Eur. J. Med. Chem. 36 (2001) 829–836

PII: S0223-5234(01)01276-4/FLA

Original article

A comparative study of quantitative structure activity relationship methods based on antitumor diarylsulfonylureas

Hea-Young Park Choo[a,*], Jae-Su Lim[a], Yurim Kam[a], Su Yeon Kim[b], Jeewoo Lee[b]

[a]School of Pharmacy, Ewha Women's University, Seoul 120-750, Republic of Korea
[b]School of Pharmacy, Seoul National University, Seoul, Republic of Korea

Received 6 July 2001; accepted 18 September 2001

Abstract – A series of 28 diarylsulfonylureas with antitumor activity was subjected to a three-dimensional quantitative activity relationship (3D-QSAR) study. Three different QSAR methods, comparative molecular field analysis (CoMFA), hologram QSAR (HQSAR) and comparative molecular similarity indices analysis (CoMSIA), were compared in terms of their potential for predictability. All three QSAR-based models had good predictability and yielded q^2 values 0.74, 0.63 and 0.72, respectively. The CoMFA model provided the highest q^2 and r^2 values, which implied the significance of correlation of steric and electrostatic fields with biological activities. The number of components was 3–4 for all three QSAR methods. The quality of HQSAR or CoMSIA was slightly lower than that of CoNFA in terms of q^2 and r^2 values. HQSAR does not require the generation of a three-dimensional structure of molecules and CoMSIA does not require molecular superposition, therefore they are faster than CoMFA in data processing. © 2001 Éditions scientifiques et médicales Elsevier SAS

antitumor activity / diarylsulfonylureas / biological activities

1. Introduction

The mathematical relationships between physico–chemical properties of molecule and the biological activities as well as the subsequent prediction for novel compounds in series have been widely investigated. The successful development and application of CoMFA make it one of most powerful tools in the drug design. CoMFA methodology is based on the assumption that changes in the biological activity correlate with changes in the steric and electrostatic fields of molecules [1, 2]. However, CoMFA requires the functionally active conformation of the molecules and molecular superposition.

In HQSAR, which recently introduced by Tripos Inc., each molecule in the data set was divided into structural fragments [3, 4]. The structural fragments are counted in the bins of a fixed length array to form a molecular hologram. The correlation of bin occupancies of the molecular hologram with variation in activity data is calculated to produce a QSAR model.

Another recently introduced method, CoMSIA uses similarity indices calculated in the space [5–8]. The similarity indices are enumerated for each aligned molecule at regularly spaced grid points. Instead of direct measure of similarity between all mutual pairs of molecule, indirect evaluation of similarity of each molecule in the data set with a common probe atom is calculated [9, 10]. Then the linear regression equation of similarity with biological activities is derived.

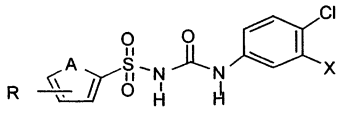
In this paper, the three QSAR methods CoMFA, HQSAR, and CoMSIA are evaluated for their predictability and performance using a series of cytotoxic diarylsulfonylureas.

Diarylsulfonylureas exhibit broad-spectrum antitumor activity against solid tumors such as in the colon, the ovary, and lung although the series had originally been synthesized as potential herbicides [11]. The mechanism of antitumor diarylsulfonylureas is unkown, but they were different from other anticancer

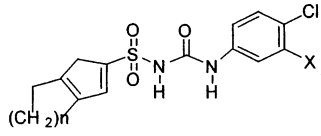
* Correspondence and reprints
E-mail address: hypark@mm.ewha.ac.kr (H.-Y. Park Choo).

agents being used and had no effect on RNA, DNA or protein synthesis or the cross resistance in multidrug resistant cell lines [12]. The lead compound *N*-(5-indanylsulfonyl)-*N'*-(4-chlorophenyl)urea (Sulofenur) has been dropped in the phase II clinical trials mainly due to anemia and methemoglobinemia as well as poor effectiveness [13]. Extensive studies have attempted at improving the antitumor activity and at decreasing the toxicity of Sulofenur [14–19]. In structure activity relationship study it has been reported that the modification of sulfonylurea linkage to sulfonylthiourea, sulfonylcarbamate, sulfonylthiocarbamate resulted in abolition of activity against the 6C3HED lymposarcoma. The effect of size, electronic character and position of substituent on each aromatic ring to the antitumor activity has been reported

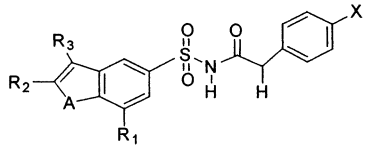
Table I. The structure and activity of diarylsulfonylurea analogs [15].



