J Math Chem (2008) 43:1549–1559
DOI 10.1007/s10910-007-9283-1

ORIGINAL PAPER

Approximate similarity and QSAR in the study of spirosuccinimide type aldose reductase inhibitors

Irene Luque Ruiz · Manuel Urbano-Cuadrado · Miguel Ángel Gómez-Nieto

Received: 9 March 2007 / Accepted: 12 July 2007 / Published online: 2 September 2007

Abstract A new application using approximate similarity (AS) measurements to the study of chemical properties of drugs is presented in this paper. A quantitative structure—activity relationship (QSAR) model for predicting spirosuccinimide fused tetrahydropyrrolo[1,2-a]pyrazine (SPPP) compounds activity as inhibitors of the aldose reductase (AR) enzyme was developed. This enzyme is involved in the transformation of glucose into sorbitol, which causes several diseases related to diabetes mellitus. AS matrices were built based on isomorphic and nonisomorphic data fusion and they were employed as representation space of SPPP data set for the development of QSAR model. For this purpose, isomorphism among all the pairs of molecules of the studied data set was extracted and Wiener and HyperWiener descriptors were used for describing the isomorphic and nonisomorphic subgraphs. Full cross validation ($Q^2 = 0.92$, Standard Error in Cross Validation = 0.10) was the strategy employed to build and validate accurate predictive similarity spaces.

Keywords Approximate similarity · Nonisomorphic fragments · Topological descriptors · QSAR · Aldose reductase inhibitors

I. Luque Ruiz (✉) · M. Á. Gómez-Nieto
Department of Computing and Numerical Analysis, University of Córdoba, Campus Universitario de Rabanales, Albert Einstein Building, 14071 Córdoba, Spain
e-mail: ma1lurui@uco.es

M. Á. Gómez-Nieto
e-mail: mangel@uco.es

M. Urbano-Cuadrado
Institute of Chemical Research of Catalonia ICIQ, Avinguda Països Catalans, 16, 43007 Tarragona, Spain
e-mail: murbano@cnio.es

1 Introduction

Different studies have proven that the spirosuccinimide fused tetrahydropyrrolo[1,2-a]pyrazine (SPPP) derivatives are potent inhibitors of the aldose reductase enzyme (AR), which is an enzyme involved in the first step of the polyol pathway reducing the glucose in blood into sorbitol [1,2]. The excess of intracellular sorbitol is one of the reasons for retinopathy, nephropathy, cataracts, etc., related to the diabetes mellitus disease [3–6]. Therefore, the design of new AR inhibitors and the study of their capacity are nowadays work topics focused on combating the hyperglycemia symptoms observed in the diabetes mellitus. This interest is also justified by the fact of obtaining a wide effect spectrum for the "in vivo" tests inhibitory capacity [7–17]. Therefore, the use of new drug-design methodologies showing minimal environment damages is pursued.

The establishment and validation of mathematical equations which match pieces of structural information and compound behavior is pursued in several scientific fields and recommended by several international organizations [18]. Thus, quantitative structure-activity relationship (QSAR) approaches have attracted the attention of pharmaceutical industries due to the modeling and predictive ability achieved by QSAR models, which permit to assist the drug design via-computer, to develop screening chemical databases methods, to predict, within a confidence interval, the behavior of new compounds, etc. [19,20]

Regarding SPPP compounds, QSAR models have been recently proposed using 3D optimized structures [21,22]. Due to the enantioselectivity shown by the AR enzyme, different models were built using R-descriptors, S-descriptors and the arithmetic mean of the descriptors resulting from each SPPP enantiomer structure (racemic descriptors). These models achieved good correlations for the inhibition activity by means of five descriptors: (a) partial negative surface area ($PNSA_1$), (b) the noncommon overlap steric volume (V_{NCOS}), (c) the z-component dipole moment (μ_Z), (d) the aqueous desolvation free energy of the molecule (F_{H_2O}), and (e) the length of the molecule in the z-direction. These descriptors included conformational, electronic, spatial, structural, thermodynamic and molecular shape analysis calculations. With these descriptors, Won et al. [21,22], proposed their SPPP approaches using genetic algorithms—linear and quadratic terms of the descriptors were considered in the independent variable set.

