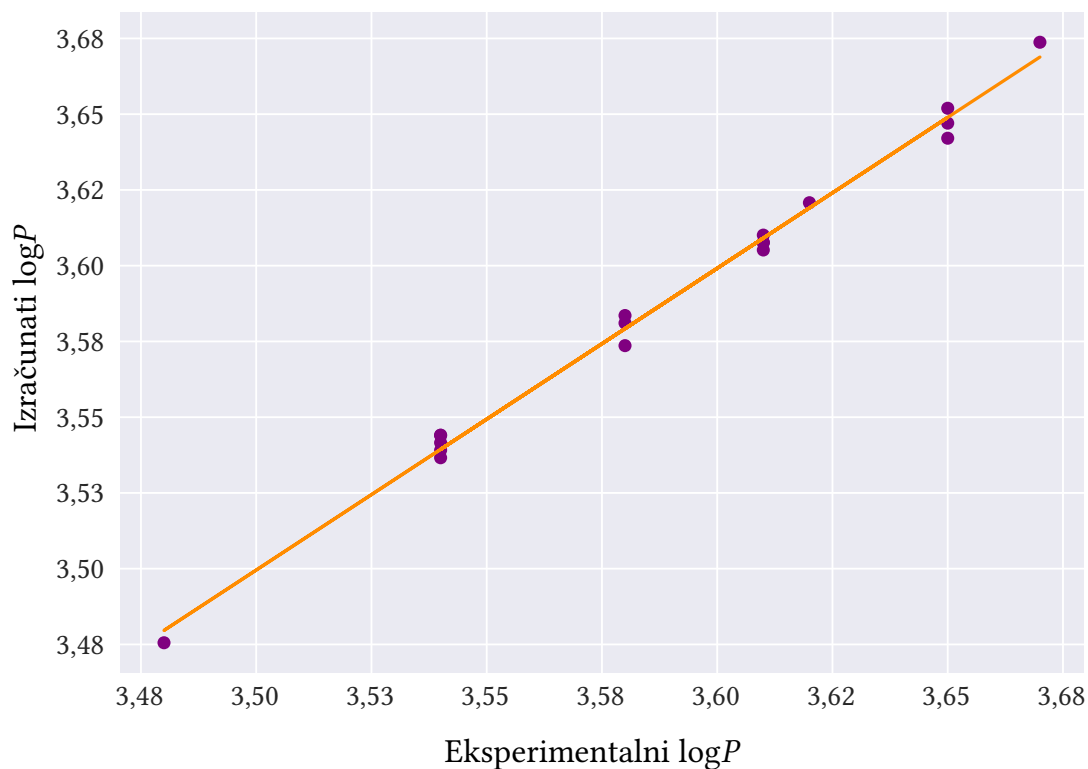
gde A , B i C predstavljaju koeficijente fitovanja koji su određeni metodom najmanjih kvadrata. U tabeli 3.18 su predstavljene njihove vrednosti kao i vrednosti koeficijenata determinacije za dobijene modele.

Tabela 3.18: Vrednosti koeficijenata fitovanja i koeficijenata determinacije predikcionih modela

Osobina	Koeficijent	E	EE	ER
$\log P$	A	-0,033	0,467	747,041
	B	-0,038	-0,165	-0,169
	C	4,030	-3,826	-764,366
R^2 , %		20,30	98,62	99,48

Koeficijenti determinacije pokazuju da postoji dobro slaganje između eksperimentalnih i izračunatih vrednosti koeficijenta $\log P$ izomera oktana. Na slici 3.30 je prikazana korelacija između eksperimentalnih vrednosti i vrednosti izračunatih pomoću modela zasnovanog na rezolventnoj energiji.

Izuzetno dobra korelacija između eksperimentalnih i izračunatih vrednosti koeficijenta $\log P$ (kod modela zasnovanih na Estradinom indeksu i rezolventnoj energiji) može se delimično objasniti činjenicom da ovi deskriptori nose informacije o voluminoznosti molekula. Kao što je ranije prikazano, Estradin indeks i rezolventna energija se mogu izraziti preko spektralnih momenata. Veza između spektralnih momenata i zagrebačkih indeksa (prvog i drugog) i F indeksa [182] je poznata. Ovi indeksi, koji su zasnovani na stepenu čvora, nose informacije o grananju u molekulu, odnosno, samim tim informacije o voluminoznosti molekula. S druge strane, veza između koeficijenta $\log P$ nepolarnih jedinjenja i njegove voluminoznosti je odavno



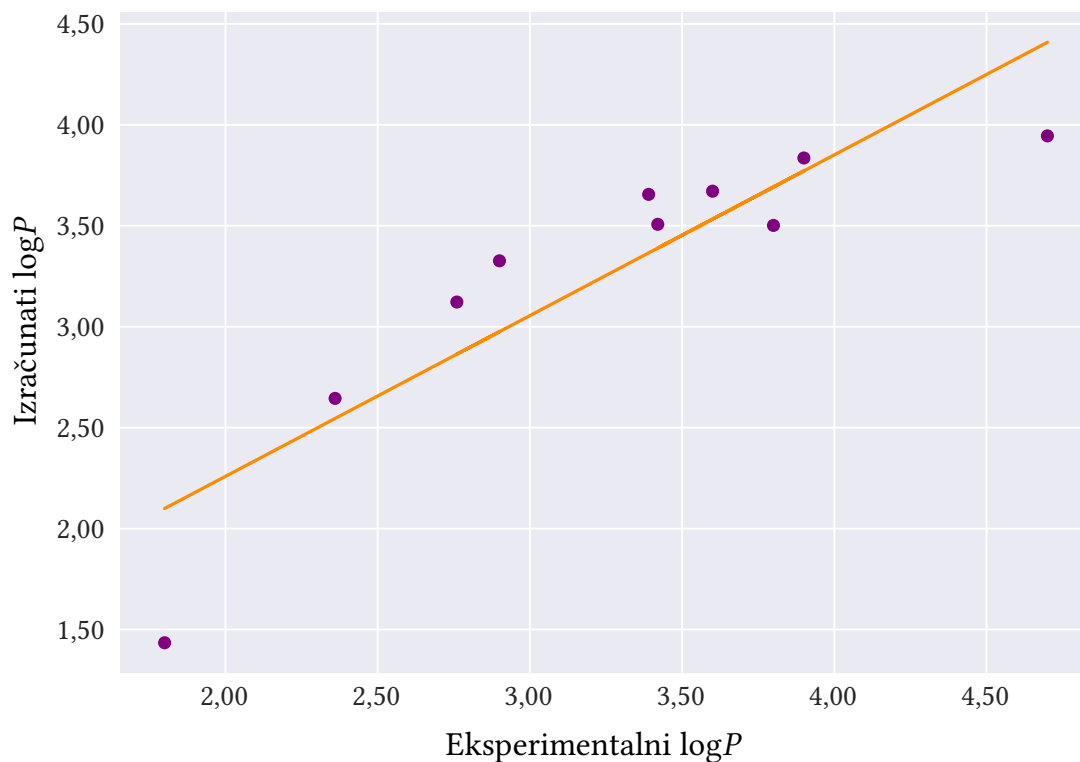
Slika 3.30: Korelacija između eksperimentalnih vrednosti iz tabele 3.13 i izračunatih vrednosti koeficijenta $\log P$

poznata [183]. Zajednička zavisnost rezolventne energije (Estradinog indeksa) i koeficijenta $\log P$ od voluminoznosti molekula donekle objašnjava fizički smisao dobre predikcione moći razvijenih modela.

Tabela 3.19: Eksperimentalne vrednosti koeficijenta $\log P$

Molekul	$\log P$
etan	1,81
propan	2,36
<i>n</i> -butan	2,89
2-metilpropan	2,76
<i>n</i> -pentan	3,39
<i>n</i> -heksan	3,90
2,2-dimetilbutan	3,82
3-metilpentan	3,60
2,3-dimetilbutan	3,42
<i>n</i> -heptan	4,66

Kako bi dodatno ispitali dobijene modele iskorišćeni su podaci iz tabele 3.19. Usled nedostatka dovoljnog broja eksperimentalnih podataka, uspele smo da sakupimo samo deset eksperimentalnih vrednosti koeficijenta $\log P$.



Slika 3.31: Korelacija između eksperimentalnih vrednosti iz tabele 3.19 i izračunatih vrednosti koeficijenta $\log P$

Na slici 3.31 je prikazana korelacija između eksperimentalnih vrednosti datih u tabeli 3.19 i izračunatih vrednosti pomoću modela 3.14. Koeficijent determinacije od 79,57% ukazuje na dobro slaganje između eksperimentalnih i izračunatih vrednosti. Međutim, kako bi se ispitala valjanost i stabilnost konstruisanog modela neophodno je da model bude testiran na skupu koji se sastoji od većeg broja podataka. Iako je naš skup mali, ipak pokazuje da molekularni deskriptori zasnovani na sopstvenim vrednostima (Estradin indeks i rezolventna energija) mogu biti upotrebljeni za modeliranje ove važne osobine jedinjenja.