Entry No.	R	A	X	Percent inhibition
1	5-CH_3	S	Cl	24
2	5-CH_2CH_3	S	H	69
3	5-$(CH_2)_2CH_3$	S	H	38
4	5-$CH(CH_3)_2$	S	H	42
5	5-O-$(CH_2)_2CH_3$	S	H	44
6	5-$C(CH_3)_3$	S	H	7
7	5-CH_2CH_3	O	H	57
8	3-CH_3	S	H	17
9	3-CH_2CH_3	S	H	21
10	4-OCH_3	S	H	36
11	4,5-Br_2	S	H	25
12	4,5-$(CH_3)_2$	S	H	42
13	5-CH_2CH_3-4-CH_3	S	H	26

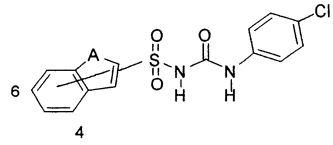


Entry	n	X	Percent inhibition
14	1	H	36
15	2	H	0
16	1	H	16

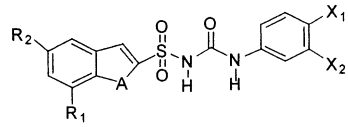


Entry	A	R_1	R_2	R_3	X	Percent inhibition
17	O	H	H	H	Cl	100
18	O	H	H	H	CH_3	84
19	O	CH_3	H	H	Cl	89
20	O	H	CH_3	H	Cl	57

Table 1. (*Continued*)



Entry	A	Substitution	Percent inhibition
21	O	4	84
22	O	6	100



Entry	A	R_1	R_2	X_1	X_2	Percent inhibition
23	O	H	H	Cl	Cl	28
24	O	CH_3	H	Cl	H	21
25	O	Br	H	Cl	H	18
26	O	H	OCH_3	Cl	H	25
27	O	H	CH_3	Cl	H	6
28	S	H	CH_3	Cl	H	10

[14]. For the ring bound to sufonyl group of sulfonylurea, phenyl ring shows good activity and both electron donating and neutral substituents on the phenyl ring are compatible with good activity, but substitutions by an electron withdrawing group are associated with poor activity regardless of position of substituents. The ring attached to the urea nitrogen of sulfonylurea, required a small neutral lipophilic group in the 4-position for good activity. The bioisosterer of aryl group such as benzothiophene, benzofuran have also been studied by Eli Lilly scientists [15]. However, no three-dimensional quantitative structure activity relationship analysis of benzothiophene and benzofuran analogs was reported and the results of the series of these analogs reported by Mohamadi et al. was employed in the evaluation of three different QSAR methods in this study.

2. Methods

2.1. Data set for analysis

A set of the antitumor activity data on human lung carcinoma (A549) cell line reported by Mohamadi et al. was used in this calculation [15]. *Table I* represents the structure, antitumor activities, and lipophilicity of 28 compounds employed. The activity is percent inhibition of tumor cell growth obtained after the mice dosed p.o. at 150 mg kg^{-1} daily for 8 days (*table I*).

2.2. Computational methods

All molecular modeling and statistical analyses were performed using SYBYL 6.5 molecular modeling software (Tripos Inc.) and Silicon Graphics Indy workstation (IRIX 6.2). The 2D structure of each compound was built using SYBYL Build program with the default SYBYL settings. The 2D structure was converted to 3D structure using CONCORD 4.0 program. The structural energy minimization was performed using the SYBYL energy minimizer (Tripos Force Field) and Gasteiger–Huckel charge, with a 0.005 kcal mol^{-1} Å energy gradient convergence criterion. Low energy conformation was searched by geometry optimization after rotating every 30° of single bond from 1 to 330° of tortional angle. All of the structures generated were aligned into a lattice box by fitting with sulfonylurea group as a common structure.

