

Különböző szerkezetű porfirin származékok anti-HIV-1 aktivitásának modellezése

VANYÚR Rozália*, HÉBERGER Károly és JAKUS Judit

MTA Kémiai Kutatóközpont, Kémiai Intézet, Pusztaszeri út 59-67, 1025 Budapest, Magyarország

1. Bevezetés

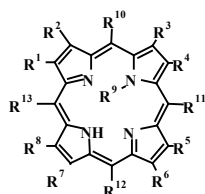
A modern orvostudomány a porfin-vázak és rokon vegyületeket egyre szélesebb körben alkalmazza terápiás célra^{1,2} (pl. szilárd tumorok, makula degeneráció, pikkelysömör kezelése) valamint vér és vértermékek fotodinamikus sterilizálására³. Bizonyos fotoszenzibilizátorként ismert vegyületek^{4,5} fotodinamikus effektus nélkül is képesek inaktiválni a HIV-1 vírust azzal, hogy a gp120-as kapszidfehérjéhez kapcsolódva megakadályozza, hogy a vírus a T limfocitákat megfertőzze^{6,7}.

A hatékony fotoszenzibilizátorok és vírusellenes szerek fejlesztésében fontos szerepet kaphatnak a nagyszámú molekula tesztelésére alkalmas szerkezet – hatás vizsgálatok, amely elméleti számítások alapján a hatástalan vegyületek a költséges és bonyolult szintetizálás, illetve biológiai rendszerekben való vizsgálat nélkül is kiszekelhetők.

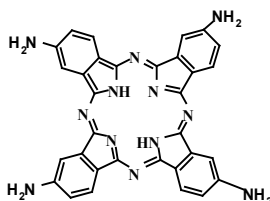
Ebben a munkánkban porfin-vázak vegyületek nem fotodinamikus anti-HIV-1 aktivitását modelleztük többszörös lineáris regresszió (multiple linear regression -MLR), részleges legkisebb négyzetek módszere⁸ (partial least squares projection to latent structures -PLS), valamint mesterséges ideghálózat⁹ (artificial neural network -ANN) segítségével. A vizsgálatba bevont vegyületek száma, valamint a szerkezetük közötti lényeges eltérés lehetőséget ad arra, hogy az épített modellekkel számos, anti-HIV-1 aktivitásra eddig nem vizsgált porfin-vázak vegyületek vírus inaktiválási képességét is megbecsüljük.

2. Anyagok és módszerek

2.1. Molekulaszerkezet



Porfin-váz



2,9,16,23-tetraamino-ftalocianin (M13)

1. Ábra. Molekulák (M) alapvázának kémiai szerkezete

2.2. Biológiai adatok

Az általunk a számításokba bevont 20 különböző tetrapirrol-vázak vegyület HIV-1 vírus inaktiválási képességét Debnath és munkatársai publikálták 1994-ben¹⁰.

1. Táblázat. Hematoporfirinek és a klorin e₆ funkciós csoportjai

	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	R ⁷	R ⁸	R ⁹	R ¹²
M1	Me	V	Me	V	Me	P	P	Me	H	H
M2	Me	V	Me	V	Me	P	P	Me	Me	H
M3	Me	Et	Me	Et	Me	P	P	Me	H	H
M4	Me	HO-Et	Me	V	Me	P	P	Me	H	H
M5	Me	V	Me	HO-Me	Me	P	P	Me	H	H
M6	Me	G	Me	G	Me	P	P	Me	H	H
M7	A	P	A	P	A	P	P	A	H	H
M8	A	P	A	P	A	P	A	P	H	H
M9	Me	P	Me	P	Me	P	A	P	H	H
M10	Me	P	A	P	Me	P	A	P	H	H
M11	A	P	Me	P	A	P	A	P	H	H
M12	Me	V	Me	Et	Me	C	P	Me	H	A

