

MODELAREA RELAȚIEI STRUCTURĂ-TOXICITATE A COMPUȘILOR METAL ALCHIL PRIN INTEGRAREA INFORMAȚIILOR STRUCTURALE COMPLEXE

Sorana Daniela Bolboacă*, L. Jäntschi**

REZUMAT

Compușii metal alchil sunt toxine ubicuitare cu proprietăți imunotoxice sau/și neurotoxice. Pornind de la informațiile complexe oferite de structura moleculară a zece compuși metal alchil, toxicitatea acestora a fost modelată prin aplicarea unei metodologii originale. Evaluarea și validarea modelelor obținute s-a realizat prin studiul coeficienților de corelație, a parametrilor statistici asociați modelelor și a coeficienților de corelație încrucișată. Modelul cu capacitatea cea mai bună de predicție ($r^2_{cv(100)} = 0.9965$) evidențiază că toxicitatea compușilor metal alchil este deopotrivă de natură topologică și geometrică, depinde de masă și sarcina parțială și este în relație directă cu lucrul și câmpul proprietății.

Cuvinte cheie: Relații Structură-Activitate (SAR); Compuși metal alchil; Integrarea informațiilor complexe, Familia Descriptorilor Moleculari (MDF)

ABSTRACT

Modeling of structure - toxicity relationship of alkyl metal compounds by integration of complex structural information

Alkyl metal compounds are ubiquitously toxins, known to be immunotoxic and/or neurotoxic. Starting with the complex information offered by the molecular structure of ten alkyl metal compounds, their toxicity was by applying an original methodology. The obtained models were evaluated and validated by means of correlation coefficients, statistical parameters of models, and by cross-validation correlation coefficients. The model with the highest predictive ability ($r^2_{cv(100)} = 0.9965$) shows that the toxicity of alkyl metal compounds is alike geometrical and topological, depend on the mass and partial charge, and is related with mechanical work of property and its field.

Key words: Structure-Activity Relationship (SAR); Alkyl metal compounds; Complex information integration, Molecular Descriptors Family (MDF)

Introducere

Compușii metal alchil sunt toxine ubicuitare cu proprietăți fungicide [Chandra & all, 1987], erbicide [Crowe, 2004], insecticide și bacteriostatice [Metrotra&Sungh, 2004]. Se cunoaște astăzi că, expunerea la compuși metal alchil este imunotoxică [Ade & all, 1996; Dacasto & all, 2001] sau/și neurotoxică [Aschner & Aschner, 1992]. De exemplu, expunerea la trietil plumb produce modificări histologice la nivelul hipocampului, în timp ce expunerea la trimetil plumb determină modificări histologice la nivelul măduvei spinării [Walsh & all, 1986]. Importanța studierii compușilor metal alchil rezidă astfel din efectele pe care aceștia le au asupra mediului și a organismului uman. O serie de cercetători au studiat corelația dintre acești compuși și efectele lor biochimice, stabilind existența corelației dintre structură și toxicitate [Eng & all, 1991; Laughlin & all, 1984].

Pornind de la informațiile complexe oferite de structura moleculară a zece compuși metal alchil, scopul cercetării a fost de a modela toxicitatea acestora prin aplicarea unei metodologii originale și de a evalua abilitățile modelelor SAR obținute în predicția toxicității compușilor metal alchil.

Material și metodă

Un eșantion de 10 compuși metal alchil au fost incluși în studiu. Denumirea compușilor, abrevierea lor și toxicitatea măsurată (exprimată ca $\log(LC_{50}) - \mu\text{mol/l}$) [Ade & all, 1996])

sunt prezentate în tabelul I.

Tabelul I Denumirea compușilor metal alchil, abrevierea și toxicitatea măsurată

Denumire compus	Abreviere	$\log LC_{50} (\mu\text{mol/l})$
Dibutil staniu	DBS	1.8457
Dietil plumb	DEP	1.8331
Tributil staniu	TSB	0.3979
Trietil plumb	TEP	1.5211
Trietil staniu	TES	2.1973
Trimetil plumb	TMP	2.4907
Trimetil staniu	TMS	3.4419
Tripenil	TPS	0.5441
Trifenil plumb	TFP	0.5315
Tripropil staniu	TPrS	0.7924

Toxicitatea celor zece compuși metal alchil a fost modelată prin integrarea informațiilor complexe oferite de structura moleculară a compușilor, prin aplicarea unei metodologii proprii, folosind familia Descriptorilor Moleculari. Metodologia aplicată în modelare a cuprins șase etape [Jäntschi, 2005]. Fiecărei etape îi corespunde unul sau mai multe programe PHP (Hypertext Pre-Processor). Toate calculele au fost realizate pe serverul <http://vl.academicdirect.org>.

