



Budapesti Műszaki és Gazdaságtudományi Egyetem

**Farkas Orsolya**

**Mennyiségi szerkezet-hatás összefüggések retenciós indexek és  
biológiai aktivitás előrejelzésére**

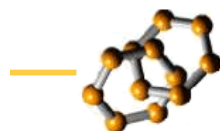
Ph.D. Tézisek

2007.

Témavezető:

**Dr. Héberger Károly**

tudományos tanácsadó



**Kémiai Kutatóközpont**

Magyar Tudományos Akadémia

*CENTER OF EXCELLENCE*

## **1. Bevezetés**

Az utóbbi évtizedekben a különféle mérőberendezések számának növekedésével, és felbontóképességük javulásával, rohamos ütemben növekszik a keletkező adatok mennyisége is. Ezek feldolgozása, értékelése hagyományos egyváltozós statisztikai módszerekkel gyakran nem oldható meg, folyamatosan nő az igény a korszerű, többváltozós statisztikai (kémiai adatok feldolgozásakor kemometriai) módszerek iránt.

Ezek közé a módszerek közé tartoznak a mennyiségi szerkezet-hatás összefüggések (Quantitative structure activity relationships, QSARs) is. A QSAR matematikai kapcsolatot ír le a molekulák szerkezeti jellemzői és valamilyen aktivitás adatuk (pl. gyógyhatás, receptorhoz való kötődés mértéke, stb.) között. A mennyiségi függvénykapcsolat segítségével könnyebben érthető, hogy milyen tulajdonságok felelősek egy aktivitás létéért vagy nemléteért, ill. mértékéért. A modell alapján költséges kísérletek nélkül új vegyületek különféle aktivitás értékeit számszerűen előre lehet jelezni. A vegyületek értékelése, valamint kívánt tulajdonságokkal rendelkező új vegyületek tervezése is lehetségessé válik. Az osztályozási módszerek hasonló célt szolgálnak, segítségükkel a molekulák hatásuk alapján (pl. toxikus vagy nem toxikus) csoportosíthatók számított vagy könnyen mérhető tulajdonságai felhasználásával.

Munkám során különböző kémiai ill. biológiai problémák megoldását végeztem el kemometriai módszerek segítségével. A QSAR modellezés részlépéseinek gyakorlati megvalósítása nem teljesen tisztázott és nincs rá egységes gyakorlat. Doktori témám egyik része a QSAR modellezés egyik fontos részlépésének, a modellbe beépítendő független változók kiválasztásának helyes módjával és az egyes módszerek összehasonlításával foglalkozott. A modellvegyületeim alkoholok, kén-, nitrogén- és oxigéntartalmú

heterociklusos molekulák, valamint zsírsav metil-észterek voltak. Modelljeimben a meghatározni kívánt függő változó a Kováts-féle retenciós index (RI) volt.

A változókiválasztás problémák tanulmányozásán kívül az RI előrejelzés más szempontból is hasznos. Segítséget nyújt olyan vegyületek szerkezetének azonosításánál, amelyek tömegspektruma nagyon hasonló (pl. azonos molekulatömegű izomerek esetében). A QSRR jelentőségét igazolják korábbi összefoglalók [Kaliszan R., *Quantitative Structure-Chromatographic Retention Relationships*, Wiley, New York, **1987**; Kaliszan R., *Structure and Retention in Chromatography - A Chemometric Approach*, Harwood, Amsterdam, **1997**] aktualitását pedig az mutatja, hogy egyidejűleg (2007-ben) a *Chemical Reviews*-ban és a *Journal of Chromatography A*-ban összefoglaló cikkek jelentek meg ebben a témában [Kaliszan R. QSRR: Quantitative Structure-Chromatographic Retention Relationships *Chem. Rev.* **2007** 107 (7) 3212-3246, és Héberger K. Quantitative structure-(chromatographic) retention relationships *J. Chromatogr. A* **2007** 1158 (1-2) 273-305.]