As Fig. 1 shows, the SPPP compounds studied in reference 22 present a common substructure (compound **1**) which practically corresponds to the 90–98% of the whole molecule. In spite of this structural similarity, the nature and position of the fragments that do not belong to the common substructure are key for the inhibition parameter ($pIC_{50} = -logIC_{50}$) variation—from 0.201 of compound **1** to 1.638 of the compound **20**.

Molecular fragments have been recently studied for developing QSAR models [23,24]. Thus, studies based on data set molecular fragmentation and compound classification as a function of the fragments characteristics have been carried out. The point was to find behavioral patterns for the different fragments regarding their properties.

Other QSAR approaches based on fragment characteristics have been developed using the approximate similarity (AS) [25–27]. Derived from chemical graphs, AS is a similarity measurement which merges, for each pair of the data set elements,

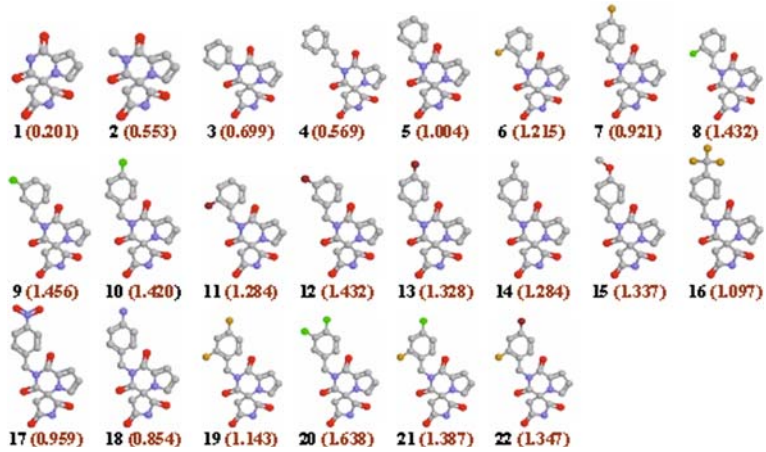

Fig. 1 Structures and inhibitory capacities of the 22 spirosuccinimide fused tetrahydropyrrolo[1,2-a] pyrazine (SPPP) compounds studied

substructure similarity and dissimilarity as follows:

$$\mathrm{AS}_{A,B} = f\left(S_{A,B}, \Gamma_{A,B}, w_{\Gamma}\right) \tag{1}$$

- Extracting the isomorphism between all pair of the data set graphs, and calculating a similarity measure based on the computed isomorphism ($S_{A,B}$).
- Using one or several topological descriptors in order to obtain distance measurements that characterize the noncommon fragments of the extracted isomorphism ($\Gamma_{A,B}$).
- Considering an adjustment factor (w_{Γ}) which weighs up the influence of the nonisomorphic fragments on the approximate similarity value.
- And combining all the above parameters in a similarity metric using a function *f()*.

Approximate similarity measurements are simple and flexible and their calculation does not require high computational resources since both isomorphism extraction and topological descriptor calculation over the isomorphic and nonisomorphic fragments are fast processes [26].

So, an N by N AS matrix can be obtained and employed as a predictive space—N is the number of elements of the data set. QSAR models based on AS matrices have been previously built for different families of compounds (biphenyls, steroids, benzodiazepines, etc.) [25–27], and high correlations were obtained.

The use of both AS matrices and partial least squares regression (PLSR) for establishing structure-activity relationships between 22 SPPP derivatives and their inhibitor capacity has been proposed. Regarding previous works, the aim was to develop QSAR approaches characterized by simplicity, fastness and robustness, thus supporting the approximate similarity concept validation. Section 2 describes the chemical

structures and their activities and, besides, the AS methodology employed for generating the structural descriptors. The statistical characterization of the models, and so, the evaluation of the predictive ability is given in Sect. 3. Finally, conclusions are exposed.

2 Materials and methods

2.1 Chemical structures and biological data

Figure 1 shows the SPPP derivatives and their inhibitor capacities (pIC_{50}). These activity values were extracted from the most recent SPPP model proposed by Won et al. [22], which, and in turn, uses the assays carried out by Negoro et al. [28] for measuring the 50% inhibition of the porcine lens AR activity—the 50% sorbitol decrease regarding to the amount accumulated in the sciatic nerve of diabetic rats.

