Quantum Mechanics Study on a Series of Steroids Relating Separation with Structure

Radu E. Sestraş, Lorentz Jäntschi*, and **Sorana D. Bolboacă**

Key Words

Separation analysis
Molecular modeling
Structure–property relationships
Steroids

Presented at the
17th International Symposium on Separation Sciences,
September 5–9, 2011,
Cluj-Napoca, Romania

Summary

A small series of steroid hormones and synthetic steroid anabolics, eluted from a sample by using classical methanol–water mobile phase, was designed at molecular level with *in vitro* geometrical modeling. Structure-based properties of the compounds were extracted for different strategies of geometry optimization and were related with the percent retardation factors observed in the experiment by using classical linear regression approach. The obtained models were assessed and the most accurate model was selected. The obtained relationship between the structure and retardation factor was interpreted in order to express the main factors driving liquid phase mobility of compounds as a function of their structure. The analysis allowed a series of general conclusions regarding the modeled property (percent of retardation factor), modeling strategy (with help of a family of molecular descriptors) as well as about the level of theory in geometrical modeling (Ab-initio 6-31G*, CLHF, core Hamiltonian guess, symmetric orthogonalization).

1 Introduction

In-silicon approaches are nowadays extremely useful methodologies for development of new chemical compounds with desired activity or property [1–3]. Energy optimization is the main application of molecular mechanics (the application of Newtonian mechanics used to model molecular systems) [4]. A series of ab initio methods was developed [4], although the time-consuming calculations are a major barrier in their utilization. The empirical or semi-empirical methods, including Austin Model 1 (AM1 [5]), CFF [6], Del Re [7], parameterized model (PM3 [8, 9], PM6 [10]), RM1 [11], Gasteiger and Marsili [12, 13], Hückel [14–17], Berthod et al. [18], optimized potentials for liquid simulations (OPLS [19]), assisted model building with energy refinement (Amber [20]), Merck molecular force field (MMFF [21–25]), are widely implemented in software due mainly to their speed.

Some studies were conducted in order to identify if the choice of energy method has or no impact on models obtained using quantitative structure-activity/property relationship (QSAR/QSPR) approaches. Rinnan et al. showed that the choice of the energy evaluation method has very limited impact [26]. Tsai and co-authors [27] evaluated twelve semi-empirical and empirical methods and showed that some methods such as Del Re and Pullman's are atom or bound-specific while the most accurate method is Austin Model 1 with Bond and Charge Correction (AM1-BCC) [28, 29]. So far, several methods for assignment to each atom the electrostatic potentials in the molecule are available, but no consent of the effects of electrostatic potential on quantitative structure-activity/property relationship models is available.

The aim of our research was to use a small series of (very) simple compounds with a (very) simple measured property to derive structure–property relationships relating compounds structure obtained at different moments during geometrical optimization process. This experimental design is meant to provide useful knowledge about the involved parameters (such as geometrical modeling method, software, method of extracting information from compounds structure, and measured chromatographic property).

2 Experimental

2.1 Compounds

The retardation factors (abbreviated as hR_F, expressed as percentage) for a sample of eleven steroid hormones, synthetic steroid anabolics (compounds with 20 to 29 heavy atoms) previously investigated [30], were included in the analysis **(Table 1).**

The 3D structure of the compounds was taken from PubChem. The following 3D properties are available for each compound [31, 32]: MMFF (Merck molecular force field) partial charges;

R.E. Sestraş and L. Jäntschi, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Mănăştur, 400372 Cluj-Napoca, Romania; L. Jäntschi, Technical University of Cluj-Napoca, 28 Memorandumului, 400114 Cluj-Napoca, Romania; and S.D. Bolboacă, "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, 13 Emil Isac, 400023 Cluj-Napoca, Cluj, Romania.
E-mail: lorentz.jantschi@gmail.com

DOI: 10.1556/JPC.25.2012.6.7

Table 1

Steroid hormones, anabolics: experimental retardation factor.