2.3. Calculation of CoMFA descriptors

Conventional CoMFA was performed with the QSAR option of SYBYL. The steric and electrostatic field energies were calculated using sp^3 carbon probe atoms with +1 charge. Maximum energy cutoff for steric and electrostatic energies was 30 kcal mol^{-1}. The CoMFA

Table II. Statistical parameters for the three QSAR-based models.

Model	q^2	S_{press}	r^2	rms error	No. of component
CoMFA	0.741	–	0.980	4.875	4
HQSAR	0.625	19.382	0.930	8.347	4
CoMSIA	0.718	17.439	0.871	11.129	3

Table III. Predictive results obtained with COMFA analysis.

Entry no.	CoMFA	Measured activity (% inhibitoin)	Predictied activity	Residual
1	118	24	25.03	−1.03
2	130	69	44.71	24.29
3	136	38	38.74	−0.74
4	138	42	40.26	1.74
5	138	44	44.13	−0.13
6	142	7	11.32	−4.32
7	124	57	33.82	23.18
8	120	17	20.71	−3.71
9	128	21	17.49	3.51
10	118	36	24.13	11.87
11	120	25	19.88	5.12
12	120	42	20.34	21.66
13	132	26	26.13	−0.13
14	128	36	34.32	1.68
15	140	0	6.9	−6.9
16	126	16	13.85	2.15
17	116	100	87.69	12.31
18	122	84	87.41	−3.41
19	124	89	92.25	−3.25
20	130	57	58.52	−1.52
21	116	84	82.43	1.57
22	118	100	104.35	−4.35
23	130	28	27.79	0.21
24	136	21	23.29	−2.29
25	130	18	22.64	−4.64
26	140	25	13.58	11.42
27	134	6	21.48	−15.48
28	136	10	4.94	5.06

grid spacing was 2.0 Å in all three dimensions within the defined region. The partial least squares (PLS) method was used for fitting the 3D structural features and their biological activities. The optimum number of components in the final PLS model was determined by the q^2 value, obtained from the leave-one-out cross validation technique.

2.4. Calculation of CoMSIA descriptors

The CoMSIA of the QSAR module of SYBYL was used for the analysis. Similarity indices between a compound and a probe atom were calculated. The common probe atom with charge +1, radius 1 Å, and hydrophobicity +1 was placed at the intersections of a regularly spaced lattice. The attenuation factor (a) was set 0.3. To determine the similarity, the mutual distance between probe atom and the atoms of the molecules of the data set was considered. The equation used to calculate the similarity indices is as follows:

$$A^{q}_{F,K,(j)} = -\sum W_{\text{probe},k}\, W_{ik} \mathrm{e}^{-\alpha\gamma^2_{iq}}$$

A is the similarity index at grid point q, summed over all atoms i of the molecule j. $W_{\text{probe},k}$ is the probe atom, W_{ik} is the value of the physicochemical property k of atom i. The distance between the probe atom at grid point q and atom i of the test molecule is represented by

r_{iq}, and α is the attenuation factor, larger values of which result in steeper Gaussian function curves and a strong attenuation of the distance-dependent effects of molecular similarity.

2.5. Calculation of HQSAR descriptors

For HQSAR, the HQSAR module of SYBYL was used. All the sub-structural fragments in the size of 4–7 atoms were generated for each molecule. The SLN (SYBYL Line Notation) for each fragment generated was mapped to an unique integer. Each integer was used to select a bin in an integer array of predetermined hologram length, which was 12 prime numbers ranging from 53 to 401. The hologram length defined the dimensionality of the descriptor space. The bin occupancies were the HQSAR descriptor variables that counted molecular structural fragments in each bin. The hydrogen atoms were included in the generation of the molecular hologram.

