

Dr. Pósa Mihály\*

## A Shannon-féle entrópia mint molekula deszkriptor a szteroid vázas vegyületeknél

### 1. Bevezetés

A Vajdasági Magyar Tudományos Társaság 2010. évi tanácskozásán Kiss Ernő professzor ismertette a termodinamika törvényeit (Hány főtétele van a termodinamikának?). Előadásában megemlíttette a termodinamika II. törvényét, amely a spontán folyamatok létrejöttét magyarázza az információs elmélet alapján. Ezek szerint, valamely rendszer állapota annál valószínűbb minél nagyobb annak az információs bizonytalansága (minél több mikro-rendszerrel jellemezhető), azaz rendezetlensége, amelyet az entrópia-állapot függvény fejez ki.

A szerves vegyületek nagyfokú szerkezeti diverzitása lehetővé teszi szerkezetük jellemzését a Shannon-féle entrópiával mint molekula deszkriptorral.

Az epesavak szteroid vázas amfifil molekulák, amelyek tulajdonosságát nagymértékben meghatározza a  $5\beta$ -kolánsav molekula geometriája. A  $5\beta$ -kolánsav molekula szteroid vázán megkülönböztethető egy domború ( $\beta$ ) és egy homorú ( $\alpha$ ) felület. Ezek a felületek az epesav származékoknál a hidrofóbia szemszögéből különböznek. A legtöbb topológiai molekula deszkriptor nem tükrözi az epesavak szteroid vázáinak kettségét. Sőt bizonyos deszkriptorok (*Winer-index*, *Balaban-index*, *shape attribute*, stb.) az epesavak szerkezeti izomérjeit sem képesek megkülönböztetni egymástól (pl. a kólsav mono-oxoszármazékainál az előbb említett molekula deszkriptorok ugyanazt a számot adják, függetlenül az oxo csoportnak a szteroid vázban elfoglalt helyétől) (1-3).

Sokszor előfordul, hogy a QSA(P)R kutatásra szánt tanítóhalmaz molekula száma (elemszám) kicsi (tanítóhalmaz: 6-12 molekula, ellenőrzőhalmaz: 2-4 molekula), így a többváltozós regressziós módszerek általában túlbecsült modellt adnak. Ezért célszerű olyan molekula desz-

---

\* Dr. Pósa Mihály, egyetemi docens, Újvidéki Egyetem, Orvostudományi Kar, Gyógyszerészeti Tanszék, A Fizikai kémiai kutató csoport vezetője, Újvidék

kriptorok szerkesztése-fejlesztése, amelyek minél jobban jellemzik (kódolják) a vizsgált molekula szerkezetét, aminek köszönhetően a QSA(P)R függvény akár egyváltozós lineáris egyenlet is lehet, nagy előjelző képességgel a vizsgált kongenerikus molekulák csoportjára tekintve (1,4). E munka célja, hogy rámutasson a *Shannon* entrópia alkalmazásának lehetőségére új epesav molekula deszkriptor(ok) kifejlesztésében.

## 2. A molekula deszkriptorok

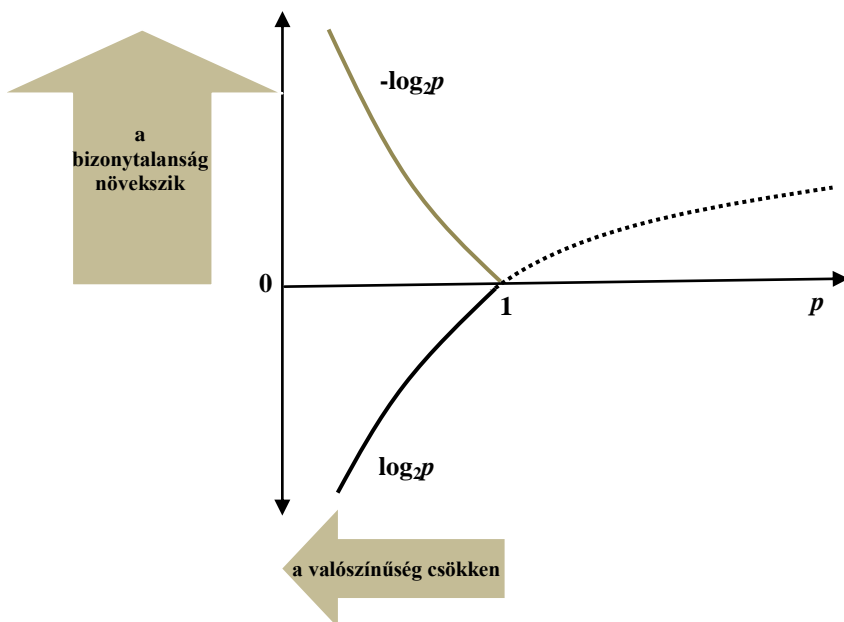
A molekula deszkriptorok a molekulák szerkezetének matematikai megjelenítése, ideális esetben, minden különböző molekula egy meghatározott számmal jellemezhető. A molekula deszkriptorok lehetnek:

- topológiai deszkriptorok, amikor a molekulákat egy 2D-es gráffal ábrázoljuk, ami alapján távolsági, szomszédsági, kötésrendi stb. mátrixokat határozzunk meg,
- kvantumkémiai deszkriptorok, amelyek a molekula elektron eloszlása alapján számított elektronsűrűségek (atomokra vetítve), energia értékek stb.,
- fizikai-kémiai jellemzők adatbázisa (fragmentációs módszerrel kiterjeszhető az ismeretlen vegyületekre is).

A hatás szerkezet (QSAR) vizsgálatoknál a molekula deszkriptorok a független változók, míg a biológiai-farmakológiai tulajdonságok (kísérleti eredmények) a függő változók. Cél megkeresni a vizsgált vegyületeknél a biológiai-farmakológiai tulajdonságuk és a megfelelő molekula deszkriptorok (szerkezetük) közötti függvény-kapcsolatot (4).

## 2. A Shannon-féle entrópia

Valamely diszkrét valószínűségi változó  $X$  a következő (különböző) értékeket veheti fel  $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ . Ha  $m$  számú kísérletben az  $X$  valószínűségi változó valamely  $i$  értéke ( $x_i$ )  $m_i$ -szer jelenik meg, akkor annak a valószínűsége, hogy  $X$  értéke  $x_i$ :  $p_i = m_i/m$ . Minél kisebb a  $p_i$  valószínűség, annál nagyobb annak a bizonytalansága (*uncertainty*), hogy valamely kísérletben  $X$  értéke  $x_i$  legyen. A bizonytalanság kifejezésére legalkalmasabb egy olyan függvény, amely folyamatosan csökken a valószínűség növekedésével: *bizonytalanság* =  $-\log_2 p$  (1. ábra).



1. ábra  
A bizonytalanság és a valószínűség

Az  $X$  valószínűségi változó mindegyik  $x_i$  értékéhez hozzárendelhető a megfelelő bizonytalanság, így valójában az  $X$  valószínűségi változót leképeztük egy új valószínűségi változóba a  $X$  határozatlanságába ( $\hat{X}$ ):

$$\begin{array}{rcl}
 X & = & \{ x_1, x_2, \dots, x_i, \dots, x_n \} \\
 \downarrow & & \downarrow \quad \downarrow \quad \dots \quad \downarrow \quad \dots \quad \downarrow \\
 \hat{X} & = & \{ -\log_2 p_1, -\log_2 p_2, \dots, -\log_2 p_i, \dots, -\log_2 p_n \} \\
 & & \downarrow \quad \downarrow \quad \dots \quad \downarrow \quad \dots \quad \downarrow \\
 & & p_1 \quad p_2 \quad \dots \quad p_i \quad \dots \quad p_m
 \end{array}$$

Mivel a  $\hat{X}$ -nek minden  $-\log_2 p_i$  értéke  $p_i$  valószínűséggel realizálódik, így átlagértéke:

$$\langle \hat{X} \rangle = -\sum_{i=1}^n p_i \log_2 p_i \quad (1).$$

Shannon szerint az  $X$  valószínűségi változó határozatlansága – entrópiája  $H(X)$  a diszkrét értékeinek a határozatlanságainak az átlagértéke:

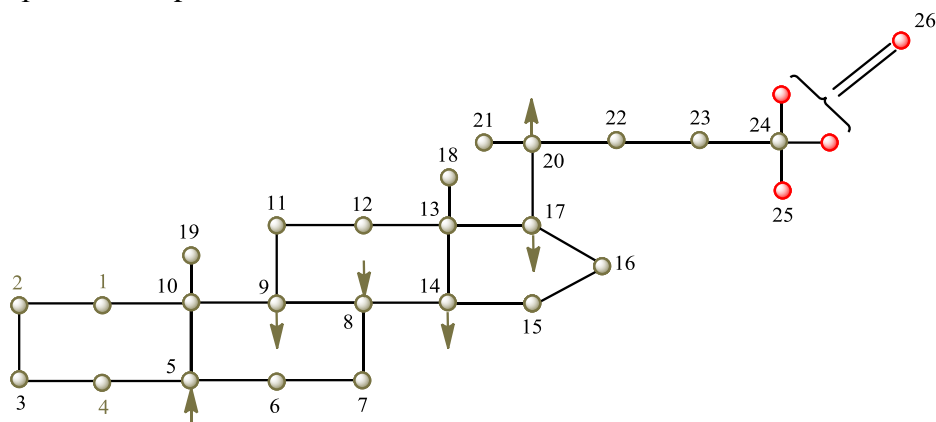
$$H(X) = \langle X \rangle \quad (3).$$

A  $H(X)$  akkor nulla, ha  $X$  lehetséges értékeinek valószínűsége mind nulla, kivéve egyet, amelyre akkor  $p=1$  ( $\log_2 1=0$ ). A  $H(X)$  maximális értékét akkor veszi fel, ha  $X$  mindegyik értéke azonos valószínűséggel  $p_i$  realizálódik ( $1=np_i \rightarrow p_i=1/n$ ) (5):