No.	Compound name	CID	hR_F (%)
1	Progesterone	5994	76.25
2	Trenbolone acetate	107845	58.75
3	Melengestrol acetate	250948	64.95
4	17β-Estradiol	5757	52.50
5	19-Nortestosterone	9904	56.25
6	Fluoxymesterone	6446	17.50
7	Norethandrolone	5858	63.75
8	4-Chloro-δ-1-methyltestosterone	227107	59.97
9	Clostebol acetate	13327	84.61
10	6β-Hydroxymethandienone	538883	19.23
11	Oxymetholone	5281034	66.66

CID = compound identification number; hR_F = retardation factor

(CID source: PubChem; available from: http://pubchem.ncbi.nlm.nih.gov/)

volume; steric monopole, quadrupole, octopole moments; MMFF94 energy [25, 33] (with coulombic terms removed); shape fingerprint; self-overlap volumes used in ST, CT similarity computation; conformer model RMSD; the conformer model diverse ordering; and pharmacophore features.

2.2 Translation of Compounds Structure into Information

The Massively Parallel Quantum Chemistry Program (abbreviated as MPQC, version 2.3.1 released on 2006-03-22, available from: http://www.mpqc.org/) was used to optimize the geometry of compounds included in the study. The MPQC program computes the properties of atoms and molecules from first principles using the time-independent Schrödinger equation [34].

Four case studies were conducted to accomplish the aim of the research:

– Case 1: 3D structures, partial charges as downloaded from PubChem (no geometry optimization).

– Case 2: compounds optimized using 20 steps/cycles (MPQC program with 6-31G*, CLHF, core Hamiltonian guess, symmetric orthogonalization).

– Case 3: compounds optimized using 30 steps/cycles with (MPQC program with 6-31G*, CLHF, core Hamiltonian guess, and symmetric orthogonalization).

– Case 4: complete optimization with MPQC program.

The PubChem MMFF94 charges (or ChemBioOffice whenever the MMFF94 charges were not available on the PubChem downloaded compounds) were used for case studies from 2 to 4.

The structural information of the compounds was translated in values of descriptors using homemade software that implemented structural atomic property family (SAPF, a total number of 144,060 descriptors) [35]. The SAPF approach is a method that cumulates atomic properties at the molecular level [36]. The SAPF descriptors calculated on a set of compounds can be available upon request. The SAPF descriptors that accomplish the following criteria were furthermore used in separation vs. structure analysis:

Criterion	Min/max	Value
Absolute variance of the descriptor value relative to the measured property	Minimum	0.02
Deviation from normality relative to measured property	Maximum	2.00
Determination between descriptor value and measured property	Minimum	0.02

2.3 Separation vs. Structure Analysis

Linear regression approach was used to relate the separation power (as dependent variable) with the compounds structure (as independent variable(s), expressed as molecular descriptors calculated with SAPF approach).

The simple linear regression was conducted to identify the "adapted" SAPF descriptors. Multiple linear regressions with two descriptors were conducted to identify the most accurate model for each investigated case. The geometric mean of Student *t*-values for coefficients (the higher the better – increases stability in prediction if $r_{cv\text{-}loo}^2$ closer to r^2) was used to identify the most accurate models.

The accuracy of multiple linear regression (MLR) models was assessed by conventional determination coefficient (r^2), standard error of estimate (se_{est}), *F*-value and its associated *p*-value, determination coefficient on leave-one-out analysis ($r^2_{cv\text{-}loo}$, a value higher than 0.5 indicate a robust model [37]), standard error of predicted (se_{pred}), F_{loo}-value and its associated p_{loo}-value (where loo = leave-one-out) [38].

The best MLR model was defined as the model with:

– Highest determination coefficient (both the conventional determination coefficient and cross validation leave-one-out determination coefficient). Steiger's *Z* test at a significance level of 5% was used to compare two correlation coefficients [39].

– Lowest standard error (both standard error of estimate and predicted).

– Lowest difference between conventional determination coefficients and leave-one-out cross validation determination coefficient.

– Lowest mean absolute error (MAE) where MAE = $(\Sigma|Y_i - \hat{Y}_i|)/n$ (where Y_i = measured hR_F, $\hat{Y}_i$ = estimated hR_F, i = the i^{th} compound in the dataset, and n = sample size).

– Lowest mean absolute percentage error (MAPE) where MAPE = $(\text{sum}|Y_i - \hat{Y}_i|/Y_i)/n$.