*Sorana Daniela Bolboacă - Msc., Asistent Universitar, Universitatea de Medicină și Farmacie Iuliu Hațieganu

**Lorentz Jäntschi - Dr., MSc., șef lucrări, Universitatea Tehnică Cluj-Napoca

Prima etapă a modelării a fost destinată reprezentării tridimensionale a compușilor metal alchil care s-a realizat cu programul HyperChem [HyperChem, 2005]. În etapa a doua s-a creat fișierul care conține toxicitatea măsurată ($\log(LC_{50})$, exprimată în $\mu\text{mol/l}$) a compușilor metal alchil. Structurile tridimensionale ale compușilor și toxicitatea măsurată asociată acestora au intrat în etapa de treia, de generare, calculare și filtrare a membrilor familiei descriptorilor moleculari [Diudea & all, 2001; Janrschi & all, 2000]. În generarea listei familiei de descriptori moleculari s-au luat în considerare următoarele caracteristici, regăsite în denumirea fiecărui descriptor: geometria sau topologia moleculei (litera șapte în denumirea descriptorului), proprietatea atomică (masa atomică relativă, sarcina atomică parțială, cardinalitatea, electronegativitatea, electronegativitatea în grup, numărul de atomi de hidrogen legați direct - litera șase), descriptorul de interacțiune (litera cinci), modelul de suprapunere a interacțiunii descriptorilor (litera patru), metoda de fragmentare moleculară (litera trei), metode de cumulare a proprietăților de fragmentare (litera a doua) și procedura de linearizare aplicată în generarea descriptorului global molecular (prima literă). Odată generată lista descriptorilor moleculari, s-a trecut la etapa a patra, de căutare și identificare a celor mai semnificative modele mono- și/sau multi-variate SAR. Cele mai performante modele SAR identificate au intrat în etapa cinci, de validare, etapă în care fiecare compus metal alchil a fost exclus pe rând din analiză, s-au recalculat valorile coeficienților, și s-a prezis pe baza modelului obținut toxicitatea compusului exclus. Analiza de validare a modelelor SAR a avut ca rezultat calcularea coeficientului de validare încrucișată (r^2_{cv}), parametrului

Fischer și semnificația analizei de validare încrucișată (F_{pred} , $p_{pred}(\%)$) [Leave-one-out Analysis, 2005]. Etapa șase a constat în analiza modelelor SAR, analiză care s-a realizat pe baza următoarelor criterii: coeficientul de determinare, probabilitatea de model SAR greșit, coeficientul de validare încrucișată, probabilitatea unui model de validare încrucișată greșit și stabilitatea modelului dată de diferența dintre coeficientul de determinare al modelului SAR și coeficientul de validare încrucișată cu valoare cât mai mică. Tot în această etapă s-a realizat compararea performanțelor modelului mono-validat și a celui bi-variat prin analiza de corelare a coeficienților de corelație (testul Steiger [Steiger, 1980]).

Rezultate

În urma integrării cunoștințelor complexe, au fost identificate cele mai performante modele mono- și bi-variate SAR. Cele două modele SAR și statisticile asociate acestora sunt în tabelul II. În tabelul II s-a notat cu Y toxicitatea estimată, și cuprinde valorile coeficienților de corelație dintre fiecare descriptor în parte și toxicitatea măsurată ($r(\text{desc}, \log(LC_{50}))$), eroarea standard (ErStd), valorile extreme ale intervalului de încredere asociat coeficienților (CI_1 95% - valoarea inferioară, CI_5 95% - valoarea superioară), parametrul Student (t) și probabilitatea testului Student (p_t).

Abrevierea compușilor metal alchil, valorile descriptorilor moleculari folosiți în modelul mono- și bi-variat și toxicitatea estimată cu modelul mono-variat = $\hat{Y}_{\text{mono-v}}$, toxicitatea estimată cu modelul bi-variat = $\hat{Y}_{\text{bi-v}}$ sunt prezentate în tabelul III.