Munkám második felében antidepresszáns hatású gyógyszerjelölt molekulák osztályozását végeztem el a hERG kálium csatornához való kötődés alapján. Ez a tulajdonság egy nem kívánt mellékhatás: a kálium csatornához való kötődés befolyásolhatja a szív működést, és hirtelen halálhoz vezethet, gyógyszerjelölt vegyületek esetén vizsgálata újabbán elengedhetetlen. A már ismert hERG aktivitású antidepresszánsok felhasználásával épített mennyiségi szerkezet-hatás modellek segítségével lehetőség nyílik olyan vegyületek tervezésére, melyek hERG gátló aktivitása elhanyagolható.

Dolgozatom harmadik részében flavonoidok antioxidáns aktivitását előrejelző QSAR modelleket mutatok be. A természetben több mint 4000 flavonoid található. Antioxidáns sajátágaikat okozó szerkezeti hatásokról sok irodalom áll rendelkezésre, ám a biológiai aktivitás számszerű mennyiségi előrejelzésére alig található modell.

## **2. Vizsgált vegyületek és számítási módszerek**

A vizsgált vegyületek Kováts indexei (RI) az irodalomból, illetve kutatópartnerünk, I. G. Zenkevich professzor méréseiből származtak. A flavonoidok antioxidáns aktivitás értékeit szintén a szakirodalomból vettem. Az antidepresszánsok hERG aktivitás értékét az ESTEVE spanyol gyógyszervállalat bocsátotta rendelkezésre.

Ha szükséges volt, a molekulák szerkezetének optimalását a HyperChem [HyperChem 7.0 program package HyperCube Inc., Canada] programmal hajtottam végre. Az optimalás a szemi-empirikus AM1 módszerrel történt. A molekulák szerkezeti jellemzőit, az ún. deskriptorokat a Dragon program segítségével számítottam [Todeschini R., Consonni V., Pavan, M. Dragon Software Version 2.1, **2002.**].

A kemometriai módszerek közül a páronkénti korrelációs (PCM), az előreirányuló változóbevonás (FS), a peremregresszió (RR), a Lasso, a legjobb alrendszer regresszió (BSS), a többszörös lineáris regresszió (MLR), a genetikus algoritmus (GA), a parciális legkisebb négyzetek módszere (PLS), a lineáris diszkriminancia analízis (LDA), valamint az osztályozási és döntési fák módszerét (CART) használtam. Ezeket a módszereket vagy változó kiválasztásra, vagy/és a modellek paramétereinek becslésére, továbbá a két legutolsót osztályozásra használtam. A számítások döntő részét a Statistica [Statistica 5.5 ill. 6 Software Package, StatSoft Inc., Tulsa, OK, USA] program felhasználásával végeztem. A PCM számításokat egy Microsoft Excel makróval [Rajkó R.] végeztem, a genetikus algoritmus számításokat pedig a MobyDigs [Todeschini R., Ballabio D., Consonni V., Mauri A., Pavan M. MobyDigs Professional version 1.0 **2004** Milano Chemometrics and QSAR Research Group] programmal hajtottam végre. Néhány számítást (Lasso, egy esetben PLS komponensek regressziós koefficienseinek számítása) nem saját kezűleg végeztem el, ezeket a dolgozatban pontosan feltüntettem.

### **3. Új tudományos eredmények, alkalmazási lehetőségek**

1. Megbízható modelleket hoztam létre alkoholok Kováts-indexének előrejelzésére.

A modellek alakkal és mérettel összefüggő WHIM indexeket tartalmaznak. Megállapítottam, hogy az alkoholok retenciós indexének előrejelzésekor a legjobb módszer a független változók kiválasztására a legjobb alrendszer módszer (BSS) és az előreirányuló változóbevonás (FS) volt.

A PCM modell az alkoholok retenciójának leírásra alkalmas, viszont előrejelzésére nem alkalmazható.

Az általam vizsgált vegyületek esetében - az irodalomtól eltérően – a peremregresszió és PLS módszer nem vezetett elfogadható eredményre a modellépítés során.