A: -CH₂CO₂H, C: -CO₂H, Et: -CH₂CH₃, HO-Et: -CHOHCH₃, Me: -CH₃, HO-Me: -CH₂OH, G: -CHOHCH₂OH, P: -CH₂CH₂CO₂H, V: -CHCH₃ M1: protoporfirin IX; M2: N-metil-protoporfirin IX; M3: mezoporfirin IX; M4: 3-hidroxi-etil-8-vinil-deuteroporfirin IX; M5: 3-vinil-8-hidroxi-metil-deuteroporfirin IX; M6: 2,4-bisglikol-deuteroporfirin IX; M7: uroporfirin III; M8: uroporfirin I; M9 pentakarboxil-porfirin I; M10: hexakarboxil-porfirin I; M11: heptakarboxil-porfirin I; M12: Klorin e₆

2. Táblázat. Mezo-tetrafenil-porfirinek funkciós csoportjai

	R ⁹	R ¹⁰	R ¹¹	R ¹²	R ¹³
M14	H	Ar	Ar	Ar	Ar
M15	H	ΦCOOH	ΦCOOH	ΦCOOH	ΦCOOH
M16	CH ₃	ΦCOOH	ΦCOOH	ΦCOOH	ΦCOOH
M17	H	ΦCOOH	ΦCOOH	ΦH	ΦCOOH
M18	H	ΦCOOH	ΦCOOH	ΦCl	ΦCOOH
M19	H	ΦCH ₃	ΦCH ₃	ΦCOOH	ΦCH ₃
M20	H	ΦCONH ₂	ΦCONH ₂	ΦCONH ₂	ΦCONH ₂

Ar: -C₆H₅, Φ: -C₆H₄-; M14: tetrafenil-porfirin; M15: tetra(4-karboxi-fenil)-porfirin; M16: N-metil-tetra(4-karboxi-fenil)-porfirin; M17: 5-fenil-10,15,20-tri(karboxi-fenil)-porfirin; M18: 5-(p-klorofenil)-10,15,20-tri(karboxi-fenil)-porfirin; M19: 5-(p-karboxi-fenil)-10,15,20-tri(tolil)-porfirin; M20: tetra(4-karboxi-amido-fenil)-porfirin.

A megadott anti-HIV-1 aktivitás (log 1/EC₅₀): azon porfirin koncentráció (mM) reciprokának logaritmus, amely mellett P24-es HIV-1 magfehérje termelése 50%-ra csökken vírussal fertőzött MT-2 humán T limfocita sejtekben. A vírusfehérje mennyiségi meghatározását ELISA (enzyme-linked immunoadsorbent assay) teszttel végezték.

*Vanyúr Rozália. Tel.: 4384141/152; Fax: 3257863; e-mail: vanyur@chemres.hu

2.3. Geometria-optimalizálás és konformáció analízis

A 3D szerkezetek geometria-optimalizálását és konformáció analízisét a HyperChem[®] programmal¹¹, azon belül az MM⁺ nevű molekula mechanikai módszer felhasználásával végeztük. Az adott molekulához tartozó legkisebb potenciális energiájú három dimenziós szerkezetet 1024 konformer közül választottuk ki. A kétdimenziós (2D), a fenti háromdimenziós (3D) szerkezeteket, valamint a biológiai aktivitásadatokat IsisBase programmal¹² rendeztük adatházisba.

2.4. Számítások

A két- és háromdimenziós szerkezetekből 87 elméleti leíró változót (deskriptort) számítottuk ki a 3DNET programmal¹³. Az eredeti 20 molekula közül hármat (**M3**, **M7** és **M17**; lásd 1. ábra) félretettünk külső ellenőrzésre, majd a maradék 17 molekula alapján a kémiai szerkezet - anti-HIV-1 aktivitás - összefüggést jól leíró modelleket építettünk mindhárom módszerrel. Az MLR számításokat a StatisticaTM programcsomaggal¹⁴, míg az ANN számításokat a 3DNET programmal végeztük. A PLS módszer megvalósítására mindkét programot alkalmaztuk.

2.5. Változókiválasztás és modellépítés

A többszörös lineáris regressziós számításoknál az előzetesen kiszámított deskriptorok közül az első modell (MLR_I) esetében előre irányuló változóbevonással (forward selection, FS), míg a második modell (MLR_II) esetében egy főkomponenselemzést (principal component analysis, PCA) követő előreirányuló változóbevonással választottuk ki azokat a deskriptorokat amelyek 5%-os szinten mégszignifikánsnak bizonyultak. Az eredeti 87 deskriptor közül kiválasztott 15 változóra vonatkozó pontok a kétdimenziós főkomponens együttható térben közel esnek az HIV-1 aktivitásra vonatkozó ponthoz.