4 ZAKLJUČCI

U okviru ove doktorske disertacije ispitivani su topološki molekulska deskriptori zasnovani na sopstvenim vrednostima matrice susedstva. Tu spadaju energija grafa, Estradin indeks i rezolventna energija. Cilj ovih testiranja je bio da se pokaže sličnost, odnosno različitost ponašanja ovih deskriptora u okviru različitih problema.

U poglavlju 3.1 su predstavljeni rezultati ispitivanja relacija između ova tri molekulska deskriptora kod hemijskih stabala. Utvrđeno je da prvi zagrebački indeks, kao mera grananja u molekulu, i broj nula u spektru utiču na veze između ovih indeksa. Na osnovu ove činjenice, konstruisane su aproksimativne jednačine. Modeli zasnovani na jednačinama 3.1 i 3.5 pokazuju da se veze između $ER(T)$ i $E(T)$, $ER(T)$ i $EE(T)$, respektivno, mogu uspešno modelirati pomoću ovih parametara grafa. Valjanost modela 3.5 za modeliranje veze između Estrading indeksa i rezolventne energije je potvrđena i kod izomernih katakondenzovanih i perikondenzovanih benzenoidnih ugljovodonika. Jednačina koja sadrži Zg_1 i n_0 (model 3.8) je upotrebljena za modeliranje veze između energije grafa i Estradinog indeksa. Iako je pomoću ovog modela objašnjeno oko 98% varijanse u podacima, utvrđeno je da pored ova dva parametra, još neki parametar utiče na vezu između E i EE .

Diskriminativna moć energije grafa i Estradinog indeksa, u slučaju hemijskih stabala, je slična i ne prelazi 21%. Međutim, senzitivnost rezolventne energije drastično opada sa porastom broja čvorova. Ovo otkriće ukazuje da postoji mnoštvo hemijskih stabala sa istom rezolventnom energijom (u ovoj disertaciji takva stabla su nazvana r -ekvienergetska). Inicijalnim pretraživanjem stabala pokazalo se da postoji mnoštvo r -ekvienergetskih stabala sa većim brojem čvorova, iako je prag podešen na 10^{-20} . Dodatna ispitivanja su pokazala da se povećanjem preciznosti računanja ER smanjuje broj neizospektralnih r -ekvienergetskih hemijskih stabala, odnosno, da su su dobijena r -ekvienergetska stabla u stvari skoro- r -ekvienergetska. To zapravo znači da ER pokazuje izrazitu senzitivnost ali da računarski resursi i visoka preciznost proračuna sprečavaju praktičnu primenu ER za razlikovanje izomera.

Strukturna osetljivost energije grafa, Estradinog indeksa i rezolventne energije je testirana na katakondenzovanim i perikondenzovanim benzenoidnim ugljovodonicima. U ovu svrhu je primenjen dostupni algoritam za određivanje strukturne senzitivnosti. Sva tri deskriptora su pokazala sličan nivo osetljivosti kod katakondenzovanih molekula. ER je sa najnižim vrednostima za $|\overline{SS} - \overline{Abr}|$ u blagoj prednosti u odnosu na E i EE . Što se tiče perikondenzovanih benzenoidnih ugljovodonika, energija grafa je pokazala superiornost u odnosu na druga dva topološka molekulska deskriptora. Ispitivanja su potom nastavljena na katakondenzovanim benzenoidnim ugljovodonicima sa različitim brojem zaliva, uvala i fjordova. Ovo ispitivanje je pokazalo da sva tri deskriptora pokazuju sličnu strukturnu senzitivnost kod ovih molekula.

Takođe, utvrđeno je da se njihove vrednosti postepeno menjaju sa povećanjem broja ovih strukturnih detalja.

Ispitivanjem uticaja veličine cikla, veličine molekula i tipa anelacije pokazano je da se ef_{EE} i ef_{ER} ponašaju znatno drugačije u odnosu na ef_E . Na osnovu ovoga se može zaključiti da Estradin indeks i rezolventna energija prikupljaju drugačije strukturne informacije u odnosu na energiju grafa. Iako se Estradin indeks i rezolventna energija u većini slučajeva slično ponašaju, istraživanjem uticaja cikla pokazalo se da ovi indeksi sadrže različite informacije koje mogu biti od koristi.

Predikcioni potencijal energije grafa, Estradinog indeksa i rezolventne energije je testiran na predviđanje tačke ključanja, entalpije obrazovanja i $\log P$ alkana. Utvrđeno je da za modeliranje tačke ključanja, pored deskriptora zasnovanog na sopstvenim vrednostima, neophodno uključiti prvi zagrebački indeks i broj nula u spektru. Kod entalpija obrazovanja, pored E , EE i ER , potrebno je uključiti prvi zagrebački indeks. Kod partitionog koeficijenta oktanol/voda uz E , EE i ER potrebno je koristiti broj metil grupa za uspešno predviđanje ove osobine. Kada su ovi parametri identifikovani konstruisani su modeli zasnovani na višestrukoj linearnoj regresiji. Dobijeni modeli su testirani i pokazali su dobro slaganje sa eksperimentalnim vrednostima. Posebno dobru predikcionu moć su pokazali Estradin indeks i rezolventna energija.

Literatura

- [1] R. Todeschini, V. Consonni: *Handbook of molecular descriptors*. Wiley–VCH, 2000.
- [2] M. Karelson, V. S. Lobanov, A. R. Katritzky: *Quantum–chemical descriptors in QSAR/QSPR studies*. Chem. Rev., **96**:1027–1044, 1996.
- [3] H. Hong, Q. Xie, W. Ge *et al.*: *Mold², molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics*. J. Chem. Inf. Model., **48**:1337–1344, 2008.
- [4] P. Gramatica, N. Navas, R. Todeschini: *Classification of organic solvents and modelling of their physico–chemical properties by chemometric methods using different sets of molecular descriptors*. Trends Anal. Chem., **18**:461–471, 1999.
- [5] A. Tropsha, Z. Weifan: *Identification of the descriptor pharmacophores using variable selection QSAR applications to database mining*. Curr. Pharm. Des., **7**:599–612, 2001.
- [6] D. Casoni, J. Petre, V. David *et al.*: *Prediction of pesticides chromatographic lipophilicity from the computational molecular descriptors*. J. Sep. Sci., **34**:247–254, 2011.
- [7] A. R. Leach, V. J. Gillet: *An introduction to chemoinformatics*. Springer Netherlands, 2007.
- [8] J. Gasteiger, T. Engel: *Chemoinformatics: a textbook*. Wiley–VCH, 2003.
- [9] J. L. Faulon, A. Bender: *Handbook of chemoinformatics algorithms*. CRC Press, 2010.
- [10] M. Dehmer, K. Varmuza, D. Bonchev: *Statistical modelling of molecular descriptors in QSAR/QSPR*. Wiley–VCH, 2012.
- [11] T. Engel, J. Gasteiger: *Chemoinformatics: basic concepts and methods*. Wiley–VCH, 2018.
- [12] T. Engel, J. Gasteiger: *Applied chemoinformatics: achievements and future opportunities*. Wiley–VCH, 2018.
- [13] D. Butina: *Unsupervised data base clustering based on daylight’s fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets*. J. Chem. Inform. Comput. Sci., **39**:747–750, 1999.
- [14] A. Cereto–Massague, M. J. Ojeda, C. Valls *et al.*: *Molecular fingerprint similarity search in virtual screening*. Methods, **71**:58–63, 2015.
- [15] P. Willett: *Similarity searching using 2D structural fingerprints*. U: J. Bajorath (urednik): *Chemoinformatics and computational chemical biology. Methods in molecular biology (Methods and Protocols)*, strane 133–158. Humana Press, 2010.