2. Zsírsvav metil-észterek retenciós index előrejelzésére megbízható modelleket hoztam létre.

A zsírsvav metil-észterek példáján igazoltam, hogy az ún. minimálev alkalmazásakor a Lasso és a PLS módszer nem vezetnek jó eredményre.

Az MLR modellek az optimális modellezésnél is előnyösebbek olyan szempontból, hogy kevesebb változót tartalmaznak. Bemutattam, hogy a fenti vegyületcsoport esetében az MLR modellek nem működnek jól olyan változókkal, amiket PLS vagy Lasso módszer választott ki.

Bizonyítottam, hogy a PCM kiválóan alkalmazható változó kiválasztásra metil-észterek RI előrejelzésekor.

Megállapítottam, hogy ha a minimálevet nem alkalmazzuk, a PLS módszer a legalkalmasabb az észterek retenciós indexének előrejelzésére.

Összesen 37, eddig ismeretlen retenciós indexű zsírsav metil-észter Kováts-indexét számítottam ki, mely jelentősen megkönnyíti ezeknek a vegyületeknek a szerkezeti azonosítását.

Megállapítottam, hogy a metil-észterek retenciós mechanizmusának leírására alkalmasak a topológiai és egyéb terjedelmet jellemző deskriptorok, valamint a flexibilitást jellemző sajátságok.

3. Megbízható modelleket építettem heterociklusos vegyületek retenciós indexének előrejelzésére.

Megállapítottam, hogy a forráspont értékek felhasználása mindenképpen szükséges, de önmagában nem elégséges e vegyületek retenciójának leírásához. A háromdimenziós független változók, mint pl. a GETAWAY és WHIM deskriptorok hasznosak az előrejelzésben, de az előrejelzés kétdimenziós deskriptorokkal is megvalósítható.

A PLS módszer nem volt alkalmas a változókiválasztásra ebben az esetben. A BSS módszerrel kombinálva azonban a legjobb előrejelző képességű modellt adta.

4. Sikerült hERG gátló és nem gátló antidepresszánsok osztályozására alkalmas modelleket létrehoznom. A modellek szerkezetileg igen változatos vegyületek esetében is jó (80% feletti), szerkezetileg egy családba tartozó vegyületek között pedig kiváló (90% feletti) elkülönítést tettek lehetővé.

A legjobb módszerek az antidepresszánsok csoportosítására az előreirányuló változóbevonás illetve a genetikus algoritmus lineáris diszkriminancia elemzéssel összekapcsolt változatai voltak.

Megmutattam, hogy CART módszer használatával egyszerű és könnyen értelmezhető modell kapható. Az összes antidepresszáns csoportosításához két független változó

szükséges és elégséges, a karbonilcsoportok száma a molekulában és egy GETAWAY deszkriptor.

A legnagyobb alcsoport osztályozásához az oxigén-oxigén kötésekben levő atomok közti geometriai távolságok összege elégséges.

5. Modellt építettem igen különböző szerkezetű flavonoidok antioxidáns aktivitásának előrejelzésére. A PLS modell egyszerűen, változókiválasztás nélkül létrehozható, és az előrejelzésen kívül alkalmas lehet különféle flavonoidcsoportok osztályozására is.

## 4. Közlemények

### 4.1. A tézisek alapjául szolgáló közlemények:

1. **Farkas O.**, Héberger K., Zenkevich I. G. Quantitative structure – retention relationships XIV. Prediction of gas chromatographic retention indices for saturated O-, N- and S-heterocyclic compounds *Chemometrics and Intelligent Laboratory Systems* **2004**, 72, 173-184. IF: 2,45

2. **Farkas O.**, Jakus J., Héberger K. Quantitative structure-antioxidant activity relationships of flavonoid compounds *Molecules* **2004**, 9, 1079-1088. IF: 0,84

3. **Farkas O.**, Héberger K. Comparison of ridge regression, partial least squares, pair-wise correlation, forward- and best subset selection methods for prediction of retention indices of aliphatic alcohols *J. Chem. Inf. Model.* **2005**, 45, 339-346. IF: 3,42