Arészlegeslegkisebbnégyzetekmódszerénekalkalmazásakor az eredeti deskriptorok lineáris kombinációjával létrehozott új változókkal építettünk modelleket. Az első lépésben nyert modellenél 3 PLS komponenst képeztünk az eredeti 87 változó kombinációiból, majd a következő modellépítési eljárásban résztvevő 18 eredeti deskriptort a regressziós koefficiensek abszolút értéke alapján (pl. >0,1) választottuk ki. A fenti módszert alkalmazva, több lépésben a változók számát tovább csökkentettük. A legjobb modellben (PLS_III_9C) 11 deskriptor kombinációjából képzett 9 komponens vett részt. A modelloptimalizálást a 3DNET program auto-QSAR egységével folytattuk. A kapott 3 komponensű modell (PLS_IV_3C) 3 deskriptort használt fel a rendelkezésre álló 11 változóból, s ezzel valójában egy MLR modellnek felel meg.

Az általunk használt háromrétegű, visszafuttatásos algoritmussal tanuló, mesterséges ideghálózat szerkezetét és működését elvét korábbi közleményünkben részletesen ismertettük¹⁵. Az alkalmazott ideghálózatok rejtett rétege három (ANN_V és ANN_VII modellek esetében) és négy (ANN_VI, ANN_VIII és ANN_IX modellek esetében) neuront tartalmaznak. A legígéretesebb két-, (ANN_VII, ANN_VIII, ANN_IX) és három deskriptoros (ANN_V, ANN_VI) modellek statisztikai jellemzőit vizsgáltuk.

A modellek aktivitást leíró képességét az R^2 -tel, a többszörös determinációs együtthatóval jellemeztük.

2.6. Modellek ellenőrzése

Az egy-elem-kihagyásos (leave-one-out, LOO) keresztellenőrzés során a modell előrejelzési képességét vizsgáltuk. A keresztellenőrzést minden egyes molekulára elvégeztük úgy, hogy kihagytuk azt a modellépítési eljárásból, majd a nélküle megépített modellel a hozzátartozó aktivitásértéket becsültük meg. A becsült aktivitást összehasonlítottuk a kísérleti biológiai aktivitással. Az előrejelzés jósága a Q^2 értékével adható meg.

$$Q^2 = 1 - \frac{\sum_{i=1}^{17} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{17} (Y_i - \bar{Y})^2}$$

ahol i = tesztelt molekulák száma, Y_i = kísérleti aktivitás, $\hat{Y}_i$ = számított aktivitás a kihagyott molekula nélkül épített modellel és $\bar{Y}$ = kísérleti aktivitások kihagyott molekula nélküli átlaga.

A bemenő adatok véletlenszerű összekeverése után is meghatároztuk az egy-elem-kihagyásos Q^2 értékeket.

A modellek előrejelző-képességét több-elem-kihagyásos (leave-n-out, LNO) keresztellenőrzéssel is ellenőriztük. A változókiválasztásban résztvevő 17 vegyület közül számításanként 4 molekulát kihagytunk a modellépítési eljárásból (az adott deskriptor kombináció változtatása nélkül), majd megbecsültük aktivitás értékeiket. Az eljárást nyolcszor ismételtük meg úgy, hogy minden egyes molekula legalább egyszer a teszt-csoportba kerüljön. Az előrejelzés jóságát számításanként meghatároztuk, majd a nyolc Q^2 érték átlagával jellemeztük az adott deskriptor kombinációt.

Legutolsó ellenőrző lépésként, mindegyik (17 molekulával betanított) modellel megbecsültük a három, változózószektálási és modellépítési eljárásból kihagyott vegyület aktivitás értékét.

3. Eredmények és értékelésük

A HIV-1 vírus inaktiválásának előrejelzésére szolgáló, ígéretes modellek deskriptorait az 3. táblázatban, a szerkezet – aktivitás – összefüggést leíró és előrejelző tulajdonságokat a 4. táblázatban foglaltuk össze.