Two dimensional structures of the SPPP derivatives built using Marvin software [29] were employed for the descriptor generation. Thus, only the type of atoms and bonds were required for developing the QSAR models, this way avoiding 3D geometry optimization methods.

2.2 Isomorphic and nonisomorphic chemical information

For the data set of compounds above described, isomorphism between all pairs of molecular graphs was extracted using an algorithm developed by the authors [30]. In this way, information of the Maximum Common Substructure (MCS) and the Non-Isomorphic Fragments (NIF) was stored in two 22 by 22 matrices, namely:

- A symmetric MCS matrix where each element (i, j) stores the MCS value of the isomorphism between the molecules i and j; the matrix elements (i, i) correspond to the structure of the molecule i.
- A nonsymmetric NIF matrix where each element (i, j) stores the value of the nonisomorphic fragments of the molecular graph i calculated considering the isomorphism extracted from the molecules i and j matching. Therefore, diagonal elements (i, i) are null.

The information of these isomorphic and nonisomorphic substructures was obtained by means of topological invariants, which describe chemical information from graph representations (isomorphic and nonisomorphic subgraphs). So, two matrices, MCS and NIF, storing the information about similarity and dissimilarity of the data set compounds were built and employed as the basis of the subsequent generation of AS values, which constitute the QSAR model input.

2.3 Approximate similarity measurements

As stated above, isomorphic and nonisomorphic data were fused in order to generate more efficient similarity spaces. This fact was achieved combining similarities of the

common substructures and distances of the noncommon subgraphs, as follows:

$$\mathrm{AS}_{A,B} = S^I_{A,B} - \left(\frac{\mathrm{NIF}_A}{TD(A)} - \frac{\mathrm{NIF}_B}{TD(B)}\right) \times \left(1 - \frac{\sqrt{[TD(\mathrm{NIF}_A) - TD(\mathrm{NIF}^P_A)]^2 + [TD(\mathrm{NIF}_B) - TD(\mathrm{NIF}^P_B)]^2}}{\sqrt{[TD(A) - TD(A^P)]^2 + [TD(B) - TD(B^P)]^2}}\right) \tag{2}$$

Expression (2) shows three components which constitute the AS measurement. The first term consists of the invariant-based similarity value $S^I_{A,B}$ obtained using one of the established similarity indexes, for instance, the cosine index computed as follows:

$$S^I_{A,B} = \frac{TD(MCS_{A,B})}{\sqrt{TD(A) \times TD(B)}} \tag{3}$$

where $TD(MCS_{A,B})$, $TD(A)$ and $TD(B)$ are the topological invariants of the maximum common substructure, and the A and B compounds, respectively. In a previous work [27], the fact of depicting finer chemical information by employing invariant-based similarity values has been proven. Expression (3) surpasses the role of constitutional data by taking into account molecular characteristics (e.g., number and nature of intramolecular bonds) instead of the number of vertexes and edges.

Although the invariant-based similarity shows a more appropriate behavior than that shown by the constitutional similarity, nonisomorphic information is required in order to achieve predictive AS values. Thus, the second and third terms of Expression (2) add the contribution of the nonisomorphic fragments by means of a product between a nonisomorphic distance value ($\Gamma_{A,B}$) and a factor (w_Γ) which weighs up the distance contribution, as shown in Expression (4) and (5), respectively.

$$\Gamma_{A,B} = \left(\frac{\mathrm{NIF}_A}{TD(A)} - \frac{\mathrm{NIF}_B}{TD(B)}\right) \tag{4}$$

$$w_\Gamma = \left(1 - \frac{\sqrt{[TD(\mathrm{NIF}_A) - TD(\mathrm{NIF}^P_A)]^2 + [TD(\mathrm{NIF}_B) - TD(\mathrm{NIF}^P_B)]^2}}{\sqrt{[TD(A) - TD(A^P)]^2 + [TD(B) - TD(B^P)]^2}}\right) \tag{5}$$