$$H(X)_{\max} = -\sum_{i=1}^n p_i \log_2 p_i = -n \left( \frac{1}{n} \log_2 \frac{1}{n} \right) = \log_2 n \quad (4).$$

### 3. Shannon-féle entrópia alapú molekula deskriptor konstrukciója az epesavaknál (szteroid vázas vegyületek)

Az epesav molekulákat 2D-es hidrogén atomok nélküli gráfnak tekintjük (2. ábra), amelyben a különböző elemek atomjai különböző típusú csomópontokat alkotnak.



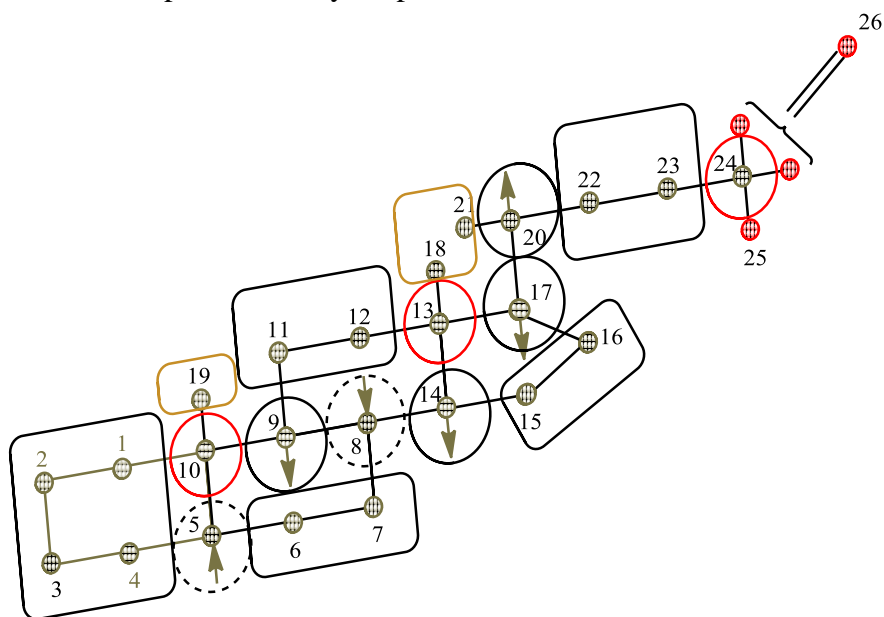
2. ábra  
A 5 $\beta$ -kolánsav molekula gráfja

A szteroid gyűrűrendszer sztereokémiai jellegzetességei oly módon konzerválódtak a 2D-es gráfban, hogy az anguláris (és a C21-es

láncbéli) terciér C atomok és a megfelelő H atom közötti kötést irányított útnak tekintjük. Ha a megfelelő H atom a gyűrűrendszer  $\beta$  oldalán helyezkedik el, akkor a molekula gráfban a terciér C atomoknál az út iránya a csomópont felé mutat. Az  $\alpha$  orientációnál pedig az út kezdőpontja a csomópontban van (2. ábra).

Az első döntéshozatalnál ( $D1$ ) megállapítjuk a különböző típusú csomópontokat (atomokat) az epesav molekula gráfban (megkapjuk a csomópontok csoportjait). Ezután meghatározzuk annak a valószínűségét, hogy egy véletlenül választott csomópont valamely csoportba tartozzon.