– Lowest standard error of prediction (SEP) where SEP = $\sqrt{((\text{sum}(Y_i - \hat{Y}_i)^2/(n-1))}$.

– Lowest relative error of prediction (REP%) where REP(%) = $100/m_Y \times \sqrt{(\text{sum}(\hat{Y}_i - Y_i)^2/n)}$ (where m_Y = the mean of observed hR_F).

– Lowest values of Akaike's information criteria: AIC (Akaike information criterion [40]); AIC_c (corrected AIC for bias-adjustment in small sample sizes, applied when the $n/k < 40$, where k = number of descriptors in the model) [41]; AIC_r^2 (AIC based on the determination coefficient); AIC_u (McQuarrie, Tsai corrected AIC) [42]; BIC (Schwarz or Bayesian information criterion, also abbreviated as SIC) [43]; APC = Amemiya prediction criterion [44]; and HQC (Hannan–Quinn criterion) [45].

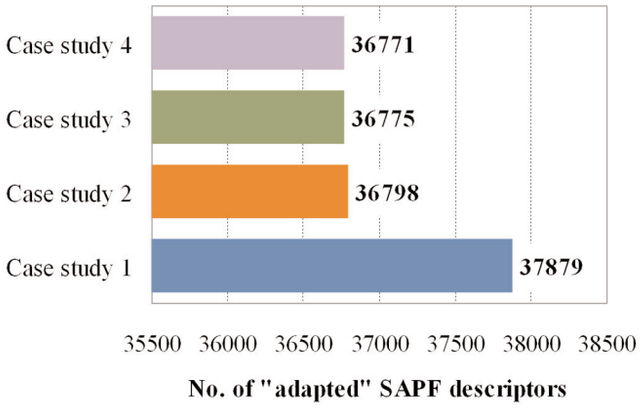

Figure 1

Distribution of "adapted" SAPF descriptors according with case study.

– Lowest relative distance from the "truth" defined as: $\Delta_i = AIC_i - \min(AIC)$, where Δ_i is the difference between the AIC of the best fitting model and the value of the i^{th} model; AIC_i = AIC corrected for model i; min(AIC) = minimum AIC value of all models) [46].

– Highest value of Kubinyi function (FIT) [47, 48].

3 Results and Discussion

Four distinct cases were investigated to identify the action of compounds geometry optimization on QSPR models. The number of SAPF descriptors that accomplished the imposed criteria proved different according to the case **(Figure 1).**

The highest number of "adapted" descriptors was identified in the first case study, when the geometry of compounds was not optimized. The highest difference (of 1081) of "adapted" descriptors was seen between second and first case study while the smallest difference (of 4) was seen between fourth and third case study. This result suggests that refinement of the semi-empirical MMFF94 model at HF-6-31G* (ab-initio) level of theory is suitable in molecular design of compounds subject to chromatographic separation as alternative to the method used in PubChem deposited 3D structures.

However, how many optimization cycles are necessary? The results regarding the cycle on which each compound achieves the optimum are presented in **Table 2.**

The analysis of the results presented in Table 2 revealed the following:

– 10 optimization cycles: any investigated compound achieves the geometry optimization.

– 20 optimization cycles (case study 2): 4 compounds achieve the optimum.

– 30 optimization cycles (case study 3): 5 compounds achieve the optimum.

– Complete optimization (case study 4): 2 compounds achieve the optimum.

A linear relationship between cycles and HOMO values with a correlation coefficient of 0.6844 ($p < 0.05$) was identified.

Two SAPF descriptors (LIMQQHTD and LIMSQHTD) were systematically identified in the most accurate models in simple linear regression analysis ($r^2 = 0.835$, were r^2 = determination coefficient).