3. Results and discussion

3.1. Comparison of different QSAR methods

The summary of the statistical results of three QSAR analyses is listed in *table II*. In this table, n represents the number of the compound, s the standard error of estimate, r^2 the correlation coefficient and q^2 the cross validation, respectively. The cross validation value q^2 and conventional correlation coefficient r^2 are normally accepted as statistical measures for the quality of QSAR models. All three QSAR models exceeded the $q^2>0.5$ criterion which indicated the stability of model and reasonable predictability [20]. The CoMFA and HQSAR showed higher r^2 value than 0.9 criterion but the r^2 value for CoMSIA was slightly lower than the criterion. The quality of CoMFA was better than the CoMSIA or HQSAR in terms of q^2 and r^2 values. The number of components was 3–4 for all three QSAR methods. The fewest number of components was reason-

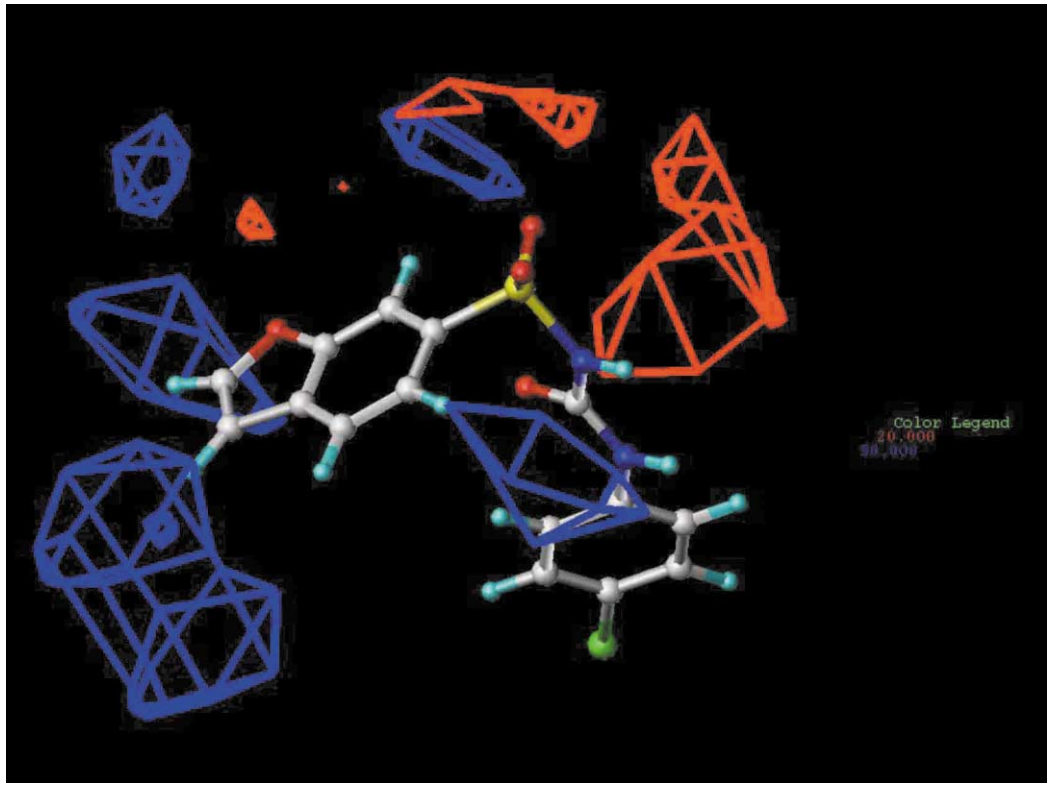


Figure 1. CoMFA contour map of electric field. Positive charge favored areas are represented in blue and negative charge favored areas are represented in red.

