



---

**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].

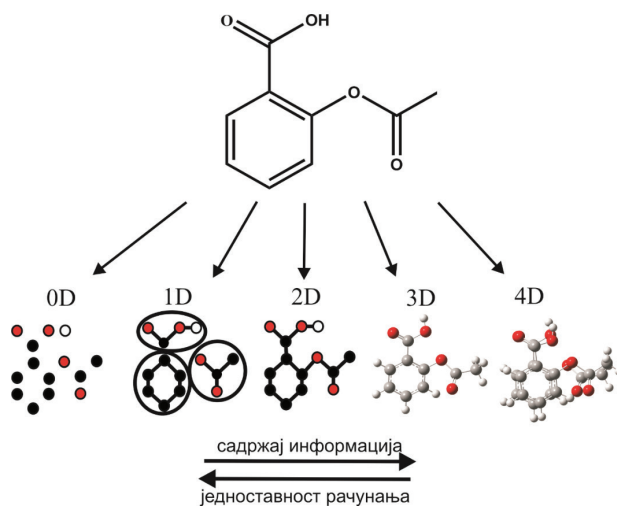
Молекулски дескриптор представља број или скуп бројева који се могу приписати неком једињењу. Они могу бити изведени из молекулске структуре применом различитих алгоритама. Због тога постоји мноштво молекулских дескриптора који су уведени са различитим циљевима. Разликују се по количини информација које 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.

С друге стране, дескриптори засновани на квантно-механичким прорачунима су врло захтевни за рачунање, али врло добро описују неке особине молекула [3].

Примене молекулских дескриптора у хемији су разноврсне, а најизраженији је њихов допринос у претраживању хемијских база и у QSPR/QSAR (quantitative structure property/activity relationships) истраживањима. Ова метода има за циљ идентификацију и квантификовање везе између структуре и посматране физичко-хемијске особине, односно, активности молекула. Крајњи производ ових студија је математички модел који у својој дефиницији садржи неки дескриптор, или још чешће, комбинацију два или више дескриптора.

Број молекулских дескриптора се мери хиљадама и могу се груписати у неколико категорија (Слика 1). Општеприхваћена класификација се заснива на „димензионалности“ молекулске структуре из које се изводи дескриптор:

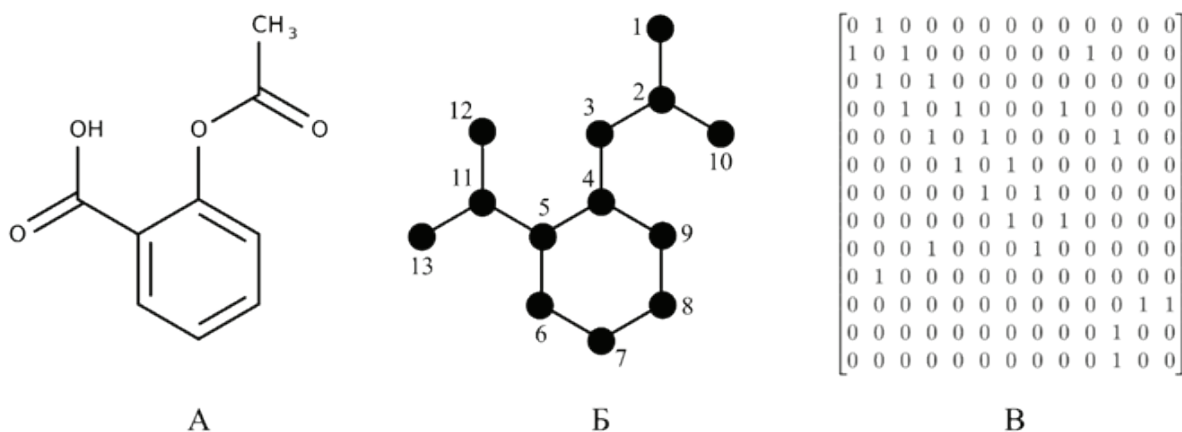
- 0D дескриптори су најједноставнији „бројачки“ дескриптори који се добијају из молекулске формуле (нпр. број азотових атома у молекулу);
- 1D дескриптори се заснивају, такође, на једноставном пребројавању одређених детаља у молекулу (нпр. број одређених функционалних група);
- 2D дескриптори се рачунају из дводимензионалних структурних формула молекула (у ову групу спадају тополошки дескриптори);
- 3D дескриптори узимају у обзир просторни распоред атома и веза у молекулу;
- 4D дескриптори, поред тродимензионалне структуре узимају у обзир још једну димензију, нпр. време.



Слика 1. Подела молекулских дескриптора.