The most accurate MLR models with two descriptors for each case study are presented in eq. (1) for case study 1 to eq. (4) for case study 4:

$$\hat{Y}_{case1} = 127.19(\pm 9.42) + SISHQHGC*(-5.71\cdot10^{-2}(\pm 7.44\cdot10^{-3}) + QQSSEHTD*(-5.71\cdot10^{4}(\pm 1.11\cdot10^{4}) \quad (1)$$

$$\hat{Y}_{case2} = -5.77\cdot10^{2}(\pm 92.13) + TESHIMGP* 1.52\cdot10^{-3}(\pm 1.86\cdot10^{-4}) + TGMSAEGC*6.77\cdot10^{4}(\pm 1.04\cdot10^{4}) \quad (2)$$

$$\hat{Y}_{case3} = -5.83\cdot10^{2}(\pm 79.27) + TESHIMGP* 1.54\cdot10^{-3}(\pm 1.60\cdot10^{-4}) + TGMSAEGC*6.83\cdot10^{4}(\pm 8.97\cdot10^{3}) \quad (3)$$

$$\hat{Y}_{case4} = -5.82\cdot10^{2}(\pm 79.31) + TESHIMGP* 1.54\cdot10^{-3}(\pm 1.60\cdot10^{-4}) + TGMSAEGC*6.83\cdot10^{4}(\pm 8.98\cdot10^{3}) \quad (4)$$

where $\hat{Y}$ = estimated the hR_F; SISHQHGC, QQSSEHTD, TESHIMGP, TGMSAEGC = SAPF descriptors; the number associated with ± is the value to be extracted, added to obtain a 95% confidence interval associated with the regression coefficients.

The same SAPF descriptors proved to provide the most accurate MLR model in 3 out of 4 cases. The exception is the case study when the geometry optimization was not performed before separation vs. structure analysis. The effect of two additive factors proved to link the hR_F of the investigated compounds with their structural information: a steric factor (the descriptor based on atomic mass, TESHIMGP) and a polarity factor (the descriptor based on atomic electronegativity, TGMSAEGC). The formulas used to calculate the descriptors found in the most accurate models presented by eqs. (2)–(4) are as follows:

– TESHIMGP descriptor:

$$TESHIMGP = (ES(array))^{-2}, \quad ES(array) = \left(\sum_{i=1}^{n}(V_i)^{-2}\right)^{-1/2}$$

$$V_i(HIMGP) = (AtomicMass(i))^{-0.5}\cdot(geom(i;P))^{-2}$$

$$geom(i;P) = \sqrt{(x_i - x_{refP})^2 + (y_i - y_{refP})^2 + (z_i - z_{refP})^2}$$

$$c_{refP} = \sum_{i=1}^{n} c_i \cdot AtomicMass(i) \Big/ \sum_{i=1}^{n} AtomicMass(i), \quad c = x, y, z$$

$$TESHIMGP = \sum_{i=1}^{n} AtomicMass(i)\cdot\left((x_i - x_{refP})^2 + (y_i - y_{refP})^2 + (z_i - z_{refP})^2\right)^2,$$

$$c_{refP} = \sum_{i=1}^{n} c_i \cdot AtomicMass(i) \Big/ \sum_{i=1}^{n} AtomicMass(i), \quad c = x, y, z$$

– TGMSAEGC descriptor:

$$TGMSAEGC = (GM(array))^{-2}, \quad GM(array) = \sqrt[n]{\prod_{i=1}^{n} V_i}$$

$$V_i(SAEGC) = (Electronegativity(i))^2\cdot(geom(i;C))^{0.5}$$

$$geom(i;C) = \sqrt{x_i^2 + y_i^2 + z_i^2}$$

$$TGMSAEGC = 1\Big/\sqrt[n]{\prod_{i=1}^{n}(Electronegativity(i))^4\cdot\sqrt{x_i^2 + y_i^2 + z_i^2}}$$

The coefficients of the regression models from eqs. (2) to (4) are slightly different with smallest differences between eqs. (3) and (4). This lowest difference between coefficients of eqs. (3) and (4) is expected since just 2 out of 11 structures were not optimized in the 3rd case study.

The parameters that describe the estimation and prediction power of the models from eqs. (1) to (4) are presented in

Table 2

The results of optimization.