- [16] M. Karelson: *Molecular descriptors in QSAR/QSPR*. Wiley, 2000.
- [17] K. Roy, S. Kar, R. N. Das: *A primer on QSAR/QSPR modeling: fundamental concepts*. Springer International Publishing, 2015.
- [18] A. Z. Dudek, T. Arodz, J. Galvez: *Computational methods in developing quantitative structure–activity relationships (QSAR): a review*. *Comb. Chem. High Throughput Screen.*, **9**:213–228, 2006.
- [19] R. Todeschini, V. Consonni: *Molecular descriptors for chemoinformatics*. Wiley–VCH, 2009.
- [20] V. Consonni, R. Todeschini: *Molecular descriptors*. U: T. Puzyn, J. Leszczynski, i M. Cronin (urednici): *Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics*, strane 29–102. Springer, 2010.
- [21] S. C. Basak, V. R. Magnuson, G. J. Niemi *et al.*: *Topological indices: their nature, mutual relatedness, and applications*. *Math. Model.*, **8**:300–305, 1987.
- [22] I. Redžepović, B. Furtula: *Topološki molekularni deskriptori*. *Hem. Preg.*, **61**:131–136, 2020.
- [23] J. Devillers, A. T. Balaban: *Topological indices and related descriptors in QSAR and QSPR*. CRC Press, 1999.
- [24] A. Talevi, C. L. Bellera, M. D. Ianni *et al.*: *An integrated drug development approach applying topological descriptors*. *Curr. Comput. Aided Drug Des.*, **8**:172–181, 2012.
- [25] R. Zanni, M. Galvez–Llompant, R. Garcia–Domenech *et al.*: *Latest advances in molecular topology applications for drug discovery*. *Expert. Opin. Drug Discov.*, **10**:945–957, 2015.
- [26] J. C. Dearden: *The use of topological indices in QSAR and QSPR modeling*. U: K. Roy (urednik): *Advances in QSAR modeling. Challenges and advances in computational chemistry and physics.*, strane 57–88. Springer, 2017.
- [27] M. Murcia–Soler, F. Perez–Gimenez, F. J. Garcia–March *et al.*: *Discrimination and selection of new potential antibacterial compounds using simple topological descriptors*. *J. Mol. Graph. Model.*, **21**:375–390, 2003.
- [28] J. R. Votano, M. Parham, L. H. Hall *et al.*: *New predictors for several ADME/Tox properties: aqueous solubility, human oral absorption, and Ames genotoxicity using topological descriptors*. *Mol. Divers.*, **8**:379–391, 2004.
- [29] R. Gozalbes, J. P. Doucet, F. Derouin: *Application of topological descriptors in QSAR and drug design: history and new trends*. *Curr. Drug Targets Infect. Disord.*, **2**:93–102, 2002.
- [30] E. Estrada, G. Patlewicz, E. Uriarte: *From molecular graphs to drugs. A review on the use of topological indices in drug design and discovery*. *Indian J. Chem. A*, **42**:1315–1329, 2003.
- [31] D. B. West: *Introduction to graph theory*. Prentice Hall, 1996.
- [32] B. Bollobas: *Modern graph theory*. Springer, 1998.

- [33] A. T. Balaban: *Applications of graph theory in chemistry*. J. Chem. Inf. Comput. Sci., **25**:334–343, 1985.
- [34] H. P. Schultz: *Topological organic chemistry. 1. Graph theory and topological indices of alkanes*. J. Chem. Inf. Comput. Sci., **29**:227–228, 1989.
- [35] I. Gutman: *Uvod u hemijsku teoriju grafova*. Univerzitet u Kragujevcu, Prirodno–matematički fakultet, 2003.
- [36] I. Gutman: *Selected theorems in chemical graph theory*. University of Kragujevac, Faculty of Science, 2017.
- [37] N. Trinajstić: *Chemical graph theory*. CRC Press, 2018.
- [38] D. Janežić, A. Miličević, S. Nikolić *et al.*: *Graph–theoretical matrices in chemistry*. CRC Press, 2015.
- [39] H. Sachs: *Beziehungen zwischen den in einem Graphen enthaltenen Kreisen und seinem charakteristischen Polynom*. Publ. Math. Debrecen, **11**:119–134, 1964.
- [40] I. Gutman: *Impact of the Sachs theorem on theoretical chemistry: a participant’s testimony*. MATCH Commun. Math. Comput. Chem., **48**:17–34, 2003.
- [41] J. R. Dias: *Molecular orbital calculations using chemical graph theory*. Springer, 1993.
- [42] D. M. Cvetković, M. Doob, H. Sachs: *Spectra of graphs: theory and application*. Academic Press, 1980.
- [43] J. K. Burdett, S. Lee: *Moments and the energies of solids*. J. Am. Chem. Soc., **107**:3050–3063, 1985.
- [44] J. K. Burdett, S. Lee: *Moments method and elemental structures*. J. Am. Chem. Soc., **107**:3063–3082, 1985.
- [45] S. Lee: *Second–moment scaling and covalent crystal structures*. Acc. Chem. Res., **24**:249–254, 1991.
- [46] G. G. Hall: *The evaluation of moments for polycyclic hydrocarbons*. Theor. Chim. Acta, **70**:323–332, 1986.
- [47] S. Marković, I. Gutman: *Dependence of spectral moments of benzenoid hydrocarbons on molecular structure*. J. Mol. Struct. THEOCHEM, **235**:81–87, 1991.
- [48] S. Marković: *The evaluation of the eighth moment for benzenoid graphs*. Theor. Chim. Acta, **81**:237–244, 1992.
- [49] I. Gutman, V. R. Rosenfeld: *Spectral moments of polymer graphs*. Theor. Chim. Acta, **93**:191–197, 1996.
- [50] S. Marković, A. Stajković: *The evaluation of spectral moments for molecular graphs of phenylenes*. Theor. Chem. Acc., **96**:256–260, 1997.
- [51] S. Marković, Z. Marković, J. P. Engelbrecht *et al.*: *Spectral moments of polycyclic aromatic hydrocarbons. Solution of a kinetic problem*. J. Chem. Inf. Comput. Sci., **42**:82–86, 2002.

- [52] I. Redžepović, S. Marković, B. Furtula: *On structural dependence of enthalpy of formation of catacondensed benzenoid hydrocarbons*. MATCH Commun. Math. Comput. Chem., **82**:663–678, 2019.
- [53] U. Debnath, S. B. Katti, Y. S. Prabhakar: *Graph theory concepts in the rationales of anti HIV-1 compounds*. Curr. Comput. Aided Drug Des., **9**:472–481, 2013.
- [54] M. Mahani, S. Sheikghomi, H. Sheikghomi *et al.*: *Quantitative structure–activity relationship study on the binding affinity of some aminothiazole derivatives with a dopamine receptor in brain*. J. Struct. Chem., **58**:344–348, 2017.
- [55] Y. Tsuji, E. Estrada, R. Movassagh *et al.*: *Quantum interference, graphs, walks, and polynomials*. Chem. Rev., **118**:4887–4911, 2018.
- [56] I. Gutman, K. C. Das: *The first Zagreb index 30 years after*. MATCH Commun. Math. Comput. Chem., **50**:83–92, 2004.
- [57] M. Randić: *Generalized molecular descriptors*. J. Math. Chem., **7**:155–168, 1991.
- [58] A. T. Balaban: *Topological indices and their uses: a new approach for the coding of alkanes*. J. Mol. Struct. THEOCHEM, **165**:243–253, 1988.
- [59] H. Wiener: *Structural determination of paraffin boiling points*. J. Am. Chem. Soc., **69**:17–20, 1947.
- [60] H. Wiener: *Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons*. J. Am. Chem. Soc., **69**:2636–2638, 1947.
- [61] H. Wiener: *Vapor pressure–temperature relationships among the branched paraffin hydrocarbons*. J. Phys. Chem., **52**:425–430, 1948.
- [62] S. Nikolić, N. Trinajstić: *The Wiener index: development and applications*. Croat. Chem. Acta, **68**:105–129, 1995.
- [63] I. Gutman, J. H. Potgieter: *Wiener index and intermolecular forces*. J. Serb. Chem. Soc., **62**:185–192, 1997.
- [64] G. Chartrand, O. R. Oellermann, S. L. Tian *et al.*: *Steiner distance in graphs*. Časopis pro pěstování matematiky, **114**:399–410, 1989.
- [65] X. Li, Y. Mao, I. Gutman: *The Steiner Wiener index of a graph*. Discuss. Math. Graph Theory, **36**:455–465, 2016.
- [66] A. T. Balaban: *Highly discriminating distance based numerical descriptor*. Chem. Phys. Lett., **89**:399–404, 1982.
- [67] D. Plavšić, S. Nikolić, N. Trinajstić *et al.*: *On the Harary index for the characterization of chemical graphs*. J. Math. Chem., **12**:235–250, 1993.
- [68] I. Gutman: *Degree–based topological indices*. Croat. Chem. Acta, **86**:351–361, 2013.
- [69] I. Gutman, J. Tošović: *Testing the quality of molecular structure descriptors. Vertex–degree–based topological indices*. J. Serb. Chem. Soc., **78**:805–810, 2013.