## ТОПОЛОШКИ ДЕСКРИПТОРИ

Молекулски дескриптори засновани на дводимензионалним структурним формулама представљају оптималан компромис између количине корисних информација које носе о структури и комплексности њиховог израчунавања. Због тога заузимају значајно место у QSPR/QSAR истраживањима и њихов број се мери хиљадама. Посебну класу ових молекулских дескриптора представљају тополошки индекси. Засновани су на молекулском графу (Слика 2), код кога су атоми представљени чворовима ( $n$ ), а хемијске везе гранама ( $e$ ) које повезују одређене парове чворова. Због тога се ови дескриптори називају још и графовске инваријанте. Област хемије која објашњава хемијске појаве применом теорије графова назива се хемијска теорија графова [4,6]. Најпопуларнији су тзв. једноставни графови, код којих се не води рачуна о типу атома и веза представљених чворовима и гранама, док су водоникови атоми потпуно занемарени. Један та-



Слика 2. А. Структурна молекулска формула аспирина; Б. Молекулски граф аспирина; В. Матрица суседства изведена из графа Б.

кав молекулски граф је представљен на слици 2.Б. Молекулски граф се може претворити у матрицу суседства (A). Ова матрица је квадратна матрица димензије  $n \times n$ , која носи информације о повезаности чворова унутар графа. Елементи матрице суседства су 0 и 1. Уколико су два чвора суседна у графу (тј. ако постоји веза између њих) онда је  $a_{i,j} = 1$ , у супротном  $a_{i,j} = 0$ .

Главна улога ових дескриптора је квантификовање информације која је садржана у дводимензионалној структури молекула представљеној помоћу графа. На тај начин се врши екстракција информација која се потом може употребити за конструкцију математичког модела ради описивања неке особине молекула. Већина тополошких дескриптора се могу сврстати у једну од следеће три групе:

1. дескриптори засновани на степену чвора (degree-based);
2. дескриптори засновани на растојању чворова (distance-based);
3. дескриптори засновани на сопственим вредностима неке од матрица (eigenvalue-based).

#### Дескриптори засновани на степену чвора

Степен чвора (deg) представља број суседних чворова са којима је дати чвор повезан. Постоји велики број тополошких дескриптора који у својој дефиницији узимају у обзир степен чвора. Први дескриптор из ове класе спада у групу најстаријих тополошких молекулских индекса. Увео га је амерички научник Џон Р. Плат (John R. Platt) 1952. године и данас он носи име по свом творцу. Међутим, овај индекс се данас ретко среће у стручној и научној литератури, иако његов потенцијал није занемарујући. Најпопуларнији тополошки молекулски дескриптори засновани на степену чворова су тзв. загребачки индекси (први  $M_1$  и други  $M_2$ ) [7,8] и Рандићев индекс ( $\chi$ ) [9]. Дефинисани су на следећи начин:

$$M_1 = M_1(G) = \sum_u \text{deg}(u)^2 \quad (1)$$

$$M_2 = M_2(G) = \sum_{u,v} \text{deg}(u) \times \text{deg}(v) \quad (2)$$

$$\chi = \chi(G) = \sum_{u,v} \frac{1}{\sqrt{\text{deg}(u) \times \text{deg}(v)}} \quad (3)$$

где  $u$  и  $v$  представљају суседне чворове.

Ови индекси углавном прикупљају информације о „разгранатости“ 2Д молекулских структурних формула. Пронашли су примене у

најразноврснијим гранама хемије, а неки од њих су коришћени и у другим областима науке. Рандићев индекс је један од најпопуларнијих молекулских дескриптора о коме је написано више књига и преко хиљаду научних радова.

#### Дескриптори засновани на растојању чворова

Растојање чворова ( $d$ ) у повезаним графовима је дефинисано као најкраћи пут (број веза) између два чвора. Први тополошки индекс икада употребљен управо припада овој групи. Назива се Винеров индекс ( $W$ ) по америчком научнику Хари Винеру (Harry Wiener), који га је увео 1947. године [10].

$$W = W(G) = \sum_{u,v} d(u,v). \quad (4)$$

Овај индекс је Винер користио за предвиђање тачке кључања алкана. Било је потребно да протекне више година од објављивања његовог рада [10] како би се прихватио концепт моделирања физичко-хемијских особина молекула помоћу параметара изведених из 2D структуре молекула. Интензивно проучавање и увођење великог броја тополошких молекулских дескриптора почело је тек крајем шездесетих и почетком седамдесетих година прошлог века.