No.	CID	Conv. cycles	Mol. no. atoms	HOMO (eV)	LUMO (eV)	GAP (eV)	SCF (Hartrees)	Nucl_Rep (Hartrees)
1	5994	18	53	–0.3580	0.1145	0.4725	–962.4	2074.8
2	107845	26	47	–0.2982	0.0766	0.3748	–995.9	1948.6
3	250948	58	61	–0.3257	0.0864	0.4121	–1264.8	2991.8
4	5757	61	44	–0.2948	0.1441	0.4390	–845.3	1623.4
5	9904	17	46	–0.3551	0.1170	0.4721	–846.5	1675.9
6	6446	21	53	–0.3547	0.1196	0.4743	–1098.3	2395.5
7	5858	23	52	–0.3549	0.1184	0.4734	–924.6	1954.8
8	227107	13	52	–0.3508	0.1040	0.4548	–1383.5	2279.2
9	13327	12	54	–0.3513	0.1033	0.4546	–1496.2	2525.9
10	538883	26	51	–0.3589	0.0969	0.4558	–998.2	2129.5
11	5281034	21	56	–0.3360	0.1102	0.4462	–1038.4	2285.3

CID = PubChem compound identification number

Conv. cycles = an mpqc software feature allowing tracking the geometrical optimization process

Mol. no. atoms = number of atoms

HOMO = energy of highest occupied molecular orbital (“nucleophilicity”)

LUMO = energy of lowest unoccupied molecular orbital (“electrophilicity”)

GAP = HOMO – LUMO (the difference between these two energies; photon transitions occurs at this level of energy)

SCF = Self Consistent Field (energy) – SCF is the level of theory (mathematical model) in calculation

Nucl_Rep = nuclear repulsion energy

Table 3. No significant difference between correlation coefficients was identified when the Steiger’s *Z* test was applied (all *p* values higher than 0.08).

The analysis of the results presented in Table 3 revealed the following:

– No differences could be observed between eqs. (3) and (4) in terms of: determination coefficient, standard error of estimate, difference between conventional and cross validation leave-one-out determination coefficient, standard error of predicted, mean absolute percentage error, as well as two Akaike’s information criterion (AIC_r^2 and AIC_r^2).

– According to the criteria presented in the Experimental section, the model obtained when 30 optimization cycles were imposed is the most accurate MLR model. Thus, the model presented by eq. (3) accomplished 15 out of 18 criteria (83%). At a significant distance, the model presented by eq. (4) is classified with 8 out of 15 criteria, followed by the model presented by eq. (1) (2 out of 18 criteria). The model presented by eq. (2) did not accomplish any criterion.

The most accurate MLR model proved the one obtained using 30 geometry optimization cycles, indicating that the most accurate model did not need the optimized geometry of all compounds in the dataset. Further analyses are needed to identify if this property is general valid regardless the number of molecules in the dataset.

The improvements of the molecular geometry with optimization cycles sustain the efficiency of the SAPF descriptors to relate the compounds structure with their properties.

The decrease of the adapted pool of SAPF descriptors according with the number of cycles used for geometry optimization of compounds revealed that specificity/selectivity of the SAPF occurs when the geometry is near full or fully optimized.

4 Conclusion

Our study revealed an improvement in determination of the structure–hR_F (separation property) relationship when the semi-empirical MMFF94 model was further refined at HF-6-31G* (ab-initio) level of theory, suggesting the suitability of this approach in molecular design of compounds subject to chromatographic separation as alternative to PubChem deposited 3D structures of the compounds.

Regardless the number of cycles used to optimize the compounds geometry with semi-empirical MMFF94 model further refined at HF-6-31G* (ab-initio) level of theory, the most accurate models proved that the additive effect of a steric factor with a polarity factor best relates the separation of the investigated compounds with their structural information.

Acknowledgments

The study was supported by European Social Fund, Human Resources Development Operational Program, project number 89/1.5/62371 through a fellowship for L. Jäntschi.

Table 3

Parameters for assessment the SAPF-MLR models.