Tabelul II. Modele SAR și statistici de regresie

	$r(\text{descr}, \log(LC_{50}))$	CI_1 95%	CI_5 95%	ErStd	t	$P_t(\%)$
Mono-variat: $\hat{Y} = 0.335 + 0.252 \cdot iFDmCg$						
Intercept	-	0.101	0.569	0.102	3.298	$1.09 \cdot 10^0$
iFDmCg	0.9830	0.213	0.290	0.017	15.165	$3.54 \cdot 10^5$
Bi-variat: $\hat{Y} = 11.210 - 1.639 \cdot IHDmWmt + 0.372 \cdot LAMrEQg$						
Intercept	-	10.413	12.008	0.337	33.239	$5.78 \cdot 10^7$
IHDmWmt	-0.8396	-1.801	-1.478	0.068	-24.009	$5.53 \cdot 10^6$
LAMrEQg	0.9246	0.346	0.397	0.011	34.348	$4.60 \cdot 10^9$

Tabelul III Abrevierea compușilor, valorile descriptorilor moleculari și toxicitatea estimată cu ajutorul acestora (cele mai bune modele obținute)

Abreviere	Mono-variat		Bi-variat		
	iFDmCg	$Y_{\text{mono-v}}$	IHDmWmt	LAMrEQg	$Y_{\text{bi-v}}$
DBS	6.3257	1.9278	5.4095	-1.4056	1.8194
DEP	5.2300	1.6518	5.1908	-2.2845	1.8513
TBS	1.1821	0.6325	5.3537	-5.2584	0.4793
TEP	4.2832	1.4134	5.1084	-3.5195	1.5275
TES	7.0121	2.1006	5.1252	-1.7060	2.1738
TMP	9.3584	2.6915	4.7398	-2.5899	2.4772
TMS	11.868	3.3236	4.7446	0.0947	3.4668
TPS	0.4109	0.4382	5.4053	-5.0309	0.4793
TFP	0.0010	0.3350	5.6086	-3.9080	0.5632
TPrS	2.9605	1.0803	5.2630	-4.9172	0.7548

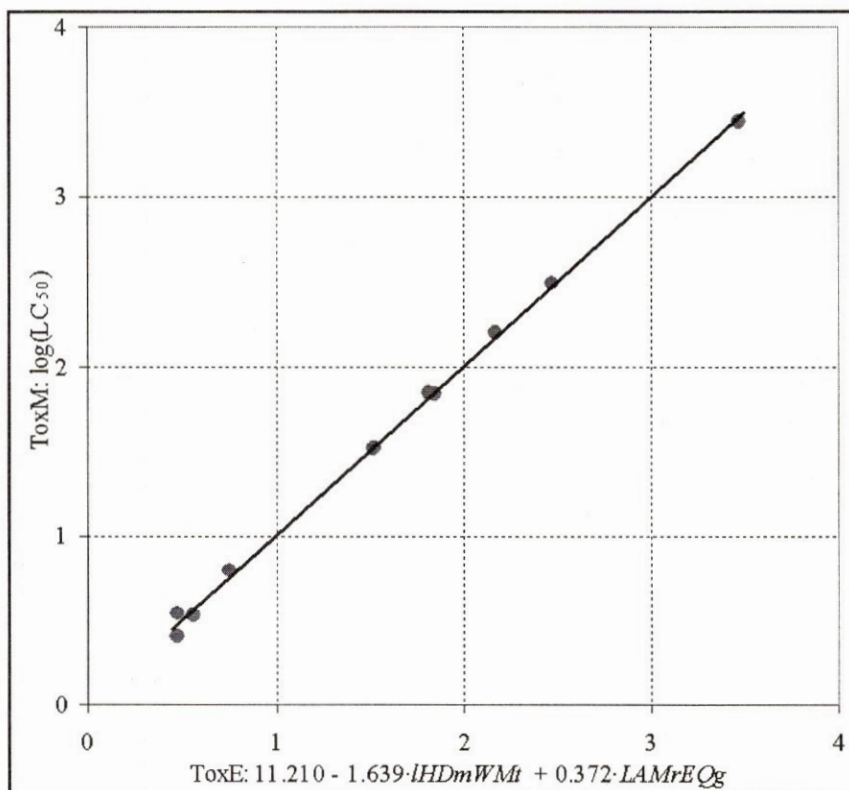


Fig. 2. Compuși metal alchil: toxicitatea măsurată vs. estimată din structură

Statisticile asociate modelului mono- și bi-variant, exprimate prin coeficientul de corelație (r), coeficientul de determinare (r^2), coeficientul de determinare ajustat (r^2_{adj}), eroarea standard a modelului (s_{est}), parametrul Fischer (F_{est}), probabilitatea unui model de regresie greșit exprimată procentual ($p_{est}(\%)$), coeficientul de validare încrucișată ($r^2_{cv(100)}$), parametru Fisher (F_{pred}) și probabilitatea unui model greșit de validare încrucișată ($p_{pred}(\%)$), eroarea standard a analizei de validare încrucișată (s_{100}) și diferența dintre coeficientul de determinare și coeficientul de validare încrucișată ($r^2 - r^2_{cv(100)}$) sunt în tabelul IV.