The greater the difference between the molecules A and B is, the higher the distance value shows. The distance value is normalized using the invariants of the molecular graphs A and B. In addition, the distance contribution in the similarity correction depends on the size and nature of the molecular graphs and the noncommon subgraphs computed over the distance matrix—$TD(A^P)$, $TD(B^P)$, $TD(\mathrm{NIF}^P_A)$ and $TD(\mathrm{NIF}^P_B)$—and over the weighted distance matrix— $TD(A)$, $TD(B)$, $TD(\mathrm{NIF}_A)$ and $TD(\mathrm{NIF}_B)$. Each element (i, j) of the weighted distance matrix stores the minimal pathway length between the atom i and the atom j using interatomic distances relatives

regarding to the simple bond C–C distance. The nonweighted distances matrix also stores the minimal pathway lengths between atoms, but considering all the interatomic distances as 1.00 independently of the bond type. Logically, the weighted matrix led to better molecular representations. The fact of employing nonweighted distance matrices in addition to the weighted ones increased the structural differentiation achieved by using only the weighted distance matrices.

2.4 Data analysis and software employed

Approximate similarity (AS) matrices with dimensions 22 by 22 were built using Expression (2). Considering multivariate regression, the *AS* matrix is considered as a set of 22 objects (rows) characterized by 22 variables (columns). Thus, an object (chemical compound) is described using a series of global variables which accounts for the approximate similarity between the molecule and a reference compound.

Partial least squares regression (PLSR) [31] was employed as multivariate regression technique. PLSR reduced the original similarity spaces considering variances of the predictor and property matrices. PLSR allowed the use of symmetric matrices—other regression techniques, e.g. multiple linear regression (MLR), requires systems with more objects than predictors— and different statistical parameters were obtained, namely: coefficient of determination, standard error in prediction, slope and bias of the correlation analysis [32,33].

Software for isomorphism extraction, invariant computation and similarity calculation was developed by authors in C programming language. Regression and validation methods were implemented using the Matlab Statistics Toolbox [34].

3 Results and discussion

3.1 Study of topological descriptors and outliers

Several AS matrices were built each of which used a given topological invariant for describing the common and noncommon substructures. For all these *AS* matrices, full-cross validation (FCV) processes were carried out and predictive models for the SPPP inhibitory capacity were obtained ($Q^2 > 0.60$). Wiener and HyperWiener descriptors, which make use of different distance matrices distributions, led to the best predictions. While the Wiener invariant consists of the half-sum of all the distance matrix elements, the HyperWiener descriptor also adds a quadratic term of the distance elements. Thus, usefulness of each invariant depends on the structural differentiation required for obtaining predictive QSAR models. In this case, SPPP derivatives only required the graph representation provided by the Wiener index, and the HyperWiener could have added a noise component to the similarity space.

Table 1 shows the experimental pIC_{50} activities for the 22 SPPP compounds and the predicted values using the best AS-based QSAR model (Wiener descriptor). The activity values predicted by the previous QSAR method proposed in bibliography are

Table 1 Experimental, predicted and reference values for the inhibitory capacity of the 22 SPPP derivatives. Reference values were predicted by the previous QSAR model (ref. 22)

Molecules	pCI_{50}		
	Experimental	Predicted	Reference
1. O=C1CC2(N3C=CC=C3C(=O)NC2=O)C(=O)N1	0.201	0.412	0.330
2. CN1C(=O)C2=CC=CN2C3(CC(=O)NC3=O)C1=O	0.553	0.379	0.365
3. CN1C(=O)C2=CC=CN2C3(CC(=O)NC3=O)C1=O	0.699	0.673	0.611
4. O=C1CC2(N3C=CC=C3C(=O)N(CCC4=CC=CC=C4)C2=O)C(=O)N1	0.569	Outlier	0.584
5. O=C1CC2(N3C=CC=C3C(=O)N(CC4=CC=CC=C4)C2=O)C(=O)N1	1.004	1.033	1.203
6. FC1=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=CC=C1	1.215	1.161	1.259
7. FC1=CC=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	0.921	1.040	1.209
8. ClC1=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=CC=C1	1.432	1.276	1.188
9. ClC1=CC=CC(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)=C1	1.456	1.429	1.519
10. ClC1=CC=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	1.420	1.315	1.308
11. BrC1=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=CC=C1	1.284	1.337	1.193
12. BrC1=CC=CC(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)=C1	1.432	1.341	1.475
13. BrC1=CC=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	1.328	1.267	1.376
14. CC1=CC=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	1.284	1.214	1.149
15. COC1=CC=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	1.337	1.352	1.159
16. FC(F)(F)C1=CC=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	1.097	Outlier	1.116
17. O=C1CC2(N3C=CC=C3C(=O)N(CC4=CC=C(C=C4)N(=O)=O)C2=O)C(=O)N1	0.959	0.854	0.994
18. NC1=CC=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	0.854	0.984	0.786
19. FC1=CC(F)=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	1.143	1.052	1.253
20. ClC1=C(Cl)C=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=C1	1.638	1.633	1.559
21. FC1=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=CC(Cl)=C1	1.387	1.484	1.291
22. FC1=C(CN2C(=O)C3=CC=CN3C4(CC(=O)NC4=O)C2=O)C=CC(Br)=C1	1.347	1.396	1.337

