

A molekula-transzform, a molekulák hasonlóságának mérőszáma

NÁRAY-SZABÓ Gábor^{a,b,**}

^aMagyar Tudományos Akadémia Könyvtára, Arany János u. 1, 1051 Budapest, Magyarország

^bELTE TTK Kémiai Intézet, Pázmány Péter sétány 1A, 1117 Budapest, Magyarország

1. Bevezetés

A molekulák hasonlósága egyre nagyobb figyelmet kelt az elméleti kémiában, különösen a számítógépes molekulatervezésben.¹ A molekuláris hasonlóság kvantitatív meghatározására számos mérőszámot dolgoztak ki, ezek közül az egyik a molekulatranszform.⁸ A számítógépes molekula-tervezésben először Soltzberg és Wilkins javasolták az alkalmazását,² majd ezt többen is követték mind a mai napig.³⁻⁷ Sűrített formában tartalmazza a molekula térszerkezetére vonatkozó valamennyi információt, ugyanakkor – egyváltozós függvény lévén – matematikai kezelése lényegesen egyszerűbb, mint pl. az elektronsűrűségek átfedésére alapozott Carbó-féle hasonlósági index.⁸

Az alábbiakban egy korábbi publikációnk⁵ eredményeit felhasználva a molekulatranszform egyszerűsített – topológiai – definícióját alkalmazzuk a molekulák hasonlóságának kvantitatív jellemzésére, így a molekula szerkezeti képletéből kiindulva, gyorsan elvégezhetjük az összehasonlítást. Jelen közleményben azt vizsgáljuk, hogy a hasonlóknak talált molekulák mely tulajdonságai állnak közel egymáshoz, valamint azt, hogy a molekulatranszform alapján definiált molekuláris hasonlósági index mennyire tekinthető univerzális mérőszámnak. Vizsgálataink néhány fizikai állandóra (forráspont, olvadáspont, sűrűség, törésmutató), néhány különböző típusú farmakológiai aktivitásra, valamint a kristályszerkezetben kialakuló molekuláris illeszkedésre terjednek ki.

2. Definíciók

Ha röntgen- vagy elektronsugárzást bocsátunk egy anyagi rendszerre, ez szóródik az atomokon. A szórt sugárzás erősségének térbeli eloszlását a kísérleti úton (pl. elektrondiffrakcióval) is meghatározható $I(s)$ molekula-transzform írja le, mely a következőképpen számítható ki

$$I(s) = K \sum_{i < j} f_i f_j \int P_{ij}(r) \sin(sr) / sr \, dr \quad (1)$$

ahol K állandó, f_i és f_j a véges és gömbszerűnek feltételezett i -edik és j -edik atomon lejátszódó szórás irányfüggését jellemző formafaktorok, $P_{ij}(r)$ valószínűségi eloszlásfüggvény, mely leírja az i és j atom távolságának a molekula rezgése következtében fellépő változását. $s = 4\pi\lambda \sin(\theta/2)$, ahol λ az elhajlított elektronsugár hullámhossza, θ az elhajlás szöge. Ha $K=1$ és $P_{ij}(r) = \delta(r-r_{ij})$, vagyis az atommagok rögzítettek, a következő egyszerűsített képlethez jutunk ($x = sr_{ij}$)

$$I(s) = K \sum_{i < j} f_i f_j \sin x / x \quad (2)$$

A molekulatranszform topológiai alakját akkor kapjuk, ha r_{ij} a molekula szerkezeti képletében az i -edik és j -edik atomot elválasztó kötések legkisebb számát jelenti. Közvetlen kötésben lévő (pl. egy benzolgyűrűben egymáshoz képest *orto* helyzetű) atomok topológiai távolsága értelemszerűen 1, a *meta* és *para* pozícióban lévő atomok topológiai távolsága 2, illetve 3.

Az egyedi atomokra jellemző formafaktorok helyett a (2) képletben más mennyiség, pl. a magtöltésszám vagy egyéb származtatott atomi tulajdonság is szerepelhet, ami lehetőséget ad arra, hogy különböző típusú molekuláris tulajdonságok esetében különböző hasonlósági indexet használjunk. Valószínű például, hogy a molekula töltéeloszlását egyetlen számmal leíró dipólusnyomaték és a molekuláris hasonlóság összefüggése könnyebben felismerhető, ha f_i -t a atomi nettó töltéssel helyettesítjük. Az alábbiakban minden esetben f_i helyett Z_i -t, az i -edik atom magtöltésszámát használjuk a (2) egyenletben.