- [70] I. Gutman, N. Trinajstić: *Graph theory and molecular orbitals. Total π -electron energy of alternant hydrocarbons*. Chem. Phys. Lett., **17**:535–538, 1972.
- [71] I. Gutman, B. Rušćić, N. Trinajstić *et al.*: *Graph theory and molecular orbitals. XII. Acyclic polyenes*. J. Chem. Phys., **62**:3399–3405, 1975.
- [72] S. Nikolić, G. Kovačević, A. Miličević *et al.*: *The Zagreb indices 30 years after*. Croat. Chem. Acta, **76**:113–124, 2003.
- [73] M. Randić: *Characterization of molecular branching*. J. Am. Chem. Soc., **97**:6609–6615, 1975.
- [74] M. Randić: *The connectivity index 25 years after*. J. Mol. Graph. Model., **20**:19–35, 2001.
- [75] M. Randić: *On history of the Randić index and emerging hostility toward chemical graph theory*. MATCH Commun. Math. Comput. Chem., **59**:5–124, 2008.
- [76] E. Hückel: *Quantentheoretische beiträge zum benzolproblem*. Zeit. Phys., **70**:204–286, 1931.
- [77] C. A. Coulson, B. O’Leary, R. B. Mallion: *Hückel theory for organic chemists*. Academic Press, 1978.
- [78] E. Heilbronner, H. Bock: *The HMO model and its application: basis and manipulation*. Wiley, 1976.
- [79] A. Streitwieser: *Molecular orbital theory for organic chemists*. U: E. T. Strom i A. K. Wilson (urednici): *Pioneers of quantum chemistry*, strane 275–300. ACS Publications, 2013.
- [80] Hs. H. Gunthard, H. Primas: *Zusammenhang von Graphentheorie und MO-Theorie von Molekeln mit Systemen konjugierter Bindungen*. Helv. Chim. Acta, **39**:1645–1653, 1956.
- [81] I. Gutman: *The energy of a graph*. Ber. Math-Stat. Sect. Forschungsz. Graz, **103**:1–22, 1978.
- [82] X. Li, Y. Shi, I. Gutman: *Graph energy*. Springer, 2012.
- [83] E. Estrada, M. Benzi: *What is the meaning of the graph energy after all?* Discret. Appl. Math., **230**:71–77, 2017.
- [84] M. V. Diudea, A. Pirvan-Moldovan, R. Pop *et al.*: *Energy of graphs and remote graphs in hypercubes, rhombellanes and fullerenes*. MATCH Commun. Math. Comput. Chem., **80**:835–852, 2018.
- [85] X. Ma: *A low bound on graph energy in terms of minimum degree*. MATCH Commun. Math. Comput. Chem., **81**:393–404, 2019.
- [86] I. Gutman, D. Vidović, N. Cmiljanović *et al.*: *Graph energy—a useful molecular structure-descriptor*. Indian J. Chem. A, **42**:1309–1311, 2003.
- [87] H. S. Ramane: *Energy of graphs*. U: M. Pal, S. Samanta, i A. Pal (urednici): *Handbook of research on advanced applications of graph theory in modern society*, strane 267–296. IGI Global, 2020.

- [88] S. Akbari, A. H. Ghodrati, M. A. Hosseinzadeh: *Some lower bounds for the energy of graphs*. Linear Algebra Its Appl., **591**:205–214, 2020.
- [89] I. Gutman, B. Furtula: *Energies of graphs—survey, census, bibliography*. Center for Scientific Research SASA and University of Kragujevac, 2019.
- [90] I. Gutman, H. S. Ramane: *Research on graph energies in 2019*. MATCH Commun. Math. Comput. Chem., **84**:277–292, 2020.
- [91] I. Gutman, B. Furtula: *The total π -electron energy saga*. Croat. Chem. Acta, **90**:359–368, 2017.
- [92] M. Gao, E-P. Lim, D. Lo: *Network data mining and analysis*. World Scientific Publishing Co., Inc., 2018.
- [93] J. Pražnikar, M. Tomić, D. Turk: *Validation and quality assessment of macromolecular structures using complex network analysis*. Sci. Rep., **9**:1–11, 2019.
- [94] D. Ganesan, P. Bashyam, C. V. Marappan *et al.*: *Bounds of spreading rate of virus for a network through an intuitionistic fuzzy graph*. U: A. K. Sangaiah, S. P. Shantharajah, i P. Theagarajan (urednici): *Intelligent pervasive computing systems for smarter healthcare*, strane 291–311. John Wiley & Sons, Inc., 2019.
- [95] S. A. Angadi, S. M. Hatture: *Face recognition through symbolic modeling of face graphs and texture*. Int. J. Pattern Recognit. Artif. Intell., **33**:1956008, 2019.
- [96] E. Estrada: *Characterization of 3D molecular structure*. Chem. Phys. Lett., **319**:713–718, 2000.
- [97] E. Estrada: *Characterization of the folding degree of proteins*. Bioinformatics, **18**:697–704, 2002.
- [98] E. Estrada: *Characterization of the amino acid contribution to the folding degree of proteins*. Proteins, **54**:727–737, 2004.
- [99] B. Li: *New bounds for Estrada index*. MATCH Commun. Math. Comput. Chem., **77**:701–706, 2017.
- [100] E. Estrada, J. A. Rodriguez–Velazquez: *Subgraph centrality in complex networks*. Phys. Rev. E, **71**:056103, 2005.
- [101] E. Estrada, J. A. Rodriguez–Velazquez: *Spectral measures of bipartivity in complex networks*. Phys. Rev. E, **72**:046105, 2005.
- [102] L. C. Freeman: *Centrality in social networks conceptual clarification*. Soc. Netw., **1**:215–239, 1979.
- [103] A. L. Barabasi, M. Posfai: *Network science*. Cambridge University Press, 2016.
- [104] I. Gutman, H. Deng, S. Radenković: *The Estrada index: an updated survey*. Zbornik radova, **14**:155–174, 2011.
- [105] E. Estrada, N. Hatano: *Statistical–mechanical approach to subgraph centrality in complex networks*. Chem. Phys. Lett., **439**:247–251, 2007.

- [106] E. Estrada, J. A. Rodriguez-Velazquez, M. Randić: *Atomic branching in molecules*. Int. J. Quantum Chem., **106**:823–832, 2006.
- [107] A. Gursoy, O. Keskin, R. Nussinov: *Topological properties of protein interaction networks from a structural perspective*. Biochem. Soc. Trans., **36**:1398–1403, 2008.
- [108] A. Ilić, D. Stevanović: *The Estrada index of chemical trees*. J. Math. Chem., **47**:305–314, 2010.
- [109] W. H. Wang, Y. S. Xue: *Extremal Estrada indices of the weighted trees with fixed total weight sum*. Appl. Math. Comput., **354**:32–41, 2019.
- [110] G. P. Clemente, A. Cornaro: *Novel bounds for the normalized Laplacian Estrada index and normalized Laplacian energy*. MATCH Commun. Math. Comput. Chem., **77**:673–690, 2017.
- [111] X. Chen, J. Qian: *On resolvent Estrada index*. MATCH Commun. Math. Comput. Chem., **73**:163–174, 2015.
- [112] J. Li, L. Qiao, N. Gao: *Bounds of the extended Estrada index of graphs*. Appl. Math. Comput., **317**:143–149, 2018.
- [113] Y. Shang: *Estrada index of general weighted graphs*. Bull. Aust. Math. Soc., **88**:106–112, 2013.
- [114] E. Estrada, A. A. Alhomaidhi, F. Al-Thukair: *Exploring the "Middle Earth" of network spectra via a Gaussian matrix function*. Chaos, **27**:023109, 2017.
- [115] Y. Shang: *Lower bounds for Gaussian Estrada index of graphs*. Symmetry, **10**:325, 2018.
- [116] I. Gutman, B. Furtula, E. Zogić *et al.*: *Resolvent energy of graphs*. MATCH Commun. Math. Comput. Chem., **75**:279–290, 2016.
- [117] T. S. Shores: *Applied linear algebra and matrix analysis*. Springer-Verlag New York, 2007.
- [118] L. E. Allem, J. Capaverde, V. Trevisan *et al.*: *Resolvent energy of unicyclic, bicyclic and tricyclic graphs*. MATCH Commun. Math. Comput. Chem., **77**:95–104, 2017.
- [119] A. Farrugia: *The increase in the resolvent energy of a graph due to the addition of a new edge*. Appl. Math. Comput., **321**:25–36, 2018.
- [120] Z. Zhu: *Some extremal properties of the resolvent energy, Estrada and resolvent Estrada indices of graphs*. J. Math. Anal. Appl., **447**:957–970, 2017.
- [121] E. Zogić, E. Glogić: *New bounds for the resolvent energy of graphs*. Sci. Pub. State Uni. Novi Pazar Ser. A: Appl. Math. Inform. Mech., **9**:187–191, 2017.
- [122] E. Zogić: *Neke osobine rezolventne i Randićeve energije grafa*. Doktorska disertacija, Univerzitet u Kragujevcu, Prirodno–matematički fakultet, 2018.
- [123] B. Furtula, I. Gutman, M. Dehmer: *On structure–sensitivity of degree–based topological indices*. Appl. Math. Comput., **219**:8973–8978, 2013.
- [124] X. Gao, B. Xiao, D. Tao *et al.*: *A survey of graph edit distance*. Pattern Anal. Appl., **13**:113–129, 2010.