#### Дескриптори засновани на сопственим вредностима неке од матрица графа

Матрица суседства је уобичајен вид претварања графичког приказа молекулских графова у облик који се може унети у рачунар и даље обрађивати. Међутим, поред ове матрице постоје и друге које су нашле примену како у хемијској теорији графова, тако и у другим областима хемије [11]. Неке од познатијих су Лапласова матрица и матрица растојања. Елемент Лапласове матрице на пресеку  $i$ -врсте и  $j$ -те колоне је одређен на следећи начин:

$$l_{ij} = \begin{cases} -1 & \text{ако су чворови } u_i \text{ и } v_j \text{ суседни; } i \neq j \\ 0 & \text{ако чворови } u_i \text{ и } v_j \text{ нису суседни; } i \neq j \\ \text{deg}(i) & \text{ако је } i = j. \end{cases}$$

Елемент матрице растојања на пресеку  $i$ -врсте и  $j$ -те колоне је растојање између чворова  $i$  и  $j$ . Наравно, неопходно је да граф буде повезан. Сопствене вредности које се добијају из ових матрица се користе за рачунање важне класе тополошких молекулских дескриптора, који су данас веома актуелни. Један од најстаријих индекса ове групе је енергија графа ( $E$ ) [12], који се интензивно проучава како у хемији тако и у математици. Друга два значајна индекса из ове групе су Естрадаин индекс ( $EE$ ) [13] и резолвентна енергија ( $ER$ ) [14]. Ови индекси су дефинисани на

следећи начин:

$$E = E(G) = \sum_{i=1}^n |\lambda_i| \quad (5)$$

$$EE = EE(G) = \sum_{i=1}^n e^{\lambda_i} \quad (6)$$

$$ER = ER(G) = \sum_{i=1}^n \frac{1}{n - \lambda_i} \quad (7)$$

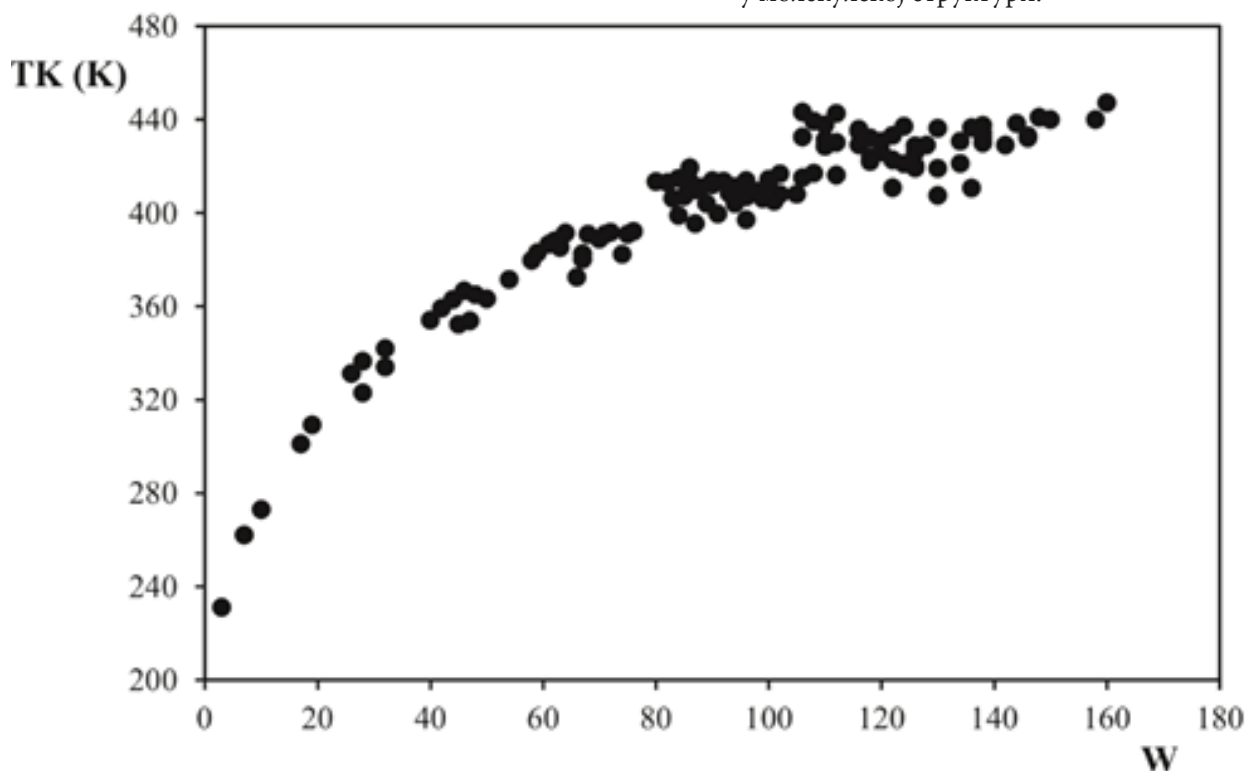
где  $\lambda_i$  представља  $i$ -ту сопствену вредност матрице суседства.