Parameter	Case study 1 eq. (1)	Case study 2 eq. (2)	Case study 3 eq. (3)	Case study 4 eq. (4)
Conventional determination coefficient	0.9772	0.9798	0.9852	0.9852
Standard error of estimate	3.52	3.32	2.84	2.84
F-value (*p*-value)	171	194	266	266
	($2.71 \cdot 10^{-7}$)	($1.67 \cdot 10^{-7}$)	($4.80 \cdot 10^{-8}$)	($4.82 \cdot 10^{-8}$)
p-Value of the intercept coefficient	$1.23 \cdot 10^{-9}$	$5.15 \cdot 10^{-7}$	$1.49 \cdot 10^{-7}$	$1.50 \cdot 10^{-7}$
p-Value of the 1st SAPF descriptor coefficient	$1.07 \cdot 10^{-7}$	$6.42 \cdot 10^{-8}$	$1.84 \cdot 10^{-8}$	$1.85 \cdot 10^{-8}$
p-Value of the 2nd SAPF descriptor coefficient	$2.30 \cdot 10^{-6}$	$3.88 \cdot 10^{-7}$	$1.13 \cdot 10^{-7}$	$1.13 \cdot 10^{-7}$
Leave-one-out determination coefficient	0.9606	0.9610	0.9715	0.9715
Standard error of predict	4.63	4.62	3.95	3.95
F-value (*p*-value) in leave-one-out	97	98	135	135
	($2.41 \cdot 10^{-6}$)	($2.39 \cdot 10^{-6}$)	($6.79 \cdot 10^{-7}$)	($6.82 \cdot 10^{-7}$)
Mean absolute error (MAE)	2.3752	2.1818	1.9517	1.9526
Mean absolute percentage error (MAPE)	0.0490	0.0486	0.0446	0.0446
Standard error of prediction (SEP)	3.1507	3.1507	2.5378	2.5393
Relative error of prediction (REP%)	5.3261	5.0136	4.2901	4.2926
Corrected AIC[a] (AIC_c)	33.63	32.30	28.87	28.88
AIC based on the r^2(AIC_r^2)	–0.18	–0.30	–0.61	–0.61
McQuarrie, Tsai corrected AIC (AIC_u)	4.85	4.73	4.42	4.42
Bayesian information criterion (BIC)	34.90	33.57	30.14	30.15
Amemiya prediction criterion (APC)	15.79	13.99	10.25	10.26
Hannan–Quinn criterion (HQC)	29.45	28.12	24.69	24.70
Akaike weight for AIC_c	$1.47 \cdot 10^{-7}$	$2.87 \cdot 10^{-7}$	$1.59 \cdot 10^{-6}$	1.00
Akaike weight for AIC_r^2	0.22	0.23	0.27	0.27
Akaike weight for AIC_u	0.22	0.23	0.27	0.27
Kubinyi function (FIT)	14.99	16.96	23.29	23.27

[a]Akaike information criterion

References

[1] *P. Vanhee, A.M. van der Sloot, E. Verschueren, L. Serrano, F. Rousseau, J. Schymkowitz,* Trends Biotechnol. **29** (2011) 231–239.

[2] *E.B. Fauman, B.K. Rai, E.S. Huang,* Curr. Opin. Chem. Biol. **15** (2011) 463–468.

[3] *J. Hu, X. Zhang, Z. Wang,* Int. J. Mol. Sci. **11** (2010) 1020–1047.

[4] *K. Namboori, K.S. Ramachandran, D. Gopakumar,* Computational Chemistry, Molecular Modeling: Principles, Applications, Berlin, Springer, 2008.

[5] *M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart,* J. Am. Chem. Soc. **107** (1985) 3902–3909.

[6] *J.R. Maple, M.-J. Hwang, T.P. Stockfisch, U. Dinur, M. Waldman, C.S. Ewig, A.T. Hagler,* J. Comput. Chem. **15** (1994) 162–182.

[7] *G. Del Re, B. Pullman, T. Yonezawa,* Biochim. Biophys. Acta **75** (1963) 153–182.

[8] *J.J.P. Stewart,* J. Comput. Chem. **10** (1989) 209–220.

[9] *J.J.P. Stewart,* J. Comput. Chem. **10** (1989) 221–264.

[10] *J.J.P. Stewart,* J. Mol. Model. **13** (2007) 1173–1213.

[11] *G.B. Rocha, R.O. Freire, A.M. Simas, J.J.P. Stewart,* J. Comput. Chem. **27** (2006) 1101–1111.

[12] *J. Gasteiger, M. Marsili,* Tetrahedron **36** (1980) 3219–3228.

[13] *J. Gasteiger, M. Marsili,* Tetrahedron Lett. **19** (1978) 3181–3184.

[14] *E. Hückel,* Z. Phys. **70** (1931) 204–286.

[15] *E. Hückel,* Z. Phys. **72** (1931) 310–337.