Az a és b molekula hasonlóságát a molekulatranszformok átfedéséből határozhatjuk meg

$$N_{ab}^2 = \int I_a(s) I_b(s) ds \quad (3)$$

Ebből a hasonlósági index Carbó nyomán⁸ a következő

$$R_{ab}^C = N_{ab}^2 / N_{aa} N_{bb} \quad (4)$$

Hodgkin és Richards⁹ az alábbi kifejezést javasolta

$$R_{ab}^H = 2N_{ab}^2 / (N_{aa}^2 + N_{bb}^2) \quad (5)$$

Egy korábbi munkánkban³ pedig egy másik képlet szerepel

$$(R_{ab}^G)^2 = \int [I_a(s) - I_b(s)]^2 ds \quad (6)$$

A (4), (5) és (6) képletekkel definiált hasonlósági indexek szoros korrelációban vannak egymással,⁵ könnyen belátható, hogy $(R_{ab}^G)^2 = 2(1 - R_{ab}^C)$.

Jelentősen megkönnyíti és felgyorsítja az számszerű összehasonlítást, hogy R_{ab}^C zárt alakban kiszámítható, miután

$$N_{ab}^2 = \sum_{i < j(a)} \sum_{k < l(b)} f_i^a f_j^a f_k^b f_l^b g_{ab}(D_{ij}^a, D_{kl}^b) \quad (7)$$

ahol

$$g_{ab}(D_{ij}^a, D_{kl}^b) = 1/2\pi [\max(D_{ij}^a, D_{kl}^b)]^{-1} \quad (8)$$

A fenti definíciók alapján két molekula annál jobban hasonlít

* Telefon: 311-4230, fax: 331-6954, drótposta: naraysza@chem.elte.hu

** Néhai Lukovits Istvánnak ajánlom ezt a munkát.

§ Az angol "molecular transform" magyar megfelelőjét a SZTAKI Szótárból (<http://szotar.sztaki.hu/index.hu.jhtml>) vettük át.

egymáshoz, minél közelebb van R_{ab}^C és R_{ab}^H az egyhez, illetve R_{ab}^G a nullához.

3. Eredmények és diszkusszió

A molekulák hasonlóságát akkor érdemes számszerűen kifejezni, ha feltételezhetjük, hogy hasonló molekulák fizikai és kémiai tulajdonságai is hasonlók egymáshoz. Az alábbiakban megmutatjuk, hogy a molekulatranszformok átfedése, mint hasonlósági kritérium alapján az a feltételezés bizonyos mértékben teljesül, azonban különböző tulajdonságok esetében különböző mértékben.

3.1. Fizikai állandók

A hasonlóságot azonos alapvázal rendelkező, különböző funkciós csoporttal szubsztituált molekulák halmazában vizsgáltuk. Minden molekulacsaládból kiválasztottunk egy vezérmolekulát, melyet 0-val jelölünk, és kiszámítottuk a család tagjainak hasonlósági indexét a (4) képlet szerint (R_{0a}^C) ehhez a vezérmolekulához viszonyítva. Azt vizsgáltuk, mennyire lehet sorba rendezni a többi molekula adott tulajdonságát R_{0a}^C függvényében, vagyis milyen szoros a korreláció a kétféle mennyiség között.

Szubsztituált alifás alkoholok esetében működik a legjobban a módszer. A forráspont, az olvadáspont, a sűrűség és a törésmutató esetében jól leírható e mennyiségek változása a metanolhoz viszonyított hasonlósági index függvényében, l. a (9) egyenletet és az 1. táblázatot.

$$P = aR_{0a}^C + b \quad (9)$$

1. Táblázat. Alifás alkoholok fizikai és biológiai tulajdonságainak függése a metanolhoz viszonyított hasonlósági indextől: a (9) regressziós egyenlet együtthatói (a és b), a korrelációs együttható (r) és az összehasonlított molekulák száma (n). IGC: a *Tetrahymena Pyriformis* növekedését 50 %-ban gátló koncentráció, LC: a *Pymphales pyriformis* számára halálos adag 50 %-a, PC: a *Madison 517* gombával szemben kifejtett mérgező hatás, LIP: a lipoxigenáz enzimre kifejtett gátló koncentráció, SHL: a birkamáj észteráz enzimre kifejtett gátló koncentráció. Számszerű adatokat l. korábbi munkánkban.⁵