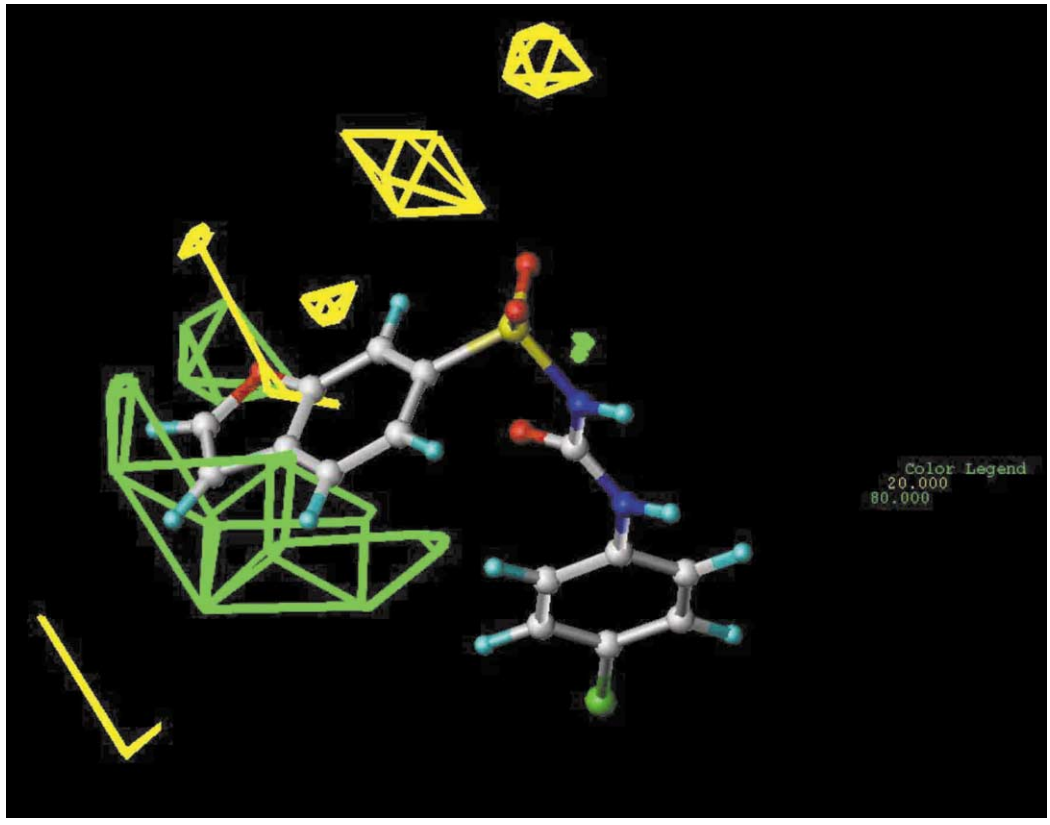


Figure 2. CoMFA contour map of steric field. Sterically favored areas are represented in green and sterically disfavored areas are represented in yellow.

able to avoid overfitting of data points since the value of r^2 generally increased as more number of components were included in the model.

3.2. CoMFA

When the CoMFA was set as descriptor and the inhibition (%) of tumor growth as a dependent column, CoMFA model provided the highest q^2 and r^2 values, which accounted for good predictability with at least 98% of the variation in biological activity. This indicates that the method is considerably reliable to predict the antitumor activities of sulfonylurea analogs.

The contributions of steric and electrostatic fields were 73.6 and 26.4%, respectively. Steric interaction of molecule with receptor could be an important factor for antitumor activity. The actual and predicted activities of the training set are reported in *table III* and plotted in *figure 1*.

The major steric and electrostatic features of the 3D QSAR derived from CoMFA study were illustrated in *figure 2* as three-dimensional solid surfaces. In CoMFA contour map, steric bulk that enhanced (green) or decreased (yellow) antitumor activity in this series of compounds are indicated. Also the location of electronegative character that enhanced/or decreased antitumor activity are depicted in red/or blue.

The blue contour located around aromatic ring bound to sulfonyl group indicates that the aromatic rings should be electron rich for its optimal activity. It is consistent with the observation that the electron donating and neutral group substitutions on the phenyl ring are compatible with good activity, but substitutions by electron withdrawing group are associated with poor activity regardless of position of substituents [14]. The red contour on phenyl ring bonded to urea nitrogen of sulfonylurea, indicates an electronegative group in the two or three-position for good activity. The substitution of aromatic group on furan was benificial for antitumor activity as indicated by the green contour.

In this series, CoMFA yielded the best QSAR models in term of self-consistency and ability to predict biologi-

Table IV. Predictive results obtained from HQSAR and CoMSIA.