Tabelul IV

Statistici asociate modelelor SAR

Caracteristica	Model SAR	
	Mono-variant	Bi-variant
r	0.9830	0.9991
r^2	0.9664	0.9983
r^2_{adj}	0.9622	0.9978
s_{est}	0.1945	0.0473
F_{est}	230	2008
$p_{est}(\%)$	$3.54 \cdot 10^{-5}$	$2.20 \cdot 10^{-8}$
$r^2_{cv(100)}$	0.9466	0.9965
F_{pred}	141	989
$p_{pred}(\%)$	$2.29 \cdot 10^{-4}$	$2.61 \cdot 10^{-7}$
s_{100}	0.2454	0.0673
$r^2 - r^2_{cv(100)}$	0.0198	0.0018

Reprezentarea grafică a dependenței dintre toxicitatea compușilor metal alchil și structura acestora exprimată prin

modelul bi-variant este în figura 1.

Rezultatele comparării dintre modelele SAR obținute prin integrarea informațiilor complexe ale structurii compușilor metal alchil, respectiv între modelul mono-variant și cel bi-variant sunt în tabelul V. În tabelul V, s-a notat cu $r(\log(LC_{50}), \hat{Y}_{bi-v})$ coeficientul de corelație dintre toxicitatea măsurată ($\log(LC_{50})$) și cea estimată cu modelul bi-variant (\hat{Y}_{bi-v}), cu $r(\log(LC_{50}), \hat{Y}_{mono-v})$ coeficientul de corelație între toxicitatea măsurată și cea estimată cu modelul mono-variant (\hat{Y}_{mono-v}), cu $r(\hat{Y}_{bi-v}, \hat{Y}_{mono-v})$ coeficientul de corelație între toxicitatea estimată cu modelul bi-variant și cea estimată cu modelul mono-variant, cu Z parametrul punctual rezultat din aplicarea testului Steiger și cu p_z probabilitatea de coincidență a coeficienților de corelație obținuți de cele două modele.

Tabelul V Rezultate ale comparării modelului bivariat cu modelul mono-variant

Caracteristica	Valoare
$r(\log(LC_{50}), \hat{Y}_{bi-v})$	0.9991
$r(\log(LC_{50}), \hat{Y}_{mono-v})$	0.9830
$r(\hat{Y}_{bi-v}, \hat{Y}_{mono-v})$	0.9821
Z	3.9038
p_z (probabilitatea de coincidență, %)	$4.47 \cdot 10^{-3}$

Discuții

Toxicitatea compușilor metal alchil a fost modelată pe baza informațiilor complexe oferite de structura acestora. Rezultatele studiului arată că există o relație între toxicitatea compușilor metal alchil și structura acestora. Trei descriptorii moleculari s-au dovedit utili în estimarea și prezicerea toxicității compușilor metal alchil, doi dintre aceștia (IFD-

mdCg, LAMrEQg) considerând geometria moleculei și unul (IHDmWMT) topologia acesteia. În ceea ce privește proprietatea atomică descriptorul modelului mono-variabil consideră cardinalitatea (iFDmDcG), în timp ce descriptorii modelului bi-variabil consideră masa atomică relativă (IHDmWMT) și sarcina atomică parțială (LAMrEQg). Modelul mono-variabil consideră ca descriptor de interacțiune inversul distanței iar modelul bi-variabil lucrul proprietății (IHDmWMT) și câmpul acesteia (LAMrEQg). Ambele modele sunt semnificative statistic, având o probabilitate de model greșit mai mică de $3,54 \cdot 10^{-9}\%$.