A PLS modellek aktivitást leíró képességei nagyon jók (PLS 9C modellre $R^2 = 0.9541$ (SD = 0.1576) és PLS 3C modellre $R^2 = 0.9982$ (SD = 0.0281)), de előrejelző képességeik nem érik el a 0.6 körüli értéket (mindkét modellre $Q^2 = 0.417$), így azokat a további vizsgálatokból kihagytuk.

Mindkét MLR modell (MLR I és MLR II) aktivitást leíró képessége elmarad az egyéb módszerekkel épített modellek hasonló tulajdonságaitól. Ennek ellenére, az egy-elem-kihagyásos keresztellenőrzés szerint, Q^2 értékeik meghaladják az irodalomban közölt, CoMFA módszerrel épített modell előrejelző képességét (0.59)¹⁰.

3. Táblázat. Az aktivitás előrejelzésére szolgáló legjobb modellek

Modellek	Deszkriptorok
MLR_I_FS	AMASS, MASS1, ESTA
MLR_II_PCA	POL1, APOL, μ
ANN_V_3N	KMASS, DF, DBE
ANN_VI_4N	HOMO, DF, DBE
ANN_VII_3N	DF, DBE
ANN_VIII_4N	DF, DBE
ANN_IX_4N	DF, ESTA

DBE: kettőskötés ekvivalens deszkriptor; DF¹⁶: kémiai kötés forgási szabadsági foka; HOMO¹⁷: legmagasabb betöltött molekulapálya energiája; KMASS¹⁸: atomi tömeggel súlyozott molekula-alak deszkriptor; AMASS, APOL¹⁸: atomi tömeggel továbbá atomi polarizációval súlyozott molekula-méret deszkriptorok; MASS1, POL1¹⁸: atomi tömeggel továbbá atomi polarizációval súlyozott irányfüggő sajátérték; ESTA¹⁷: elektrosztatikus teljes hidrogén kötés savasság; μ ¹⁹: számított dipólusmomentum.

4. Táblázat. Az aktivitást leíró modellek jellemzése

Modellek	R^2	(SD)	Q^2
MLR_I_FS	0.8275	(0.4440)	0.715
MLR_II-PCA	0.7898	(0.4901)	0.650
ANN_V_3N	0.9654	(0.1242)	0.770 ^a
ANN_VI_4N	0.9877	(0.0787)	0.846
ANN_VII_3N	0.9654	(0.1368)	0.673 ^a
ANN_VIII_4N	0.9040	(0.2280)	0.475
ANN_IX_4N	0.6766	(0.4183)	0.567 ^b

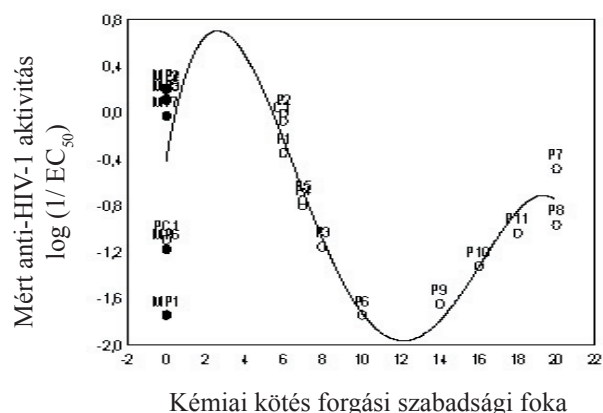
PCA: főkomponens-elemzés (principal component analysis), FS: előre irányuló változóbevonás (forward selection), N: rejtett idegsejtek száma (neuron), R^2 : többszörös determinációs együttható, SD: reziduális szórás, Q^2 : az előrejelzés jósága. ^a: M8-as vegyület kihagyásával számított érték, ^b: M13-es vegyület kihagyásával számított érték.