Tulajdonság	a	b	r	n
forráspont	-1092	1063	-0,9473	29
olvadáspont	-625,1	610,7	-0,8446	16
sűrűség	-0,2373	0,2397	-0,8230	29
törésmutató	-0,4251	0,4664	-0,9458	29
IGC	-26,52	26,27	-0,9947	12
LC	-28,94	28,86	-0,9956	12
PC	-25,48	24,98	-0,9862	19
LIP	-24,06	23,82	-0,9555	11
SHL	-19,64	19,49	-0,9010	13

Nem meglepő, hogy az olvadáspont és a sűrűség esetében gyengébb a korreláció, ezeket a mennyiségeket ugyanis jelentős mértékben az illető anyag kristályszerkezete határozza meg, mely bonyolult módon függ az egyes molekulák szerkezetétől.

Szubsztituált fenolok esetében nincs korreláció az olvadáspont és törésmutató, valamint a hasonlósági index között (l. a 2. táblázatot). Lehetséges, hogy ez a molekulák között kristályos fázisban kialakuló, esetenként meglehetősen bonyolult hidrogénhidas szerkezeteknek tudható be, melyek hatása csak rejtett módon jelentkezik a fizikai tulajdonságok kialakításában.

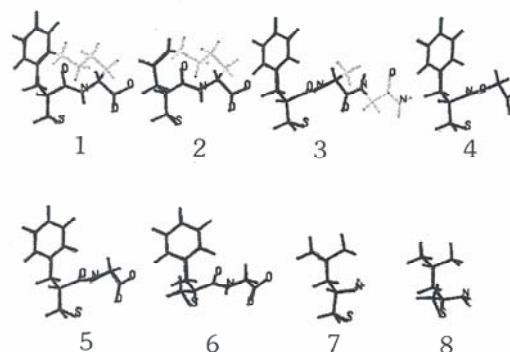
2. Táblázat. Szubsztituált fenolok fizikai és biológiai tulajdonságainak függése a fenolhoz viszonyított hasonlósági indextől: a (9) regressziós egyenlet együtthatói (a és b), a korrelációs együttható (r) és az összehasonlított molekulák száma (n). AC: akut toxicitás, HA: hemolitikus aktivitás, AA: baktériumölő hatás. Számszerű adatokat l. korábbi munkánkban.⁵

Tulajdonság	a	b	r	n
forráspont	337,1	-3,72	0,8120	20
olvadáspont	-199,1	59,15	-0,2485	22
törésmutató	0,0949	0,0179	0,2365	14
AC	-5,75	3,31	-0,8136	25
HA	-15,37	21,90	-0,7934	25
AA	-16,77	23,54	-0,8391	25

3.2. Biológiai aktivitás

A molekulák hasonlóságát jellemző mérőszámok használata elsősorban a kvantitatív hatás-szerkezet összefüggések felismerését célozza. Szubsztituált alkoholok esetében az 1. táblázat tanúsága szerint szoros korreláció figyelhető meg a metanolhoz viszonyított hasonlósági index és több, eltérő jellegű biológiai hatás között. Szubsztituált fenolok esetében gyengébb, de mégis felismerhető a korreláció (l. 2. táblázat), ezzel szemben szubsztituált acetilén-szulfon származékok gombaölő hatása nem állítható párhuzamba a hasonlósági indexszel.

Az 1. ábrán látható a termolizin enzim néhány, rokon szerkezetű gátló molekulája, melyekre Boháček és Martin végzett kvantitatív hatás-szerkezet vizsgálatokat.¹⁰ Az ábrán l-gyel



1. Ábra. A termolizin néhány inhibitora.¹⁰ A hipotetikus aktív fragmenst vastag vonallal jelöltük.

jelölt vezérmolekulához viszonyított hasonlósági index és az aktivitás (ΔpK_{ia}) között kapcsolatot teremtő, regressziós egyenlet állandóit a 3. táblázatban tüntettük fel.