Енергија графа је дескриптор који је добио име по укупној  $\pi$ -електронској енергији коњугованих молекула израчунатој на нивоу Хикелове молекулско орбиталне теорије. Наиме, вредности овог дескриптора се подударају са вредностима укупне НМО  $\pi$ -електронске енергије за алтернативне угљоводонике у основном енергетском стању. Елегантност његове дефиниције се огледа у томе што се он може применити и на друге хемијски интересантне класе молекула. Естрадин индекс је уведен како би се моделовала увијања у биолошки важним макромолекулима. Резолвентна енергија је уведена пре кратког времена, али прелиминарни тестови указују на значајну предикциону моћ овог дескриптора.

## КРИТЕРИЈУМИ ЗА УВОЂЕЊЕ НОВИХ ТОПОЛОШКИХ ДЕСКРИПТОРА

Број тополошких дескриптора је огроман. Велики број је уведен као модификација већ познатих дескриптора. Због тога је неопходно применити неке критеријуме за филтрирање најбољих међу њима. Милан Рандић је у [5] формулисао ове критеријуме:

- могућност непосредне структурне интерпретације;
- корелација са макар једном физичко-хемијском особином;
- могућност разликовања изомера (дегенеративност молекулских дескриптора);
- могућност његовог израчунавања на деловима молекула (локална дефинисаност);
- могућност да се уведе низ аналогних дескриптора;
- линеарна независност;
- да није заснован на физичко-хемијским особинама;
- једноставност;
- да није на тривијалан начин повезан са другим дескрипторима;
- да се може ефикасно израчунати;
- да је дефинисан на основу лако разумљивих структурних концепата;
- да има коректну зависност од величине молекула;
- да се постепено мења са малим променама у молекулској структури.



Слика 3. Корелација Винеровог индекса ( $W$ ) са тачком кључања алкана ( $C_nH_{2n+2}$ ,  $n=3-10$ ).

Број критеријума које треба да испуни тополошки дескриптор је велики и због тога мали број дескриптора испуњава све ове услове. Међутим неколико критеријума горе наведених морају неизоставно бити испуњени да би се тополошка инваријанта могла сматрати молекулским дескриптором. Најважнији критеријум је да постоји добра корелација тополошког индекса са макар једном физичко-хемијском особином неке класе једињења. На слици 3 је приказана корелација између Винеровог индекса и тачке кључања алкана.

Важна особина дескриптора је и његова дегенеративност. Под дегенерацијом тополошког дескриптора подразумева се појава да дескриптор даје исту вредност за различите молекулске графове. „Добар“ тополошки молекулски дескриптор би требало да има што мању дегенеративност, односно што мањи број парова различитих молекула са истом вредношћу. Ако је  $M$  скуп молекулских графова, а  $I$  скуп свих вредности коју дескриптор поприма, онда се степен дегенерације ( $\delta$ ) може дефинисати као:

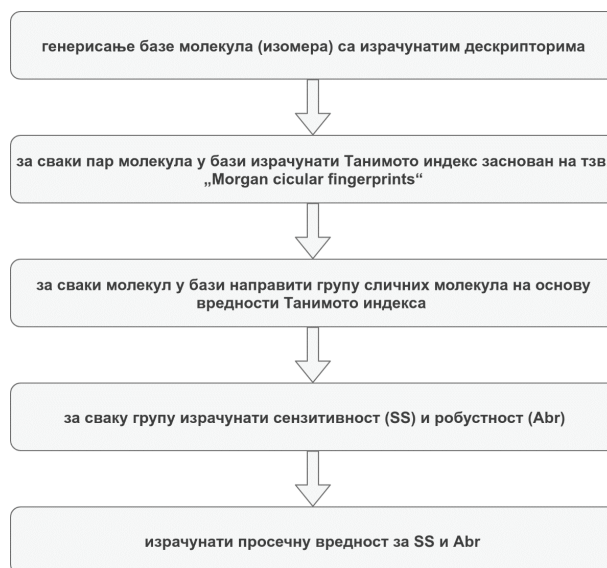
$$\delta = \frac{|M| - |I|}{|M| - 1} \quad (8)$$

где  $|M|$  и  $|I|$  представљају број елемената скупа  $M$  и  $I$ . Ако сви графови имају различите вредности, онда је  $\delta=0$ , а ако сви графови имају исту вредност за  $I$ , онда је  $\delta=1$ .