- [125] M. Rakić, B. Furtula: *A novel method for measuring the structure sensitivity of molecular descriptors*. J. Chemom., **33**:e3138, 2019.
- [126] D. Bajusz, A. Racz, K. Heberger: *Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?* J. Cheminformatics, **7**:20, 2015.
- [127] D. Rogers, M. Hahn: *Extended-connectivity fingerprints*. J. Chem. Inf. Model., **50**:742–754, 2010.
- [128] R. Breslow, E. Mohacsi: *Studies on d-orbital conjugation. III. Non-aromaticity of a derivative of the 1, 3-dithiepinyl anion, a ten π -electron conjugated system*. J. Am. Chem. Soc., **85**:431–434, 1963.
- [129] H. Hosoya, K. Hosoi, I. Gutman: *A topological index for the total π -electron energy*. Theor. Chim. Acta, **38**:37–47, 1975.
- [130] I. Gutman, S. Bosanac: *Quantitative approach to Hückel rule the relations between the cycles of a molecular graph and the thermodynamic stability of a conjugated molecule*. Tetrahedron, **33**:1809–1812, 1977.
- [131] I. Gutman: *Effect of cycles on total π -electron energy of alternant conjugated hydrocarbons*. J. Chem. Soc., Faraday Trans., **75**:799–805, 1979.
- [132] I. Gutman: *Mathematical modeling of chemical phenomena*. U: *Advances in Nonlinear Sciences II*, strane 108–134. Academy of Nonlinear Sciences, 2008.
- [133] S. Bosanac, I. Gutman: *Effect of a ring on the stability of polycyclic conjugated molecules*. Z. Naturforsch., **32a**:10–12, 1977.
- [134] J. J. Morris, P. P. Bruneau: *Prediction of physicochemical properties*. U: H. J. Bohm i G. Schneider (urednici): *Virtual screening for bioactive molecules*, strane 33–58. Wiley-VCH, 2000.
- [135] I. V. Tetko: *The prediction of physicochemical properties*. U: S. Ekins (urednik): *Computational toxicology: risk assessment for pharmaceutical and environmental chemicals*, strane 241–275. John Wiley & Sons, Inc., 2007.
- [136] J. Dearden, A. Worth: *In silico prediction of physicochemical properties*. JRC Sci. Tech. Rep., **23051**:1–68, 2007.
- [137] J. C. Dearden: *Prediction of physicochemical properties*. U: B. Reisfeld i A. Mayeno (urednici): *Computational toxicology*, strane 93–138. Humana Press, 2012.
- [138] I. V. Tetko, A. Yan, J. Gasteiger: *Prediction of physicochemical properties of compounds*. U: T. Engel i J. Gasteiger (urednici): *Applied chemoinformatics: achievements and future opportunities*, strane 53–81. Wiley-VCH, 2018.
- [139] A. A. Toropov, A. P. Toropova: *QSPR/QSAR: state-of-art, weirdness, the future*. Molecules, **25**:1292, 2020.
- [140] C. Hansch, T. Fujita: *p - σ - π analysis. A method for the correlation of biological activity and chemical structure*. J. Am. Chem. Soc., **86**:1616–1626, 1964.
- [141] *PubChem*. <https://pubchem.ncbi.nlm.nih.gov>.

- [142] S. Kim, P. A. Thiessen, E. E. Bolton *et al.*: *PubChem substance and compound databases*. *Nucleic Acids Res.*, **44**:D1202–D1213, 2016.
- [143] *NIST Chemistry Webbook*. <https://webbook.nist.gov/chemistry>.
- [144] P. J. Linstrom, W. G. Mallard: *The NIST Chemistry WebBook: a chemical data resource on the internet*. *J. Chem. Eng. Data*, **46**:1059–1063, 2001.
- [145] W. Sippl, D. Robaa: *QSAR/QSPR*. U: T. Engel i J. Gasteiger (urednici): *Applied chemoinformatics: achievements and future opportunities*, strane 9–52. Wiley–VCH, 2018.
- [146] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*: *Scikit–learn: machine learning in Python*. *J. Mach. Learn. Res.*, **12**:2825–2830, 2011.
- [147] G. B. Luilo, S. E. Cabaniss: *Quantitative structure–property relationship for predicting chlorine demand by organic molecules*. *Environ. Sci. Technol.*, **44**(7):2503–2508, 2010.
- [148] M. Fernandez, J. Caballero: *QSAR models for predicting the activity of non–peptide luteinizing hormone–releasing hormone (LHRH) antagonists derived from erythromycin A using quantum chemical properties*. *J. Mol. Model.*, **13**:465–476, 2007.
- [149] E. Papa, S. Kovarich, P. Gramatica: *QSAR modeling and prediction of the endocrine–disrupting potencies of brominated flame retardants*. *Chem. Res. Toxicol.*, **23**:946–954, 2010.
- [150] E. Pourbasheer, S. Riahi, M. R. Ganjali *et al.*: *QSAR study on melanocortin–4 receptors by support vector machine*. *Eur. J. Med. Chem.*, **45**:1087–1093, 2010.
- [151] N. R. Draper, H. Smith: *Applied regression analysis*. John Wiley & Sons, Inc., 1998.
- [152] C. Gourieroux, A. Holly, A. Monfort: *Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters*. *Econometrica*, **50**:63–80, 1982.
- [153] S. Sommer, R. M. Huggins: *Variables selection using the Wald test and a robust Cp*. *Appl. Stat.*, **45**:15–29, 1996.
- [154] P. Gramatica: *Principles of QSAR models validation: internal and external*. *QSAR Comb. Sci.*, **26**:694–701, 2007.
- [155] R. Kiralj, M. Ferreira: *Basic validation procedures for regression models in QSAR and QSPR studies: theory and application*. *J. Braz. Chem. Soc.*, **20**:770–787, 2009.
- [156] R. Veerasamy, H. Rajak, A. Jain *et al.*: *Validation of QSAR models–strategies and importance*. *Int. J. Drug Des. Discov.*, **2**:511–519, 2011.
- [157] Organisation for Economic Co–operation and Development: *Guidance document on the validation of (quantitative) structure–activity relationship [(Q)SAR] models*. *OECD Env. Health Safety Pub.*, **69**, 2007.
- [158] D. L. J. Alexander, A. Tropsha, D. A. Winkler: *Beware of R^2 : simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models*. *J. Chem. Inf. Model.*, **55**:1316–1322, 2015.

- [159] R. D. Cramer III, J. D. Bunce, D. E. Patterson *et al.*: *Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies*. *Quant. Struct. Act. Rel.*, **7**:18–25, 1988.
- [160] R. Wehrens, W. E. van der Linden: *Bootstrapping principal component regression models*. *J. Chemom.*, **11**:157–171, 1997.
- [161] R. Wehrens, H. Putter, L. M. C. Buydens: *The bootstrap: a tutorial*. *Chemometr. Intell. Lab. Syst.*, **54**:35–52, 2000.
- [162] I. Redžepović, B. Furtula: *On relationships of eigenvalue-based topological molecular descriptors*. *Acta Chim. Slov.*, **67**:312–318, 2020.
- [163] I. Redžepović, B. Furtula: *Resolvent energy and Estrada index of benzenoid hydrocarbons*. *J. Serb. Soc. Comput. Mech.*, **special issue**:37–44, 2020.
- [164] I. Gutman, S. J. Cyvin: *Introduction to the theory of benzenoid hydrocarbons*. Springer Verlag, 1989.
- [165] G. Brinkmann, O. D. Friedrichs, S. Lisken *et al.*: *CaGe—a virtual environment for studying some special classes of plane graphs—an update*. *MATCH Commun. Math. Comput. Chem.*, **63**:533–552, 2010.
- [166] I. Redžepović, B. Furtula: *On degeneracy of \mathcal{A} -eigenvalue-based molecular descriptors and r -equienergetic chemical trees*. *MATCH Commun. Math. Comput. Chem.*, **84**:385–397, 2020.
- [167] M. Randić: *Molecular bonding profiles*. *J. Math. Chem.*, **19**:375–392, 1996.
- [168] E. V. Konstantinova: *The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems*. *J. Chem. Inf. Comput. Sci.*, **36**:54–57, 1996.
- [169] O. Miljković, B. Furtula, S. Radenković *et al.*: *Equienergetic and almost-equienergetic trees*. *MATCH Commun. Math. Comput. Chem.*, **61**:451–461, 2009.
- [170] F. Johansson: *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)*, 2013. <http://mpmath.org/>.
- [171] I. Gutman, I. Redžepović, B. Furtula: *Two stability criteria for benzenoid hydrocarbons and their relation*. *Croat. Chem. Acta*, **92**:473–475, 2019.
- [172] J. Langner, H. A. Witek: *Interface theory of benzenoids*. *MATCH Commun. Math. Comput. Chem.*, **84**:143–176, 2020.
- [173] J. Langner, H. A. Witek: *Interface theory of benzenoids: basic applications*. *MATCH Commun. Math. Comput. Chem.*, **84**:177–215, 2020.
- [174] I. Redžepović, B. Furtula, I. Gutman: *Relating total π -electron energy of benzenoid hydrocarbons with HOMO and LOMO energies*. *MATCH Commun. Math. Comput. Chem.*, **84**:229–237, 2020.
- [175] I. Redžepović, B. Furtula: *Comparative study on structural sensitivity of eigenvalue-based molecular descriptors*. *J. Math. Chem.*, **59**:476–487, 2021.