Entry no.	Measured activity (% inhibition)	Predicted activity by HQSAR	Residual by HQSAR	Predicted activity by CoMSIA	Residual by CoMSIA
1	24	23.25	0.75	29.25	−5.25
2	69	60.85	8.15	65.41	3.59
3	38	41.24	−3.24	27.97	10.03
4	42	43.32	−1.32	30.18	11.82
5	44	58.95	−14.95	56.81	−12.81
6	7	5.72	1.28	20.73	−13.73
7	57	50.52	6.48	45.3	11.7
8	17	19.05	−2.05	37.92	−20.92
9	21	31.73	−10.73	23.72	−2.72
10	36	32.75	3.25	36.01	−0.01
11	25	34.97	−9.97	20.64	4.36
12	42	29.25	12.75	28.34	13.66
13	26	32.84	−6.84	21.6	4.4
14	36	20.14	15.86	17.77	18.23
15	0	0.46	−0.46	5.15	−5.15
16	16	26.25	−10.25	24	−8
17	100	92.25	7.75	86.07	13.93
18	84	78.01	5.99	94.51	−10.51
19	89	90.49	−1.49	79.45	9.55
20	57	69.03	−12.03	75.04	−18.04
21	84	83.18	0.82	86.58	−2.58
22	100	94.73	5.27	90.91	9.09
23	28	24.67	3.33	21.16	6.84
24	21	26.5	−5.5	26.84	−5.84
25	18	26.13	−8.13	15.76	2.24
26	25	18.9	6.1	24.94	0.06
27	6	0.28	5.72	20.58	−14.58
28	10	6.54	3.46	9.38	0.62

cal activity. Although CoMFA models are of high quality and can give indication of structural differences responsible for differing biological activities, they can be time-consuming to construct because the models require determination of suitable molecular conformations and a structural alignment of the molecules.

3.3. CoMSIA

The predicted activities of molecules by CoMSIA were fairly well correlated with the actual activities except for a few compounds as shown in *table III*. High cross validation value q^2 (0.72) was obtained as shown in *table II*. From this result the usefulness of similarity as a descriptor was demonstrated but the error between predicted and actual activity was biggest among three QSAR methods compared (*table IV*). This result can be also noticed from the highest rms error value in *table II*. However CoMSIA could find its value in not requiring molecular superposition and very fast data processing.

3.4. HQSAR

Table V shows a summary of the results of the HQSAR calculations. For a given hologram length, the

Table V. Result of the HQSAR calculations.

Hologram length	q^2	No. of components
53	0.556	4
59	0.467	3
61	0.410	3
71	0.538	5
83	0.385	3
97	0.319	4
151	0.542	4
199	0.408	4
257	0.500	4
307	**0.625**	**4**
353	0.567	4
401	0.614	4

The optimal HQSAR model, shown in bold typeface, is a four-component PLS model based on the molecular hologram generated using a length of 307.

q^2 value of the QSAR model with the optimal number of components was given. The 12 hologram length values listed are the default prime numbers employed by the HQSAR module in SYBYL. The optimal HQSAR method was generated using a hologram length of 307 with four components. This implies that each component consisting of some linear combinations of the 307 descriptors encoded in the hologram had good q^2 value with smallest standard deviation. HQSR method yielded high r^2 value (0.93) and reasonably high cross validation value q^2 (0.625) for the fitting of all 28 compounds in training set. The quality of the HQSAR model was comparable with that of CoMFA. Even though the CoMFA model showed the highest predictability in most of compounds in this series, some compounds such as **2**, **7**, **12** showed high error (over 20) between predicted and actual activity. The highest error in HQSAR model was 15.86 and the HQSAR was found to be reliable to predict the biological activity.

HQSAR model included only elemental and bond type information, it was 2D in nature. HQSAR employed counts of sub-structural molecular fragments as descriptor and no requirement for 3D structures or molecular alignment, it was both fast and reproducible.