A legjobb modellt (ANN_VI) alkotó deszkriptorok a kémiai kötés forgási szabadsági foka (DF), a kettőskötés ekvivalens (DBE= 1x a gyűrűk száma + 1x a kettőskötések száma + 2x a hármaskötések száma) deszkriptor, valamint a legmagasabb betöltött molekulapálya (HOMO) energiája. A legjobb, 3 rejtett neuron alkalmazásával megépített modellben (ANN_V) a HOMO energia helyett az atomi tömegekkel súlyozott, molekula-alak deszkriptor szerepel (KMASS). Ez a modell azonban az LOO ellenőrzés alapján nem alkalmas az uroporfirin I (M8) aktivitásának becslésére. A deszkriptor kombináció előrejelző képessége ($Q^2=0.549$ -ről) nagymértékben javult ($Q^2=0.77$ -re), amikor a vegyület szerepelt a tanító halmazban, de aktivitás értékét nem becsültük meg a keresztellenőrzéskor.

Mivel mindkét három deszkriptoros ANN modellben szerepelnek a DF és DBE változók, ezért három és négy rejtett idegsejt szám mellett is teszteltük kombinációjukat. A három idegsejtet tartalmazó modellnél (ANN_VII) ugyancsak akkor kaptunk elfogadható Q^2 értéket, ha az uroporfirin I (M8) vegyületet kihagytuk a tesztcsoportból, de a tanító csoportból nem (Q^2 érték 0.345-ről 0.673-ra javult).

Az egyes deszkriptorok és a vizsgált biológiai aktivitás közötti összefüggést ábrázolva jól látható, hogy a kémiai

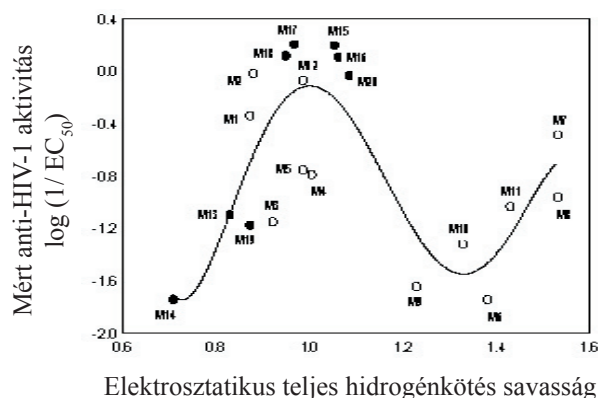
kötés forgási szabadsági fokával nagyon jól leírható a hematóporfirin származékok és a klorin anti-HIV-1 aktivitása, de nem képes megkülönböztetni egymástól az egyes tetrafenil-porfirineket és a ftalocianint (2. ábra).



2. Ábra. DF-anti-HIV-1 aktivitás összefüggés

A kettőskötés ekvivalens deszkriptor (DBE) és a legmagasabb betöltött molekulapálya (HOMO) energiája az egyes tetrafenil-porfirinek elkülönítéséhez járul hozzá, míg a molekula-alak deszkriptor (KMASS) a ftalocianint választja el ettől a vegyületcsoporttól (nem ábrázoltuk).

Az ANN_IX modellben a kémiai kötés forgási szabadsági foka (DF) mellett az elektrosztatikus teljes hidrogén kötés savasság deszkriptor szerepel (ESTA) (3. ábra). Az ESTA deszkriptor és az anti-HIV-1 aktivitás közötti összefüggés nemlineáris. A függvény, bár jól megkülönbözteti a porfirineket, a tetemes reziduális szórások miatt előrejelzésre nem alkalmas önmagában.



3. Ábra. ESTA – anti-HIV-1 aktivitás összefüggés

3.1. Modellek ellenőrzése

Az n-elem-kihagyásos keresztellenőrzés, valamint a három nem vizsgált porfirin aktivitásának előrejelzésére kapott eredményeket az 5. táblázatban foglaltuk össze. Az MLR_I modell az n-elem-kihagyásos ellenőrzéskor nagyon jó eredményt adott, de a külső minták aktivitásértékét rosszabbul jelezte előre, mintha azt a 17 molekula kísérleti aktivitásainak átlagával becsültük volna.