## СТРУКТУРНА ОСЕТЉИВОСТ ДЕСКРИПТОРА

Особина које би дескриптор требао да поседује односи се на постепену промену његове вредности са „постепеним“ променама у структури молекула. Идеја да се структурна осетљивост дескриптора квантификује датира још од самих почетака примене тополошких дескриптора у хемији. У раду [15] је представљен алгоритам за одређивање ове величине.

## АЛГОРИТАМ



Сензитивност и робустност се могу израчунати помоћу формула 9 и 10. У овим једначинама  $TI$ ,  $mr$  и  $M$  представљају тополошки индекс, референтни молекул из корака 3 алгоритма и број молекула у тој подгрупи. Молекулски дескриптор је осетљивији што има већу вредност за  $SS$  и вредност за  $Abr$  што ближе вредности за  $SS$  (која је по дефиницији увек већа од  $SS$ ).

$$SS(TI, mr) = \sqrt{\frac{1}{M} \sum_{m \in M} (TI_m - TI_{mr})^2} \quad (9)$$

$$Abr(TI, mr) = \max_{m \in M} |TI_m - TI_{mr}| \quad (10)$$

## ЗАКЉУЧАК

Квантификовањем информације садржане у дводимензионалној структури молекула омогућено је да се структура преведе у број или низ бројева којима је могуће даље манипулисати. Те величине су тополошки молекулски дескриптори. Постоји мноштво ових дескриптора и већина се може сврстати у једну од три групе. Употребом ових графовских инваријанти је могуће предвидети физичко-хемијске особине и активности непознатих једињења, чиме се олакшава њихова синтеза и даља испитивања.

## ABSTRACT

**Izudin Redžepović and Boris Furtula**, Faculty of Science, University of Kragujevac

### TOPOLOGICAL MOLECULAR DESCRIPTORS

Nature is diverse and perplexed. The extraction of the information from such an environment requires parameters by which the objects can be differentiated. In order to accomplish this task, the properties of an object need to be clearly defined. The same way of reasoning can be applied to chemistry. An idea that physico-chemical properties depend on molecular structure is a cornerstone of modern chemistry. Usually, it is difficult to identify and quantitatively determine this dependence. Topological molecular descriptors are of enormous help because they facilitate gathering the information encoded into two-dimensional molecular structure.

### РЕФЕРЕНЦЕ

- 1 J. L. Faulon, A. Bender, *Handbook of Chemoinformatics Algorithms*, CRC Press, Boca Raton, 2010.
- 2 R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley, New York, 2009.
- 3 A. Leach, V. Gillet, *An Introduction to Chemoinformatics*, Springer, Berlin, 2007.
- 4 I. Gutman, *Uvod u hemijsku teoriju grafova*, PMF Kragujevac, Kragujevac, 2003.
- 5 M. Randić, Generalized molecular descriptors, *J. Math. Chem.*, **7** (1991) 155–168.
- 6 N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, 1992.
- 7 I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total  $\pi$ -electron energy of alternant hydrocarbons, *Chem. Phys. Lett.*, **17** (1972) 535–538.
- 8 I. Gutman, B. Ruščić, N. Trinajstić, C. F. Wilcox, Graph theory and molecular orbitals. XII. Acyclic polyenes, *J. Chem. Phys.*, **62** (1975) 3399–3405.
- 9 M. Randić, On characterization of molecular branching, *J. Am. Chem. Soc.*, **97** (1975) 6609–6615.
- 10 H. Wiener, Structural determination of paraffin boiling points, *J. Amer. Chem. Soc.*, **69** (1947) 17–20.
- 11 D. Janežič, A. Miličević, S. Nikolić, N. Trinajstić, *Graph-Theoretical Matrices in Chemistry*, Univ. Kragujevac, Kragujevac, 2007.
- 12 X. Li, Y. Shi, I. Gutman, *Graph Energy*, Springer, New York, 2012.
- 13 E. Estrada, Characterization of 3D molecular structure, *Chem. Phys. Lett.*, **319** (2000) 713–718.
- 14 I. Gutman, B. Furtula, E. Zogić, E. Glogić, Resolvent energy of graphs, *MATCH Commun. Math. Comput. Chem.*, **75** (2016) 279–290.
- 15 M. Rakić, B. Furtula A novel method for measuring the structure sensitivity of molecular descriptors, *J. Chemometrics*, **33** (2019) #e3138.