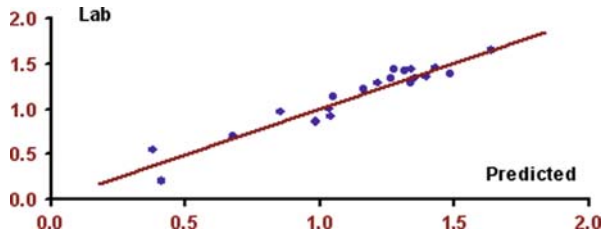

Fig. 2 Experimental versus predicted values plot and statistical parameters obtained in full cross validation process

also given. Statistical parameters obtained were as follows:

$$Q^2 = 0.92,\ \text{Standard Error in Cross Validation (SECV)} = 0.10,$$
$$\text{slope} = 1.01,\ \text{intercept} = 0.00$$

Thus, accuracy (low SECV in addition to *slope* and *bias* close to 1.00 and 0.00, respectively) and precision ($Q^2 > 0.90$) were both excellent for the model developed. Figure 2 shows the lab *versus* predicted activities plot and its statistical characterization. It is interesting to remark that it SECV was computed over predicted values (prediction error).

Considering the T parameter for each one of the 22 SPPP derivatives (computed the individual residual regarding to the SECV value), compounds **4** and **16** were detected as outliers. The reason for detecting the compound **4** as an outlier could be the fact of employing only a compound showing a $–(CH_2)_2–$ bridge group between the pyrazine and the $–C_6H_5$ substituent. Thus, when the activity of compound **4** was predicted in its corresponding FCV cycle, modeling of this structural characteristic was not achieved with the remaining compounds which train the individual model, and therefore, a high deviation was obtained.

The other outlier detected was the compound **16**. As Fig. 1 shows, this SPPP molecule has a $–CF_3$ group attached to the $–C_6H_5$ substituent. Possible reasons for the anomalous behavior of this compound within the model could lie on the bulky substructure of the $–CF_3$ group, in addition to its high electronegativity value.

3.2 Analysis of the AS-QSAR equation coefficients

Figure 3 shows the equation coefficients of the QSAR model, built by FCV, for the 22 SPPP compounds. Each coefficient measures the influence of the similarity to the derivative represented by the corresponding variable on the inhibitory capacity: the higher the absolute value of the coefficient shows, the greater the influence on the activity presents—if the coefficient has positive sign, the influence will be positive, and vice versa.

As can be observed in Fig. 3, the lowest coefficient values were those of the compounds **1**, **7**, **17** and **18**. Thus, high similarities to these compounds will produce low activity values. As expected, these molecules show low activities and common

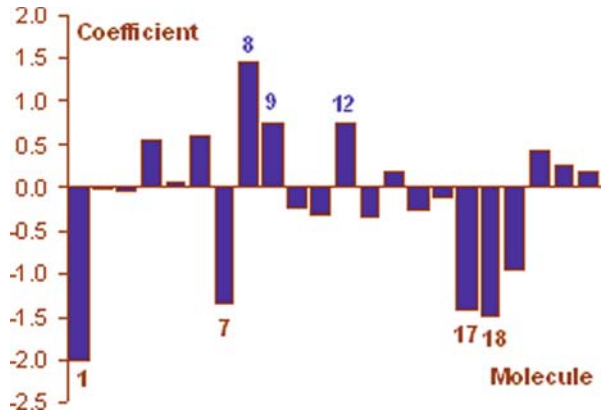

Fig. 3 Coefficients of the QSAR model. Each coefficient represents the similarity to each one of the 22 SPPP derivatives

structural characteristics. With the exception of the compound **1**, whose structure is the isomorphism extracted for most of the SPPP derivatives pairs, compounds **7**, **17** and **18** presents the $-C_6H_5$ group attached to the pyrazine substructure substituted in *para* positions by –F, $-NO_2$ and $-NH_2$, respectively.