Valoarea coeficientului de corelație al modelului mono-variabil ($r = 0.9830$, tabelul II) susține existența corelației dintre descriptorul iFDmDcG și toxicitatea compușilor metal alchil. Aproape 97% din variația toxicității poate fi explicată prin relația lineară dintre aceasta și descriptorul iFDmDcG. Modelul SAR mono-variabil este un model stabil ($r^2 - r_{cv(100)}^2 = 0.0198$, tabelul IV), capabil să prezică toxicitatea compușilor ($r_{cv(100)}^2 = 0.9466$, tabelul IV) cu o probabilitate de a greși egală cu $2.29 \cdot 10^{-4}$ (tabelul IV). Privind modelul mono-variabil în ansamblu putem spune că, toxicitatea compușilor metal alchil este de natură geometrică, depinde de cardinalitatea compușilor și de inversul distanței metrice.

Modelul bi-variabil prezintă un coeficient de corelație multiplă foarte aproape de valoarea 1 ($r = 0.9991$, tabelul IV), corelația dintre descriptorii modelului și toxicitatea compușilor fiind puternică. Valoarea coeficientului de determinare a modelului bi-variabil ne arată că, aproape sută la sută din variația toxicității compușilor metal alchil poate fi atribuită relației lineare cu descriptorii moleculari IHDmWMT și LAMrEQg. Valoarea coeficientului de validare încrucișată ($r_{cv(100)}^2 = 0.9466$, tabelul IV), a probabilității unui model de validare încrucișată greșit ($p_{pred} = 2.61 \cdot 10^{-7}\%$, tabelul IV) și a diferenței dintre coeficientul de determinare și coeficientul de validare încrucișată ($r^2 - r_{cv(100)}^2 = 0.0018$, tabelul IV) susțin validarea modelului bi-variabil și abilitatea acestuia în prezicerea toxicității compușilor metal alchil. Rezultatele analizei de regresie a modelului (valorile intervalului de încredere asociat coeficienților modelului, eroarea standard, valoarea parametrului Student și probabilitatea de a greși asociată testului Student - tabelul II) vin să susțină validitatea acestuia. Dacă ne uităm la valorile coeficienților de corelație dintre toxicitatea compușilor metal alchil cu fiecare descriptor molecular în parte, putem observa că, ambele corelații sunt semnificative ($r > 0.8$, tabelul II). Dar, valoarea coeficientului de corelație al modelului bi-variabil este mai mare în comparație cu valoarea coeficienților de corelație individuali, ceea ce vine să susțină superioritatea modelului bi-variabil care ia în considerare descriptorii IHDmWMT și LAMrEQg în comparație cu posibilele modele mono-variabile care iau în considerare descriptorul IHDmWMT sau descriptorul LAMrEQg. Din punct de vedere al modelului bi-variabil, putem spune că toxicitatea compușilor metal alchil este deopotrivă de natură geometrică și topologică, depinde de masa relativă, sarcina parțială, lucrul și câmpul proprietății.

Atât modelul mono-variabil, cât și cel bi-variabil sunt capabile să estimeze și să prezică toxicitatea compușilor metal alchil. Pentru a vedea însă dacă există o diferență semnificativă între valorile coeficienților de corelație obținuți cu cele două modele, s-a aplicat testul Steiger. Valoarea coeficientului de corelație obținut de modelul bi-variabil este semnificativ mai mare în comparație cu valoarea obținută de modelul mono-variabil ($p_z = 4.74 \cdot 10^{-3}\%$, tabelul V). Astfel, dacă dorim o predicție a toxicității cât mai apropiată de valoarea reală,

vom folosi modelul bi-variabil.

În lucrarea [Ade & all, 1996] au fost construite modele distincte pentru fiecare metal. Prin modelul obținut (modelul bi-variabil, tabelul II) se arată că modelul structură-activitate este independent de metal - și astfel se integrează informația structurală complexă într-o ecuație capabilă de predicții independente de ligand și de metal.

Modelele de predicție a toxicității compușilor metal alchil își găsesc utilitatea în prezicerea toxicității a compușilor noi, pe baza structurii acestora. Pentru obținerea toxicității unui nou compus este necesar în primul rând să desenăm, folosind programul HyperChem, structura tridimensională a acestuia. Odată ce avem structura compusului metal alchil ca fișier *.hin, folosind facilitățile programului MDF SAR Predictor [MDF SAR Predictor, 2005], prin alegerea setului corespunzător compușilor metal alchil (setul 52730) și al modelului SAR (mono-variabil sau bi-variabil) putem prezice toxicitatea compusului de interes. Se deschide astfel calea spre obținerea de informații utile în ceea ce privește toxicitatea noilor compuși metal alchil, informații obținute strict pe baza structurii compușilor, fără experimente directe, asistată de calculator, modalitatea mai ieftină și mai puțin consumatoare de timp.