[16] *E. Hückel,* Z. Phys. **76** (1932) 628–648.

[17] *E. Hückel,* Z. Phys. **83** (1933) 632–668.

[18] *H. Berthod, C. Giessner-Prettre, A. Pullman,* Theor. Chim. Acta **8** (1967) 212–222.

[19] *W.L. Jorgensen, J. Tirado-Rives,* J. Am. Chem. Soc. **110** (1988) 1657–1666.

[20] *S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P.K. Weiner,* J. Am. Chem. Soc. **106** (1984) 765–784.

[21] *T.A. Halgren,* J. Comput. Chem. **17** (1996) 520–552.

[22] *T.A. Halgren,* J. Comput. Chem. **17** (1996) 553–586.

[23] *T.A. Halgren,* J. Comput. Chem. **17** (1996) 587–615.

[24] *T.A. Halgren,* J. Comput. Chem. **17** (1996) 616–641.

[25] *T.A. Halgren,* J. Comput. Chem. **17** (1996) 490–519.

[26] *Å. Rinnan, N.J. Christensen, S.B. Engelsen,* J. Comput.-Aided Mol. Des. **24** (2010) 17–22.

[27] *K.-C. Tsai, Y.-C. Chen, N.-W. Hsiao, C.-L. Wang, C.-L. Lin, Y.-C. Lee, M. Li, B. Wang,* Eur. J. Med. Chem. **45** (2010) 1544–1551.

[28] *A. Jakalian, B.L. Bush, D.B. Jack, C.I. Bayly,* J. Comput. Chem. **21** (2000) 132–146.

[29] *A. Jakalian, D.B. Jack, C.I. Bayly,* J. Comput. Chem. **23** (2002) 1623–1641.

[30] *M. Leki, F. Kora, M. Šober, A. Marjanovi,* Acta Chim. Slov. **54** (2007) 88–91.

[31] OEShape, version 1.8.0. OpenEye Scientific Software Inc., Santa Fe, NM, USA, 2010.

[32] OEOmega, version 2.4. OpenEye Scientific Software Inc., Santa Fe, NM, USA, 2009.

[33] *T.A. Halgren,* J. Comput. Chem. **20** (1999) 720–729.

[34] *J.P. Kenny, C.L. Janssen, E.F. Valeev, T.L. Windus,* J. Comput. Chem. **29** (2008) 562–577.

[35] *L. Jäntschi,* Genetic Algorithms, their Application. PhD Thesis in Horticulture, completed at University of Agricultural Sciences, Veterinary Medicine Cluj-Napoca, (PhD Advisor: Prof. Dr. R.E. Sestraş), 2010.

[36] *R.E. Sestraş, L. Jäntschi, S.D. Bolboacă,* Int. J. Mol. Sci. **13** (2012) 5207–5229.

[37] *A. Golbraikh, A. Tropsha,* J. Mol. Graph. Model. **20** (2002) 269–276.

[38] *S.D. Bolboacă, L. Jäntschi,* Environ. Chem. Lett. **6** (2008) 175–181.

[39] *J.H. Steiger,* Psychol. Bull. **87** (1980) 245–251.

[40] *H. Akaike,* Ann. Inst. Stat. Math. **21** (1969) 243–247.

[41] *C.M. Hurvich, C.-L. Tsai,* Biometrika **76** (1989) 297–307.

[42] *A.D.R. McQuarrie, C.-L. Tsai,* Regression and Time Series Model Selection in Small Samples, World Scientific Publishing Co, Singapore, 1988.

[43] *G. Schwarz,* Ann. Stat. **6** (1978) 461–464.

[44] *T. Amemiya,* J. Econ. Lit. **19** (1981) 1483–1536.

[45] *E.J. Hannan, B.G. Quinn,* J. R. Stat. Soc. Ser. B **41** (1979) 190–195.

[46] S*.T. Buckland, K.P. Burnham, N.H. Augustin,* Biometrics **53** (1997) 603–618.

[47] *H. Kubinyi,* QSAR Comb. Sci. **3** (1994) 393–401.

[48] *H. Kubinyi,* QSAR Comb. Sci. **13** (1994) 285–294.

Ms received: December 28, 2011
Accepted: July 4, 2012