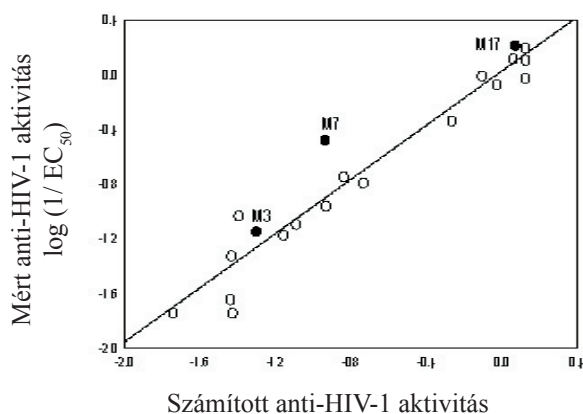
5. Táblázat. Az aktivitást leíró modellek ellenőrzése

Modellek	N-elemes Q^2 ^a (SD)	Külső Q^2
MLR_I_FS	0.730 (0.094)	-1.253
MLR_II-PCA	0.390 (0.368)	-
ANN_V_3N	0.747 (0.122)	0.788
ANN_VI_4N	0.879 (0.136)	0.612
ANN_VII_3N	0.584 (0.766)	0.773
ANN_VIII_4N	0.342 (1.124)	-
ANN_IX_4N	0.619 (0.279)	0.742

^a : nyolcszor, különböző vegyület kombinációval megismételt négy-elem-kihagyásos keresztellenőrzéskor kapott értékek átlaga, s a hozzá tartozó reziduális szórás.

A két három-deszkriptoros ANN modell (ANN_V és ANN_VI) több mint 70%-os megbízhatóságot mutatott az n-elem-kihagyásos ellenőrzés szerint.

A külső mintás vizsgálatnál három modell Q^2 értéke is meghaladja a 0.7-et (ANN modell V,VII és IX). A 4. ábrán bemutatjuk a három tesztvegyület aktivitását legjobban előrejelző modellel készült becslést.



4. Ábra. Külső ellenőrzés az ANN_V modellel

A modell előrejelző képességét jellemző $Q^2 = 0.788$ érték azt jelenti, hogy az adott deszkriptor-kombináció a három vegyület közötti különbséget 78.8%-ban képes a kémiai szerkezetükbeli különbségek alapján megmagyarázni.

4. Összefoglalás

Célunk volt, hogy mennyiségi szerkezet-hatás-összefüggést mutassunk ki fotoszenzibilizátorként megismert tetrapirrol-vázás molekulák háromdimenziós szerkezete és nem fotodinamikus anti-HIV-1 aktivitása között.

A modellépítési eljárások összehasonlításával megállapítottuk, hogy mindhárom módszer (MLR, PLS és ANN) alkalmas a porfirinek kémiai szerkezete és HIV-1 vírus inaktíváló hatása közötti kapcsolat leírására. A belső és külső ellenőrzések szerint megbízható előrejelzést csak a mesterséges ideghálózattal épített modellekkel kaptunk. A

legfontosabb paraméter mindegyik ANN modell szerint a kémiai kötés forgási szabadsági foka (DF). Ez a változó azonban önmagában nem alkalmas a vizsgált aktivitás becslésére, még legalább egy, a kettőskötés ekvivalens deszkriptor (DBE) vagy az elektrosztatikus teljes hidrogénkötés savasság deszkriptor (ESTA) hozzájárulása szükséges. A belső ellenőrzési mutatók még jobban javíthatók egy harmadik változó alkalmazásával (HOMO vagy KMASS).

Mindent összevetve elmondhatjuk, hogy a mesterséges ideghálózattal épített modelljeink alkalmasak arra, hogy egy nagyobb mintaszámú, ebben a vizsgálatban nem szereplő, porfin-vázás csoport anti-HIV-1 aktivitását előrejelezzük. A külső ellenőrzési eljárásba bevont molekulák kis száma miatt ajánljuk, hogy a négy legjobb ANN modellel becsült $\log(1/EC_{50})$ értékek átlagát fogadjuk el várható aktivitásértéknek.

Köszönetnyilvánítás

A munka az Országos Tudományos Kutatási Alap (T034986) támogatásával készült.