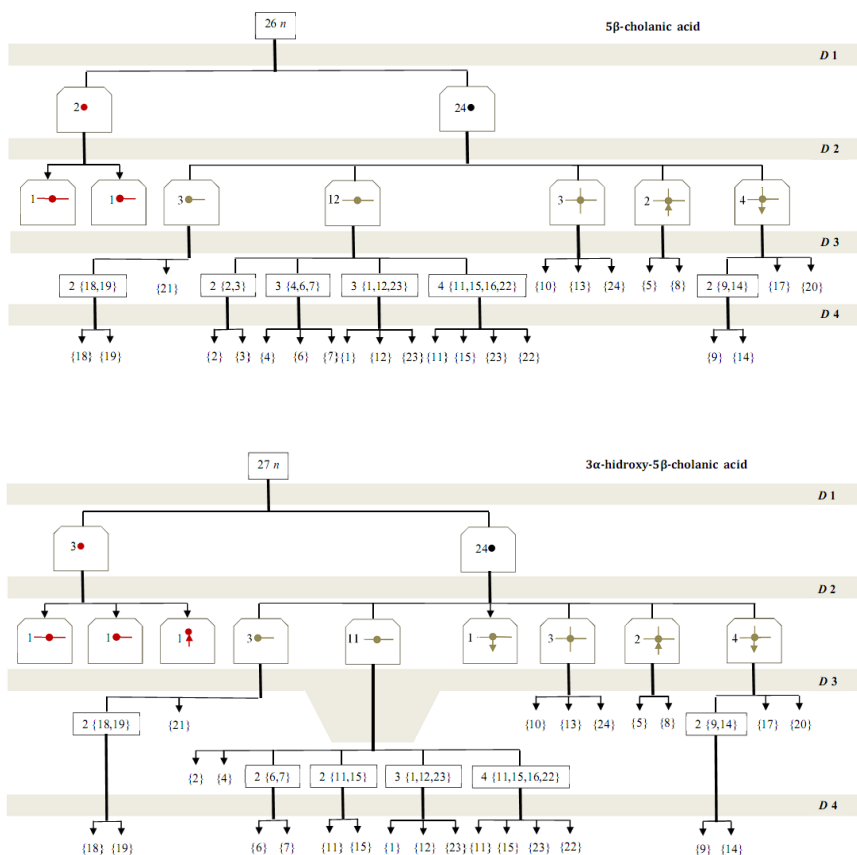
A második döntéshozatalnál ( $D2$ ) a  $D1$ -nél kapott csoportok további osztályozása (a C atom és O atomok csoportjainak további osztályozása) történik. Mindegyik csomóponthoz hozzárendeljük a csomópontból kiinduló utakat, így az azonos típusú csomópontokon belül különböző csomópont-út rendszerek jönnek létre (3. ábra). Hasonlóan mint a  $D1$ -nél, meghatározzuk annak a valószínűségét, hogy egy véletlenül választott csomópont valamely csoportba tartozzon.



3. ábra  
A 5 $\beta$ -kolánsav C atomok csoportjai a  $D2$ -ben

A harmadik döntési lépés azokat a csoportokat öleli fel, amelyeknél a  $D_2$ -ben egynél több elem található. Mindegyik ilyen  $D_2$ -es csoport csomópont-út rendszer szabad útjára ráillesztjük (hozzá csatoljuk) a szomszédos csomópont-út rendszert, majd meghatározzuk az azonos szubgráfok csoportjait. Szintén meghatározzuk annak a valószínűségét, hogy egy véletlen (mag)csomópont valamely szubgráf csoportjába tartozzon.

A további döntéshozási lépések addig folytatódnak, amíg az összes (mag)csomópont nem kerül külön csoportba (4. ábra).



4. ábra  
Az  $5\beta$ -kolánsav és a  $3\alpha$ -hidroxi- $5\beta$ -kolánsav döntési spektrumai

Az epesavak döntési spektrumaiban mindegyik döntési lépésben meghatározzuk a felbontás határozatlanságát (entrópiáját), majd az entropiákat összegezzük a döntési szintek szerint:

$$H_{cum} = -\sum_{\forall D} \sum_{i=1}^g p_i \log_2 p_i.$$

Néhány epesav  $H_{cum}$  értéke: 5 $\beta$ -kolánsav (1.96), 3 $\alpha$ -hidroxi-5 $\beta$ -kolánsav (2.08), 3-oxo-5 $\beta$ -kolánsav (2.15).

#### 4. Összegzés

A munkában bemutatottuk a *Shannon*-féle entrópia alkalmazásának lehetőségét az epesavak döntési spektrumának számszerűsítésében. Az így kapott szám, a  $H_{cum}$  felhasználható molekula deskriptorként az epesavak QSAR vizsgálatainál.

#### Felhasznált irodalom:

1. M. Poša: QSPR study of the effect of steroidal hydroxy and oxo substituents on the critical micellar concentrations of bile acids, *Steroids* 76(1-2) (2011) 85-93.
2. M. Poša: Fizičko-hemijske osobine žučnih kiselina sa osvrtom na okso derivate 5 $\beta$ -holanske kiseline, Medicinski fakultet, Novi Sad, 2011.
3. M. Poša, A. Pilipović, M. Lalić, J. Popović: Determination and importance of temperature dependence of retention coefficient (RPHPLC) in QSAR model of nitrazepam's partition coefficient in bile acid micelles, *Talanta* 83(5) (2011) 1634-1642
4. M. Poša: Osnovne metode u hemometriji, Medicinski fakultet, Novi Sad, 2010
5. Z.I Ivkovic: Teorija verovatnoće sa matematičkom statistikom, Gradjevinska Knjiga, Beograd 1976.