3. Táblázat. Termolizin működését gátló molekulák hasonlósága és aktivitása közötti összefüggés. A (9) regressziós egyenlet együtthatói (a és b), a korrelációs együttható (r) és az összehasonlított molekulák száma (n).

modell	a	b	r	n
teljes molekula	5,68	1,75	0,6867	7
aktív fragmens	6,91	1,74	0,7993	7

A korreláció lényegesen javul, ha nem az egész molekulát, hanem csak a hatásért felelős részletét (aktív fragmens, az 1. ábrán vastagítva) vesszük figyelembe a hasonlósági index számításánál.

Utolsó példánk szubsztituált benzamidin származékok tripszinre kifejtett gátló hatására vonatkozik.¹¹ A molekulák körül kialakuló elektrosztatikus potenciál-mintázat alapján a szubsztituenseket két csoportra oszthatjuk. Az egyikben (3-Me, 3-OH, 3-OMe, 3-OEt, 4-Me, 4-NH₂, 4-OH, 4-OMe, 4-OEt) a szubsztituens körül pozitív potenciál alakul ki, míg a másikban (3-NO₂, 3-COMe, 3-COOMe, 3-CONHMe, 4-NO₂, 4-COMe, 4-COOMe, 4-COOEt, 4-CONHMe) a potenciál legalábbis részben negatív annak ellenére, hogy az egész molekula egységnyi pozitív töltést hordoz. Azt találtuk, hogy két kivétellel (3-OMe és 4-OEt) az I csoportba tartozó molekulák kötődési állandója, $pK_i > 4,2$, míg a II csoportba tartozó molekulák esetében egy kivétellel (3-COMe) $pK_i < 4,2$.

A (3) és (4) egyenletben definiált R_{ab}^C hasonlósági index segítségével alátámasztható a fenti felosztás. Ha ugyanis két molekula azonos csoportba tartozik, $R_{ab}^C < 0,5$ (az I csoportban 21 lehetséges párból 1, a II csoportban 66 lehetséges párból 10 kivételével), ha viszont különböző csoportba tartoznak, a 84 lehetséges párból 15 kivételével $R_{ab}^C > 0,5$. Két molekula hasonlósága tehát abban nyilvánul meg, hogy a szubsztituens körül kialakuló elektrosztatikus potenciál-mintázatuk hasonló-e vagy sem. Ha igen, hasonló a tripszin-gátló hatásuk, ha nem ez a hatás eltérő. A csoportosítás visszatükröződik a hasonlósági indexből is.

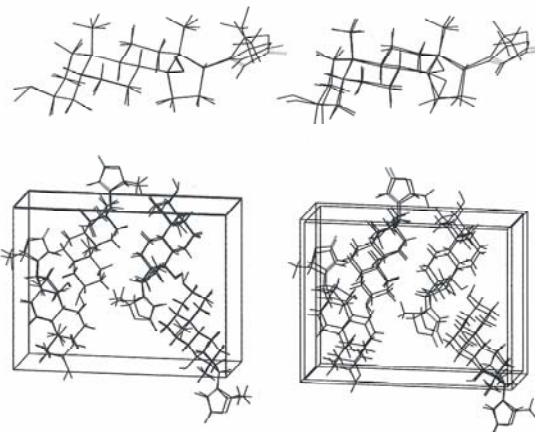
3.3. Illeszkedés kristályokban

A molekulák illeszkedésében fellelhető hasonló vonások különböző kristályos szerkezetekben jól nyomon követhetők a Kálmán és társai által javasolt *izostrukturális index* alapján.¹² Ha összevetjük az A és a B elemi cellát, az index az alábbi képlettel határozható meg

$$I_{AB}(n) = 100[|\sum(\Delta r_{AB}^i)^2/n|^{1/2} - 1] \quad (10)$$

Δr_{AB}^i a kristálytani aszimmetrikus egység megfelelő molekuláihoz tartozó i -edik, A és B nem-hidrogén atom alkotta pár frakcionált koordinátákban mért távolságát jelenti. A az egyik, B a másik elemi cellában helyezkedik el. Az összegezés egy tetszés szerint definiált molekularészlet atompárjaira vagy akár az elemi cellában lévő összes atompárra terjedhet ki, n az összehasonlított atompárok száma. Akkor nagyobb az elemi cellák hasonlósága, ha kisebb I_{AB} .