- [176] I. Redžepović, S. Radenković, B. Furtula: *Effect of a ring onto values of eigenvalue-based molecular descriptors*. *Symmetry*, **13**:1515, 2021.
- [177] I. Redžepović, B. Furtula: *Predictive potential of eigenvalue-based topological molecular descriptors*. *J. Comput. Aided Mol. Des.*, **34**:975–982, 2020.
- [178] S. C. Basak, G. J. Niemi, G. D. Veith: *Predicting properties of molecules using graph invariants*. *J. Math. Chem.*, **7**:243–272, 1991.
- [179] S. C. Basak, S. Nikolić, N. Trinajstić *et al.*: *QSPR modeling: graph connectivity indices versus line graph connectivity indices*. *J. Chem. Inf. Comput. Sci.*, **40**:927–933, 2000.
- [180] N. Trinajstić, S. Nikolić, S. C. Basak *et al.*: *Distance indices and their hyper-counterparts: intercorrelation and use in the structure-property modeling*. *SAR QSAR Environ. Res.*, **12**:31–54, 2001.
- [181] *Molecular descriptors*. <https://www.moleculardescriptors.eu>.
- [182] B. Furtula, I. Gutman: *A forgotten topological index*. *J. Math. Chem.*, **53**:1184–1190, 2015.
- [183] D. E. Leahy: *Intrinsic molecular volume as a measure of the cavity term in linear solvation energy relationships: octanol-water partition coefficients and aqueous solubilities*. *J. Pharm. Sci.*, **75**:629–636, 1986.

Indeks

- graph edit distance* (GED), 13
- Balabanov indeks, 7
- Butstrap tehnika, 23
- cikl, 3
- Efekat cikla, 15
- Energija grafa, 10
- Estradin indeks, 11
- Graf, 3
- grafovske invarijante, 3
- Hamiltonov operator, 9
- Hararijev indeks, 7
- Hemijska teorija grafova, 3
- Hikelova molekulska-orbitalna (HMO) teorija, 9
- Hikelovo pravilo, 16
- integral rezonancije, 9
- karakteristični polinom, 4
- kenogram, 3
- Koeficijent determinacije, 22
- koren kvadrata srednje greške, 23
- Laplasova matrica, 4
- Linijski graf, 11
- matrica preklapanja, 9
- matrica rastojanja, 4
- matrica susedstva, 3
- molekulski deskriptor, 1
- molekulski graf, 3
- Morganovi kružni fingerprintovi, 14
- Plerogram, 3
- poboljšani koeficijent determinacije, 22
- prosečna apsolutna greška, 23
- prosečna relativna greška, 23
- put grafa, 5
- QSAR, 19
- QSPR, 19
- QSTR, 19
- Randićev indeks, 8
- rastojanje čvorova, 3
- Rezolventna energija, 12
- Rezolventna matrica, 12
- robustnost, 13
- Saksova teorema, 5
- Saksovi grafovi, 5
- Sopstvene vrednosti matrice susedstva, 4
- spektar grafa, 4
- spektralni momenti, 6
- stablo, 3
- stepen čvora, 3
- strukturna osetljivost, 13
- Tanimoto indeksa, 13
- Teorija grafova, 3
- test skup, 19
- topološki molekulska deskriptori, 3
- training skup, 19
- Valdov test, 21
- Vinerov indeks, 7
- Višestruka linearna regresija, 21
- zagrebački indeksi, 8
- Kulonov integral, 9



A

Publikovani radovi u kojima su
prezentovani rezultati doktorske
disertacije

- Rahgozar, S; Khalil; Alifrangis, M; Natah, S; Groop, L; Shaat, N; Kozlov, A; Verschubskaya, G; Comas, D; Bulayeva, K; Mehdi, S. Q; Terwilliger, J D.; Sahi, T; Savilahti, E; Perola, M; Sajantila, A; Järvelä, I; Peltonen, L. Evidence of Still-Ongoing Convergence Evolution of the Lactase Persistence T-13910 Alleles in Humans. *Am J Hum Genet* (2007) 81 (3): 615–25.
- 13 Itan Y1, Jones BL, Ingram CJ, Swallow DM, Thomas MG. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol. Biol.* (2010) 9: 10:36
- 14 Hoyt G, Hickey MS, Cordain L. Dissociation of the glycaemic and insulinaemic responses to whole and skimmed milk. *Br. J. Nutr.* (2005) 93:175–7.
- 15 Nilsson M, Stenberg M, Frid AH, Holst JJ, Björck IM. Glycemia and insulinemia in healthy subjects after lactose-equivalent meals of milk and other food proteins: the role of plasma amino acids and incretins. *Am. J. Clin. Nutr.* (2004) 80:1246–53.
- 16 Melnik B. Milk consumption: aggravating factor of acne and promoter of chronic diseases of Western societies. *J. Dtsch. Dermatol. Ges.* (2009) 4:364–70.
- 17 Hoppe C, Mølgaard C, Vaag A, Barkholt V, Michaelsen KF: High intakes of milk, but not meat, increases s-insulin and insulin resistance in 8-year-old boys. *Eur. J. Clin. Nutr.* (2005) 59:393–8.
- 18 Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function". *Cell.* (2004) 116 (2): 281–97.
- 19 Chen X, Liang H, Zhang J, Zen K, Zhang CY. Horizontal transfer of microRNAs: molecular mechanisms and clinical applications. *Protein Cell* (2012), 3:28–37.
- 20 Izumi H, Kosaka N, Shimizu T, Sekine K, Ochiya T, Takase M. Bovine milk contains microRNA and messenger RNA that are stable under degradative conditions. *J. Dairy Sci.* (2012) 95:4831–41.
- 21 Meng F, Henson R, Wehbe-Janeck H, Ghoshal K, Jacob ST, Patel T. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* (2007) 133:647–58.
- 22 Asangani IA, Rasheed SA, Nikolova DA, Leupold JH, Colburn NH, Post S, Allgayer H. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* (2008) 27:2128–36.
- 23 Julia Ramírez-Moya and Pilar Santisteban. miRNA-Directed Regulation of the Main Signaling Pathways in Thyroid Cancer. *Front Endocrinol (Lausanne)*. (2019) 10: 430
- 24 Foster KG, Fingar DC. Mammalian target of rapamycin (mTOR): conducting the cellular signalling symphony. *J. Biol. Chem.* (2010) 285:14071–7.
- 25 Xu G, Kwon G, Cruz WS, Marshall CA, McDaniel MK. Metabolic Regulation by Leucine of Translation Initiation Through the mTOR-signaling Pathway by Pancreatic beta-cells. *Diabetes*. (2001) 50(2): 353-60.
- 26 Carayol N, Katsoulidis E, Sassano A, Altman JK, Druker BJ, Plataniias LC. Suppression of programmed cell death 4 (PDCD4) protein expression by BCR-ABL-regulated engagement of the mTOR/p70 S6 kinase pathway. *J. Biol Chem* (2008) 283:8601–10.



ИЗУДИН РЕЦЕПОВИЋ, БОРИС ФУРТУЛА,

Природно-математички факултет, Универзитет у Крагујевцу;

(e-mail: izudin.redzepovic@pmf.kg.ac.rs, boris.furtula@pmf.kg.ac.rs)

ТОПОЛОШКИ МОЛЕКУЛСКИ ДЕСКРИПТОРИ

УВОД

Пресудан корак у објашњавању природних појава се заснива на препознавању или дефинисању параметара који их описују и јасно диференцирају од других. Овакав приступ се примењује и у хемији [1]. Идеја да молекулска структура носи информације везане за сам молекул заузима централно место у хемији. Иако ова веза између структуре и особина молекула изгледа логична и једноставна, неретко ју је тешко објаснити. Екстраховање и манипулација информацијама које се добијају из саме структуре молекула је омогућено употребом тзв. молекулских дескриптора [2].

Молекулски дескриптор представља број или скуп бројева који се могу приписати неком једињењу. Они могу бити изведени из молекулске структуре применом различитих алгоритама. Због тога постоји мноштво молекулских дескриптора који су уведени са различитим циљевима. Разликују се по количини информација које носе, што ствара и разлику у времену потребном за њихово израчунавање. Молекулски дескриптори који носе више информација углавном су засновани на комплексним математичким дефиницијама и изискују много компјутерског времена за њихово рачунање. Молекулска тежина, на пример, представља врло једноставан молекулски дескриптор који се лако и брзо рачуна, али, сем масе молекула, она нам не нуди никакве додатне корисне информације.