4. **Farkas O.**, Zenkevich I.G., Stout F., Kalivas J.F., Héberger K. . Prediction of Retention Indices for Fatty Acid Methyl Esters *J. Chromatogr. A* (bírálat alatt)

5. **Farkas O.**, Héberger K. Chemometric analysis of antidepressants based on their hERG K<sup>+</sup> channel activity  
(közlésre előkészítve)



#### 4.2. Egyéb közlemények:

1. **Farkas O.**, Gere-Paszti E., Forgacs E. Study of the interaction of structurally similar bioactive compounds by thin-layer chromatography *Journal of Chromatographic Science* **2003**, 41 (4) 169-172. IF: 0,88

2. Gere-Paszti E., **Farkas O.**, Prodan M., Forgacs E. Molecular mapping of interaction between cholesterol and model drugs traced by reversed phase bioaffinity chromatography *Chromatographia* **2003**, 57 (9/10) 599-604. IF: 1,17

3. Prodan M., Gere-Paszti E., **Farkas O.**, Forgacs E. Validation and simultaneous determination of caffeine and paracetamol in pharmaceutical preparations *Chemia Analytyczna - Chemical Analysis* 2003, 48 (6) 901-907. IF: 0,56

4. [Forgacs E.](#), [Cserhati T.](#), **Farkas O.**, [Eckhardt A.](#), [Miksik I.](#), [Deyl Z.](#) Interaction between cholesterol and non-ionic surfactants studied by thin-layer chromatography *Journal of liquid chromatography & related technologies* **2004**, 27 (13) 1981-1992.  
IF: 0,88

5. Jakus J., **Farkas O.** Photosensitizers and antioxidants: a way to new drugs? *Photochem. Photobiol. Sci.* **2005**, 4, 694-968. IF: 2,42

### 4.3. Előadások, poszterek:

1. **Farkas O.**, Héberger K. Comparison of variable selection methods: Prediction of retention indices of saturated alcohols *Hungarian Chemometric Workshop Kemometria Tata*, 2002. szeptember 29-október 1. (poszter)
2. **Farkas O.**, Héberger K., Zenkevich I. G. Quantitative structure – retention relationships XIV. Prediction of gas chromatographic retention indices for cyclic compounds containing nitrogen and sulphur atoms *Advances in Chromatography and Electrophoresis – Conferentia Chemometrica 2003 Budapest*, 2003. október 27-29. (poszter)
3. **Farkas O.**, Jakus J., Héberger K. Quantitative structure–antioxidant activity relationship of flavonoids *Computer Applications and Chemometrics in Analytical Chemistry Balatonfüred*, 2004. augusztus 31-szeptember 3. (poszter)
4. **Farkas O.**, Héberger K. Classification of antipsychotic drug candidates on the basis of their hERG activity *Workshop on Chemometrics and Molecular Modelling Szeged*, 2005, április 6-8. (előadás, magyar)
5. **Farkas O.**, Stadler K., Héberger K. Classification of antidepressants based on their hERG activity *Conferentia Chemometrica 2005 – Chemometrics VII Hajdúszoboszló*, 2005. augusztus 28-30. (poszter)
6. **Farkas O.**, Zenkevich I.G., Stout F., Kalivas J.H., Héberger K. Prediction of gas chromatographic retention indices for fatty acid methyl esters *Comparison of variable*

selection methods *Conferentia Chemometrica 2005 – Chemometrics VII* Hajdúszoboszló, 2005. augusztus 28-30. (poszter)

7. **Farkas O.**, Zenkevich I.G., Stout F., Kalivas J.H., Héberger K. Hogyan hasonlítsunk össze változó-kiválasztási módszereket? „Legjobb” vagy egyszerű modellek *KeMoMo* Szeged, 2006. április 27-28. (előadás, magyar)

8. **Farkas O.**, Zenkevich I.G., Stout F., Kalivas J.H., Héberger K. Prediction of Kováts indices for fatty acid methyl esters Optimal or parsimonious models? *Conferentia Chemometrica* , Budapest, 2007. augusztus 31-szeptember 3. (poszter)