Néhány kardiotóniás hatású szteroidmolekula esetében kimutattuk,¹³ hogy az izostrukturális index és a molekulatranszformmal definiált hasonlósági index



2. Ábra. Kardiotóniás szteroid származékok szuperpozíciója. Fent: izolált molekulák, lent: elemi cellák a kristályban. Balra: 21R- és 21S-metildigitoxigenin ($R_{ab}^G=5,37$; $I_{AB}=83,9$), jobbra digirezigenin és 21S-metildigitoxigenin ($R_{ab}^G=20,4$; $I_{AB}=62,5$).

viszonylag jó korrelációban van egymással, minél közelebb van R_{ab}^G a nullához, annál nagyobb I_{AB} :

$$I_{AB} = -1,265R_{ab}^G + 88,9 \quad (11)$$

A korrelációs együttható, $r=0,7935$, az összehasonlított molekulák száma 10.

A fentiek alapján megállapítható, hogy (legalábbis a vizsgált vegyületek körében) hasonló molekulák hasonló elemi cellákat alakítanak ki a kristályban. Ez jól követhető a 2. ábrán, ahol két hasonló szteroid-pár molekuláját és kristályos fázisban kialakuló elemi celláját hasonlíthatjuk össze. A 21R- és 21S-metildigitoxigenin molekulája meglehetősen hasonló egymáshoz, ezzel párhuzamosan az elemi cellák is hasonlóak. A digirezigenin és 21S-metildigitoxigenin alkotta pár esetében a molekulák kevésbé hasonlítanak egymásra, ezzel párhuzamosan az elemi cellák is szemmel láthatóan jobban eltérnek egymástól.

4. Következtetések

A molekulatranszform alkalmas arra, hogy kvantitatív formában kifejezze a molekulák hasonlóságát. Különböző tulajdonságok és különböző molekulatípusok esetében eltérő sikerrel alkalmazható. Azokban az esetekben hasznos, melyekben más módszer nem áll rendelkezésre a hasonlóság kifejezésére. Ilyen például a molekulák illeszkedése a kristályos fázisban, melynek függése az egyedül álló molekula szerkezetétől más módszerrel nehezen követhető (l. a 2. ábrát).

Hivatkozások

1. Carbó-Dorca, R.; Mezey, P.G. *Fundamentals of Molecular Similarity*, Springer: Berlin, 2001.
2. Soltzberg, L.J.; Wilkins, C.L. *J. Am. Chem. Soc.* **1977**, *99*, 439-443.
3. Gabányi, Z.; Surján, P.R.; Náray-Szabó, G. *Eur. Med. Chem.* **1982**, *17*, 307-311.
4. King, J.W.; Kassel, R.J. *Int. J. Quant. Chem. Quant. Biol. Symp.* **1991**, *18*, 289-297.

5. Csorvássy, I.; Tózsér, L.; Kárpáti, L.; Náray-Szabó, G. *J. Math. Chem.* **1993**, *13*, 343-357.
6. Schuur, J.; Selzer, P.; Gasteiger, J. *J. Chem. Inf. Model.* **1996**, *36*, 334-344.
7. Molnar, S.P.; King, J.W. *Int. J. Quant. Chem.* **2001**, *85*, 662-675.
8. Carbó, R.L.; Arnau, M. *Int. J. Quant. Chem.*, **1980**, *17*, 1185-1189.
9. Hodgkin, E.E.; Richards, W.G. *Int. J. Quant. Chem. Quant. Biol. Symp.* **1987**, *14*, 105-110.
10. Bohacek, R.S.; McMartin, C. *J. Am. Chem. Soc.* **1994**, *116*, 5560-5571.
11. Náray-Szabó, G. *J. Mol. Struct. THEOCHEM* **1986**, *134*, 401-409.
12. Kálmán, A.; Párkányi, L.; Argay, G. *Acta Cryst.* **1993**, *B49*, 1039-1049.
13. Náray-Szabó, G.; Harmat, V. *MATCH Comm. Math. Comp. Chem.* **1997**, *35*, 29-40.

The molecular transform as a similarity measure

We call the attention to the experimentally measurable molecular transform as a basis for the quantitative description of molecular similarity. Simplified topological forms can be derived, which allows to calculate various metric distances between molecular transforms very rapidly. Such distances can be used for defining

various groups of molecules possessing similar physical, chemical or pharmacological properties. One of the interesting applications is crystal isostructurality, where unit cells and asymmetric units can be compared quantitatively and some general rules for crystal packing can be derived.