Scientific paper

On Relationships of Eigenvalue–Based Topological Molecular Descriptors

Izudin Redžepović and Boris Furtula *

Faculty of Science, University of Kragujevac, P. O. Box 60, 34000 Kragujevac, Serbia

* Corresponding author: E-mail: boris.furtula@pmf.kg.ac.rs

Received: 08-23-2019

Abstract

Three eigenvalue-based topological molecular descriptors are compared using several datasets of alkanes. Two of them are well-known and frequently employed in various QSPR/QSAR investigations, and third-one is a newly derived whose predictive potential is yet to be proven. The relations among them are found and discussed. Structural parameters that govern these relations are identified and the corresponding formulas based on multiple linear regression have been obtained. It has been shown that all three investigated indices are encoding almost the same structural information of a molecule. They differ only by the extent of the sensitivity on a structural branching of a molecule and on the number of non-bonding molecular orbitals.

Keywords: Graph energy, Estrada index, Resolvent energy of a graph, eigenvalues, adjacency matrix.

1. Introduction

Molecular descriptors are the fundamental tools in QSPR/QSAR modeling, which are frequently employed in diverse fields of chemistry.^{1–3} Among them, topological indices are the usual choice, because of their low computational complexity and fairly simple identification of structure–property relationships.^{4–7} There are hundreds of topological descriptors.² A natural way for their classification is by the origin of parameters that are used in their definitions. Thus, one differentiates degree–, distance–, and eigenvalue–based topological molecular descriptors, although there is a couple of them that cannot be strictly designated as members of any of the above-mentioned classes.

Interest for the eigenvalue-based topological molecular descriptors had been aroused after the explanation of the physical meaning of eigenvalues in HMO theory.⁸ This happened in the seventies of the last century. Probably the first eigenvalue-based topological descriptor that had been introduced is the graph energy. This index is defined using the eigenvalues of an “ordinary” adjacency matrix in the following way:

$$E(G) = \sum_{i=1}^n |\lambda_i| \quad (1)$$

where λ_i is the i -th eigenvalue of a graph G .

The graph energy is tightly connected to the total π -electron energy of alternate conjugated molecules. It is a popular research topic both in chemically, and in mathematically oriented investigations. Several books and numerous papers are devoted to this particular topological invariant.⁹ Nowadays, there are numerous eigenvalue-based topological indices, but just a couple of them are based on the eigenvalues that come from the adjacency matrix. These indices have been used as molds for defining almost all other topological invariants belonging to this class. Thus, beside graph energy, one could find indices like Laplacian energy, distance energy, Randić energy, etc. (e.g. see^{10–12}).

Next to graph energy, the second most investigated topological molecular descriptor based on eigenvalues of an adjacency matrix is Estrada index. It was designed to model the folding in some biomolecules.¹³ Estrada index is defined as follows:

$$EE(G) = \sum_{i=1}^n e^{\lambda_i} \quad (2)$$

where λ_i is the i -th eigenvalue of a graph G .

Its undeniable success led to a vigorous research of this quantity (see^{14–16} and references cited therein). This invoked the introduction of many other Estrada-like invariants.^{10,17–19}

Resolvent Energy and Estrada Index of Benzenoid Hydrocarbons

Izudin Redžepović¹, Boris Furtula^{1*}

¹ Faculty of Science, University of Kragujevac, P. O. Box 60, 12 Radoja Domanovica Street, 34000 Kragujevac, Serbia
e-mail: izudin.redzepovic@pmf.kg.ac.rs, boris.furtula@pmf.kg.ac.rs

**corresponding author*

Abstract

The relationship between the resolvent and Estrada indices for the alkanes has been recently demonstrated. That relationship involved the first Zagreb index besides these two eigenvalue-based molecular descriptors. In this paper, the quality of the correlation is tested in the case of isomeric benzenoid hydrocarbons, where the first Zagreb index is constant. Extraordinary linear correlations are identified for all studied groups of isomeric benzenoid hydrocarbons. Additionally, the relationship of these indices with the boiling points of a set of benzenoid hydrocarbons is presented.

Keywords: molecular descriptors, eigenvalues, adjacency matrix, computational modeling, molecular structure

1. Introduction

Information, stored within the molecular structure, is usually accounted for determining physico-chemical features of the corresponding compound. Extracting this information represents an important task when it comes to the modeling of molecular physico-chemical properties. This is ordinarily done by using molecular descriptors (Todeschini and Consonni 2009). Nowadays, there are a lot of such descriptors, and their number is still increasing. The most prominent group of molecular descriptors are those derived from 2D molecular structure, namely, topological descriptors. Transforming structural information into a topological index causes the loss of pieces of information carried by the molecular structure. Nevertheless, topological descriptors found remarkable applications in diverse fields of chemistry (Dudek et al. 2006, Mauri et al. 2017). One of the most common utilization of the topological indices is in QSPR/QSAR modeling (Devillers and Balaban 1999, Talevi et al. 2012, Roy et al. 2015, Zanni et al. 2015, Dearden 2017).

The graph theory found a significant application in chemistry through the use of topological descriptors. Namely, in chemical graph theory, a molecule is represented by the corresponding graph G in which vertices stand for non-hydrogen atoms and edges for bonds. Topological indices are also known as graph invariants. They can be computed from the molecular graph by applying different algorithms. Thus, depending on what parameters are used

On Degeneracy of \mathcal{A} -Eigenvalue–Based Molecular Descriptors and r -Equienergetic Chemical Trees

Izudin Redžepović, Boris Furtula*

Faculty of Science, University of Kragujevac

P. O. Box 60, 34000 Kragujevac, Serbia

izudin.redzepovic@pmf.kg.ac.rs , furtula@uni.kg.ac.rs

(Received May 9, 2020)

Abstract

Among more than two hundreds of eigenvalue–based topological indices only a couple of them are defined using the eigenvalues devised from the adjacency matrix of a graph. The resolvent energy is probably the most recent-one such an index. In this article, the degeneracy of the energy, Estrada index, and the resolvent energy is presented. The specious degeneracy of the resolvent energy in the case of chemical trees is discussed. Then, the data on searching for resolvent equienergetic chemical trees is given.

1 Introduction

A need for the quantification of a molecular structure can be traced back to the very beginnings of the structural and theoretical chemistry. Numerous counting, topological, geometric, and quantum–based molecular descriptors have been devised for this purpose [36]. A prominent and probably the largest group among them is the topological indices. They can be assorted into two major classes of degree– and distance–based descriptors [10–12]. However, there are existing topological indices that cannot be fitted into any

*Corresponding author



Comparative study on structural sensitivity of eigenvalue–based molecular descriptors

Izudin Redžepović¹ · Boris Furtula¹

Received: 10 October 2020 / Accepted: 30 November 2020 / Published online: 1 January 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG part of Springer Nature 2021

Abstract

Structural sensitivity is one of the most important and the least investigated property of the topological molecular descriptors. This paper reports results on the structural sensitivity of graph energy, Estrada index, and resolvent energy on several series of catacondensed and pericondensed isomeric benzenoid hydrocarbons. Recently, a novel method for assessing the structural sensitivity of topological molecular descriptors was put forward, which is applied here. This algorithm consists of several different steps and it is based on Tanimoto index and Morgan circular fingerprints. It was found that graph energy, Estrada index, and resolvent energy exert similar structural sensitivity on catacondensed isomers. The graph energy showed the best performance on pericondensed molecules. Additionally, the sensitivities of these descriptors were tested on the catacondensed isomers with the increasing number of bays, coves, and fjords. It was revealed that these descriptors gradually change with the increasing number of these structural features. The Estrada index and resolvent energy perform similarly and in some cases with the same structural sensitivity. This may be attributed to the high correlation between them. The graph energy showed superiority over the Estrada index and resolvent energy on pericondensed isomers.

Keywords Graph energy · Estrada index · Resolvent energy · Benzenoid hydrocarbons · Eigenvalues · Molecular structure

✉ Boris Furtula
furtula@uni.kg.ac.rs

Izudin Redžepović
izudin.redzepovic@pmf.kg.ac.rs

¹ Department of Chemistry, Faculty of Science, University of Kragujevac, P.O. Box 60, Kragujevac 34000, Serbia

Article

Effect of a Ring onto Values of Eigenvalue–Based Molecular Descriptors

Izudin Redžepović [†], Slavko Radenković [†] and Boris Furtula ^{*,†}

Department of Chemistry, Faculty of Science, University of Kragujevac, 34000 Kragujevac, Serbia; izudin.redzepovic@pmf.kg.ac.rs (I.R.); slavko.radenkovic@pmf.kg.ac.rs (S.R.)

* Correspondence: furtula@uni.kg.ac.rs

† These authors contributed equally to this work.