Hivatkozások

- Pandey R. K., *J. Porphyr. Phthalocyan.*, **2000**, 4, 368-373.
- Gál D., *Magy. Kém. Folyóirat*, **1995**, 101, 381-387.
- Moor A. C. E., *Photodynamic sterilisation of cellular blood products* –Proefschrift–, Sen Haag, **1998**, 46-65.
- Neurath A. R., Strick N., Debnath A. K., *J. Mol. Rec.*, **1995**, 8, 345-357.
- Song R., Witrouw M., Schols D., Robert A., Balzani J., De Clercq E., Bernadou J., Meunier B., *Antivir. Chem. Chemother.*, **1997**, 8, 85-97.
- Neurath A. R., Haberfield P., Joshi B., Hewlett I. K., Strick N., Jiang S., *Antivir. Chem. Chemother.*, **1991**, 2, 303-312.
- Neurath A. R., Strick N., Haberfield P., Jiang S., *Antivir. Chem. Chemother.*, **1992**, 3, 55-63.
- Pap T., Szilágyi A., *Magy. Kém. Folyóirat*, **2001**, 107, 60-70.
- Kövesdi I., Dominguez-Rodriguez L., Örfi G., Náray-Szabó G., Varró A., Papp J. Gy., Mátyus P., *Med. Res. Rev. Des.*, **1999**, 19, 249-269.
- Debnath A. K., Jiang S., Lin K., Haberfield P., Neurath A. R., *J. Med. Chem.*, **1994**, 37, 1099-1108.
- HyperChem version 4.0 program, HyperCube Inc., Canada, **1997**.
- IsisBase version 1.2 program, Molecular Design Ltd., **1996**.
- 3DNET 1.0 program, ViChem Kft., Magyarország, **1998**.
- Statistica™ version 5.5 program, Statsoft Inc., USA, **1999**.
- Vanyúr R., Héberger K., Kövesdi I., Jakus J., *Photochem. Photobiol.*, **2002**, 75, 471-478.
- Andrews P. R., Craik D. J., Martin J. L., *J. Med. Chem.*, **1984**, 27, 1648-1657.
- Cronce T. D., Famini G. R., De Soto J. A., Wilson L. Y., *J. Chem. Soc. Perkin Trans. 2*, **1998**, 2, 1293-1301.
- Todeschini R., Gramatica P., *Perspect. Drug. Discov.*, **1998**, Dec., 9/19/11, 355/380.
- Mortier W. J., van Genechten K., Gasteiger J., *J. Am. Chem. Soc.*, **1985**, 107, 829-835.

Modeling of anti-HIV-1 activity of a diverse set of porphyrin derivatives

Various modeling methods {multiple linear regression (MLR), projection of latent structures (PLS) and artificial neural networks (ANN)} were compared in the modeling of anti-HIV-1 activity of porphyrin derivatives (see Figure 1, Table 1 and Table 2).

The molecular structures were characterized by various topological and quantumchemical descriptors. Based on the three-dimensional structure of the geometry-optimized molecules, 87 theoretical descriptors were calculated by 3DNET program. These input descriptors were tested as independent variables and used for model building.

All methods were suitable to build models to describe the studied biological activity (high R^2 values in Table 4).

The predictive abilities of the best fitting models were checked by leave-one-out (LOO) (Table 3 and Table 4), leave-n-out (LNO) cross-validation and also by external validation (Table 5). The models built by MLR could not predict the biological activities of compounds in the external validation set, while PLS models also gave bad results during the validation processes (LOO $Q^2=0.417$). Only ANN models have shown good predictive ability in all validation steps, due to the inherent nonlinearity of the data sets (see figures).

The degree of chemical bond rotational freedom descriptor (DF) was the most important variable, but could not predict the activity without other descriptors (Figure 2). The double bond equivalent descriptor (DBE) and WHIM descriptor weighted by atomic mass (KMASS), or energy of highest occupied molecular orbital (HOMO), or electrostatic total hydrogen bond acidity (ESTA) were also needed to predict anti-HIV-1 activity.

The best prediction was achieved by a three-descriptor ANN model, model ANN_V (Figure 4). The external Q^2 value of this model was 0.788, explaining 78.8 % of the changes in the activity. Nevertheless, all of the ANN models were able to predict the studied activity with a $Q^2 > 0.6$.

Our ANN models were able to predict biological activities of a wide range of nontested tetrapyrrole molecules on the basis of their three-dimensional structures. Because of the low number of compounds in external validation set, we suggest to use the average value of biological activities calculated with the best four ANN models to predict anti-HIV-1 activity of similar compounds.

Beérkezett: 2003. VIII. 1.