Concluzii

Toxicitatea compușilor metal alchil poate fi modelată plecând de la structura acestora. Toxicitatea obținută cu cel mai performant model (modelul bi-variabil) este deopotrivă de natură topologică și geometrică, depinde de masa relativă și sarcina parțială, fiind în relație directă cu lucrul proprietății și câmpul acesteia.

Aplicarea metodologiei SAR permite obținerea de modele exacte care deschide calea spre prezicerea toxicității a noi compuși plecând de la structura acestora.

Notă

Cercetarea a fost suportată parțial de către Ministerul Educației și Cercetării prin proiectul ET36/UEFISC-SU/2005.

BIBLIOGRAFIE

1. Chandra, S., James, B.D., Macaulez, B.J. & Magee, R.J. (1987), "Studies on the fungicidal properties of some organo-tin compounds", J Chem. Technol. Biotechnol., vol. 39, no.1, pp. 65-73
2. Crowe, S.A. J. (2004), "Organotin compounds in agriculture since 1980. Part I. GFungicidal bactericidal and herbicidal properties", Appl. Organomet. Chem., vol. 1, no. 2, pp. 143-55
3. Methrotra, R.C. & Singh, A. (2004), Biological Applications and Environmental Aspects of Organometallic Compounds, In: Organometallic Chemistry. A Unified Approach. New Age International, New Delhi, p. 517-534
4. Ade, T., Zaucke, F. & Krug, H.F. (1996), "The structure of organometals determines cytotoxicity and alteration of calcium homeostasis in HL-60 cells", Fresenius J. Anal. Chem., vol. 354, pp. 609-614
5. Dacasto, M., Cornaglia E., Nebbis, C. & Bollo, E. (2001), "Triphenyltin acetate-induced cytotoxicity and

CD4 and CD8 + depletion in mouse thymocyte primary cultures", *Toxicology*, vol. 169, no. 3, pp. 227-238

6. **Aschner, M. & Aschner J.L. (1992)**, "Cellular and molecular effects of trimethyltin and triethyltin: Relevance to organotin neurotoxicity", *Neurosci. Biobehav. Rev.*, vol. 16, pp. 427-435

7. **Walsh, T.J., McLamb, R.L. & Bondy, S.C. (1986)**, "Triethyl and trimethyl lead: Effects on behavior, CNS morphology and concentrations of lead in blood and brain of rat", *NeuroToxicology*, vol. 7, no. 3, pp. 21-33

8. **Eng, G., Brinckman, F.E., Olson, G.J. Tierney, E.J. & Bellama, J.M. (1991)**. "Total surface areas of Group IVA organometallic compounds: Predictors of toxicity to algae and bacteria", *Appl. Organomet. Chem*, vol. 5, no. 10) pp. 33-37

9. **Laughlin, R.B., French, E.W., Johannesen, R.B., Guard, H.E. & Brinckman, F.E. (1984)**, "Predicting toxicity using computed molecular topologies: The example of triorganotin compounds", *Chemosphere*, vol. 13, no. 4 pp. 575-584

10. **Jantschi, L. (2005)**, "Molecular Descriptors Fa-

mily on Structure Activity Relationships 1. Review of the Methodology", *Leonardo Electronic Journal of Practices and Technologies*, vol. 6, pp. 76-98

11. HyperChem, Molecular Modelling System 2005; Hypercube, Inc., viewed 13 November, 2005, <<http://hyper.com/products>>.

12. **Diudea, M., Gutman, I. & Jantschi, L. (2001)**, *Molecular Topology*, Nova Science, Huntington, New York, pp. 332

13. **Jantschi, L., Katona, G. & Diudea, M. (2000)**, "Modeling Molecular Properties by Cluj Indices", *Commun Math Comput Chem (MATCH)*, Bayreuth, Germany, vol. 41, pp. 151-188

14. Leave-one-out Analysis 2005, Virtual Library of Free Software, view 20 November, 2005, <http://vl.academicdirect.org/molecular_topology/mdf_findings/loo>

15. **Steiger, J.H. (1980)**, "Test for comparing elements of a correlation matrix", *Psychol Bull*, vol. 87, pp. 245-251

16. MDF SAR Predictor 2005, Virtual Library of Free Software, viewed 20 November, 2005, <http://vl.academicdirect.org/molecular_topology/mdf_findings/sar>.