Abstract: The eigenvalues of the characteristic polynomial of a graph are sensitive to its symmetry-related characteristics. Within this study, we have examined three eigenvalue–based molecular descriptors. These topological molecular descriptors, among others, are gathering information on the symmetry of a molecular graph. Furthermore, they are being ordinarily employed for predicting physico–chemical properties and/or biological activities of molecules. It has been shown that these indices describe well molecular features that are depending on fine structural details. Therefore, revealing the impact of structural details on the values of the eigenvalue–based topological indices should give a hunch how physico–chemical properties depend on them as well. Here, an effect of a ring in a molecule on the values of the graph energy, Estrada index and the resolvent energy of a graph is examined.

Keywords: molecular descriptor; graph energy; Estrada index; resolvent energy; effect of a ring

MSC: 05C09; 05C92



Citation: Redžepović, I.; Radenković, S.; Furtula, B. Effect of a Ring onto Values of Eigenvalue–Based Molecular Descriptors. *Symmetry* **2021**, *13*, 1515. <https://doi.org/10.3390/sym13081515>

Academic Editors: Jose M. Rodriguez and José M. Sigarreta

Received: 18 June 2021

Accepted: 26 July 2021

Published: 18 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Molecular descriptors are quantities that may be regarded as numerical representations of a molecule. They are carriers of information about some parts or the whole structure of a molecule and, therefore, implicitly about its physico–chemical properties. Because of their success in predicting the behavior of molecules, investigations of molecular descriptors have been intensive over the years. Such an activity resulted in introducing thousands of molecular descriptors [1]. Among them, a salient place is reserved for topological indices [2,3]. Reasons for this will not be detailed here, but, in a nutshell, the topological indices represent a good balance between the amount of information they are harvesting from a molecular structure and the computational complexity of their algorithms. The main applications of topological indices represent construction of mathematical predictive models (QSPR/QSAR) and in virtual screening [4]. In addition, some of them have proved to be useful in molecular design and in drug discovery [5]. They can be roughly classified into three major groups: degree–based [6], distance–based [7], and eigenvalue–based topological indices [8].

The eigenvalue–based topological molecular descriptors emerged in the seventies of the last century when a linkage between one of them and the total π -electron energy of conjugated molecules was established [9]. Nowadays, there are hundreds of these indices [10], but only a couple of them originate from the eigenvalues obtained from the adjacency matrix of a graph. These particular indices are mainly used in modeling the physico–chemical properties of molecules. In the rest of the text, we shall refer to them as the eigenvalue–based indices.



Predictive potential of eigenvalue-based topological molecular descriptors

Izudin Redžepović¹ · Boris Furtula¹

Received: 13 February 2020 / Accepted: 7 June 2020 / Published online: 13 June 2020
© Springer Nature Switzerland AG 2020

Abstract

This study is directed toward assessing the predictive potential of eigenvalue-based topological molecular descriptors. The graph energy, Estrada index, resolvent energy, and the Laplacian energy were tested as parameters for the prediction of boiling points, heats of formation, and octanol/water partition coefficients of alkanes. It was shown that an eigenvalue-based molecular descriptor cannot be individually used for successful prediction of these physico-chemical properties, but the first Zagreb index, the number of zeros in the spectrum and the number of methyl groups must be also involved in the models. Performed statistics show that the models constructed using the Estrada index and resolvent energy are significantly better than ones with the energy of a graph and the Laplacian energy. Such a trend is even more noticeable in the case of octanol/water partition coefficients of alkanes.

Keywords Boiling points · Heats of formation · Octanol/water partition coefficients · Molecular modeling · Graph invariants

Introduction

An idea that the molecular structure encodes information on the physico-chemical properties of the corresponding compound is a cornerstone of modern chemistry. Although the above sentence seems self-explanatory, it is not an easy task to fully understand connection between a physico-chemical feature and the structural details that determine it [1]. To unveil this relationship experimental data are needed, but unfortunately, they are not always available. Therefore, there are many studies with the aim to clarify this dependence, and to predict missing data (e.g. heat of formation) [2–6]. Those predictions are made by applying diverse methodologies, and as more advanced techniques become available (e.g. machine learning [7]), new approaches to an old problem are being developed.

Molecular descriptors are frequently employed tools for predicting various properties of a molecule. They are especially useful for constructing QSPR/QSAR models. Nowadays, a huge number of molecular descriptors are available [8–10]. Among them, an important place belongs to the topological molecular descriptors. These are mathematical quantities derived from a molecular structure. In chemical graph theory, a compound is represented by corresponding molecular graph G , in which vertices stand for non-hydrogen atoms and edges for bonds. From such an object, topological descriptors a.k.a. graph invariants are obtained by applying different algorithms. They have proved to be valuable measures because they are capable to encrypt considerable amount of structural information, and yet, they are not computationally demanding [11–14].

These indices can be classified into several groups by the nature of graph parameters used in their definitions. One of these groups is the so-called eigenvalue-based topological molecular indices (EI). A descriptor from this set is defined using eigenvalues that come from one of the graph matrices (e.g. adjacency matrix).

Probably, the very first such index is the graph energy. It is defined in the following way:



Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10822-020-00320-2>) contains supplementary material, which is available to authorized users.

✉ Boris Furtula
boris.furtula@pmf.kg.ac.rs

¹ Faculty of Science, University of Kragujevac, P. O. Box 60, 34000 Kragujevac, Serbia

B Biografija autora



Izudin Redžepović je rođen 26. decembra 1993. godine u Novom Pazaru. Osnovne akademske studije hemije je završio 2016. godine na Departmanu za hemijsko–tehnološke nauke, Državnog univerziteta u Novom Pazaru sa prosečnom ocenom 9,54. Master akademske studije je završio 2017. godine na Prirodno–matematičkom fakultetu, Univerziteta u Kragujevcu sa prosečnom ocenom 10,00. Naziv master rada je "Antioksidativna aktivnost kafeinske kiseline–mehanistička DFT studija". Doktorske akademske studije hemije je upisao školske 2017/18. godine na Prirodno–matematičkom fakultetu u Kragujevcu, pod mentorstvom prof. dr Borisa Furtule. Izudin Redžepović do sada ima 13 objavljenih radova u časopisima referisanim u Web of Science. Kompletan spisak njegovih radova je dostupan u bazama ORCID  i Google Scholar .

ИЗЈАВА АУТОРА О ОРИГИНАЛНОСТИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Ја, Изудин Реџеповић, изјављујем да докторска дисертација под насловом:

Компаративно испитивање молекулских дескриптора заснованих на сопственим вредностима

која је одбрањена на Природно-математичком факултету, Универзитета у Крагујевцу представља *оригинално ауторско дело* настало као резултат *сопственог истраживачког рада.*

Овом Изјавом такође потврђујем:

- да сам *једини аутор* наведене докторске дисертације,
- да у наведеној докторској дисертацији *нисам извршио/ла повреду* ауторског нити другог права интелектуалне својине других лица,
- да умножени примерак докторске дисертације у штампаној и електронској форми у чијем се прилогу налази ова Изјава садржи докторску дисертацију истоветну одбрањеној докторској дисертацији.

У Крагујевцу, 10. 09. 2021. године,

Изудин Реџеповић

потпис аутора

ИЗЈАВА АУТОРА О ИСКОРИШЋАВАЊУ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Ја, Изудин Реџеповић,

дозвољавам

не дозвољавам

Универзитетској библиотеци у Крагујевцу да начини два трајна умножена примерка у електронској форми докторске дисертације под насловом:

Компаративно испитивање молекулских дескриптора заснованих на сопственим вредностима

која је одбрањена на Природно-математичком факултету,

Универзитета у Крагујевцу, и то у целини, као и да по један примерак тако умножене докторске дисертације учини трајно доступним јавности путем дигиталног репозиторијума Универзитета у Крагујевцу и централног репозиторијума надлежног министарства, тако да припадници јавности могу начинити трајне умножене примерке у електронској форми наведене докторске дисертације путем *преузимања*.

Овом Изјавом такође

дозвољавам

не дозвољавам¹

¹ Уколико аутор изабере да не дозволи припадницима јавности да тако доступну докторску дисертацију користе под условима утврђеним једном од *Creative Commons* лиценци, то не искључује право припадника јавности да наведену докторску дисертацију користе у складу са одредбама Закона о ауторском и сродним правима.

припадницима јавности да тако доступну докторску дисертацију користе под условима утврђеним једном од следећих *Creative Commons* лиценци:

1) Ауторство

2) Ауторство - делити под истим условима

3) Ауторство - без прерада

4) Ауторство - некомерцијално

5) Ауторство - некомерцијално - делити под истим условима

6) Ауторство - некомерцијално - без прерада²

У Крагујевцу _____, 10. 09. 2021. године,

Izudin Redžepović
потпис аутора

² Молимо ауторе који су изабрали да дозволе припадницима јавности да тако доступну докторску дисертацију користе под условима утврђеним једном од *Creative Commons* лиценци да заокруже једну од понуђених лиценци. Детаљан садржај наведених лиценци доступан је на: <http://creativecommons.org.rs/>