4. Conclusion

The 3D QSAR analyses of a set of diarylsulfonylureas with antitumor activity on human lung cancer cell (A529) were compared using three different methods. All three QSAR models gave good statistical results in terms of q^2 and r^2 values. The CoMFA model provided the most significant correlation of steric and electrostatic fields with the biological activities.

The performances of the new QSAR methods, HQSAR and CoMSIA were found to be comparable in statistical terms to that of COMFA. In this study HQSAR showed its effectiveness and predictability on compounds with various substituents. HQSAR did not require the generation of a three-dimensional structures, and it was fast and reproducible. The statistical results obtained from CoMSIA were slightly lower than CoMFA but CoMSIA did not require molecular superposition and very fast in data processing.

Acknowledgements

This work was supported by a grant from the Korea Research Foundation (00-B-WB-06-A-07).

References

[1] Cramer III R.D., Patterson D.E., Bunce J.D., J. Am. Chem. Soc. 110 (1988) 5959–5967.

[2] Cramer III R.D., DePriest S.A., Patterson D.E., Hecht P., The developing practice of comparative molecular field analysis, in: Kubiny H. (Ed.), 3DQSAR in Drug Design. Theory Methods and Applications, ESCOM Science Publishers, Leiden, The Netherlands, 1993, pp. 443–485.

[3] Tong W., Lowis D.R., Perkins R., Chen Y., Welsh W.J., Goddette D.W., Heritage T.W., Sheehan D.M., J. Chem. Inf. Comput. Sci. 38 (1998) 669–677.

[4] So S.S., Karplus M., J. Comput. Aided Mol. Des. 13 (1999) 243–258.

[5] Klebe G., Abraham U., Mietzner T., J. Med. Chem. 37 (1994) 4130–4146.

[6] Dean P.M., Molecular similarity, in: Kubunyi H. (Ed.), 3D QSAR in Drug Desugn: Theory, Methods and Applications, ESCOM Science Publishers, Leiden, The Netherlands, 1993, pp. 150–172.

[7] Richards W.G., Molecular similarity and dissimilarity, in: Pullman A., Jortner J., Pullman B. (Eds.), Modeling of Biomolecular Structures And Mechanisms, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995, pp. 365–369.

[8] Kubinyi H., A general view on similarity and QSAR studies. In: van de Waterbeemd H., Testa B., Folkers G. (Eds.), Computer-Aided Lead Finding and Optimization: Current Tools for Medicinal Chemistry, VHCA, Basel, and Wiley-VCH, Weinheim, 1997, pp. 7–28.

[9] So S.S., Karplus M., J. Med. Chem. 40 (1997) 4347–4359.

[10] So S.S., Karplus M., J. Med. Chem. 40 (1997) 4360–4371.

[11] Howbert J.J., Drug Future 16 (1991) 517–520.

[12] Houghton P.J., Houghton J.A., Investig. New Drugs 14 (1996) 271–280.

[13] Ehjardt W.J., J. Drug Metab. Dispos. 19 (1991) 370–375.

[14] Howbert J.J., Grossman C.S., Crowell T.A., Rieder B.J., Harper R.W., Kramer K.E., Tao E.V., Aikins J., Poore G.A., Rinzel S.M., Grindey G.B., Shaw W.N., Todd G.C., J. Med. Chem. 33 (1990) 2393–2407.

[15] Mohamadi F., Spees M.M., Grindey G.B., Bioorg. Med. Chem. Lett. 2 (1992) 987–992.

[16] Mohamadi F., Spees M.M., Grindey G.B., J. Med. Chem. 35 (1992) 3012–3016.

[17] Toth J.E., Grindey G.B., Ehlhardt W.J., Ray J.E., Boder G.B., Bewley J.R., Klingerman K.K., Gates S.B., Rinzel S.M., Schultz R.M., Weir L.C., Worzalla J.F., J. Med. Chem. 40 (1997) 1018–1025.

[18] Jung S.H., Song J.S., Lee H.S., Choi S.U., Lee C.O., Arch. Pharm. Res. 19 (1996) 570–580.

[19] Jung S.H., Kwak S.J., Arch. Pharm. Res. 20 (1997) 283–287.

[20] Cho S.J., Tropsha A., J. Med. Chem. 38 (1995) 1060–1066.