On other side, SPPP derivatives 8, 9 and 12 showed, as descriptor variables, the highest QSAR equation coefficients. In these cases, the $-C_6H_5$ group has substituents different of the fluoride (–Cl, –Br) in *ortho* and *meta* positions. In addition, other compounds with high inhibitory capacities also have substituents in the same *ortho* and *meta* positions. Thus, the new SPPP derivatives design should provide compounds with these substituents position characteristics.

Thus, similarity-based QSAR equations permitted to assess the influence of the degree similarity on the inhibitory capacity of derivatives and to provide mathematical support for logical chemical concepts.

4 Conclusions

A high predictive QSAR model for the activity of spirosuccinimide fused tetrahydropyrrolo[1,2-a]pyrazine (SPPP) compounds as inhibitors of the aldose reductase enzyme has been presented in this work. The QSAR modeling makes use of the approximate similarity methodology, which uses isomorphic and nonisomorphic data fusion.

Based on the use of topological chemical representations, structure-behavior developments characterized by accuracy, precision and robustness were achieved by means of isomorphic and nonisomorphic data fusion. In spite of obtaining several predictive models, the optimal spaces outcame from employing weighted and nonweighted distance matrices and the Wiener descriptor as the topological invariant. Thus, the Wiener index applied separately to isomorphic and nonisomorphic fragments allowed to model internal and external substructures.

Results could be reasonably well compared with the previous QSAR approaches for the SPPP compounds and the interpretation of models should permit to extract some

interesting structural requirements in order to achieve high inhibitory capacities. The fact of using graph topological measurements involves simplicity and fastness which can provide pharmaceutical industries with a powerful tool for predicting and screening the behavior of new SPPP derivatives.

Acknowledgements We would like to thank the Comisión Interministerial de Ciencia y Tecnología (CiCyT) and FEDER for their financial support (Project: TIN2006-02071).

References

1. P.F. Kador, Med. Res. Rev. **8**, 325–352 (1988)
2. D.R. Tomlinson, E.J. Stevens, L.T. Diemel, Al. Trends Pharmacol. Sci. **15**, 293–297 (1994)
3. C. Yabe-Nishimura, Pharmacol. Rev. **50**, 21–33 (1999)
4. M. Laasko, Diabetes **48**, 937–942 (1998)
5. A.Y.W. Lee, S.S.M. Chung, FASEB **13**, 23–30 (1999)
6. O. El-Kabbani, F. Ruiz, C. Darmanin, R.P.T. Chung, Cell. Mol. Life Sci. **61**, 750–762 (2004)
7. M. Yamagishi, Y. Yamada, K. Ozaki, M. Asao, R. Shimizu, M. Suzuki, M. Matsumoto, Y. Matsuoka, K. Matsumoto, J. Med. Chem. **35**, 2085–2094 (1992)
8. L. Costantino, G. Rastelli, K. Vescovini, G. Cignarella, P. Vianello, A. Del Corso, M. Cappiello, U. Mura, D. Barlocco, J. Med. Chem. **39**, 4396–4405 (1996)
9. T. Kotani, Y. Nagaki, A. Ishii, Y. Konishi, H. Yago, S. Suehiro, N. Okukado, K. Okamoto, J. Med. Chem. **40**, 684–694 (1997)
10. P. Fresneau, M. Cussac, J.M. Morand, B. Szymonski, D. Tranqui, G. Leclerc, J. Med. Chem. **41**, 4706–4715 (1998)
11. L. Costantino, G. Rastelli, M.C. Gamberini, J.A. Vinson, P. Bose, A. Iannone, M. Staffieri, L. Antolini, A. Del Corso, U. Mura, A. Albasini, J. Med. Chem. **42**, 1881–1893 (1999)
12. L. Costantino, G. Rastelli, M.C. Gamberini, M.P. Giovannoni, V.D. Piaz, P. Vianello, D. Barlocco, J. Med. Chem. **42**, 1894–1900 (1999)
13. L. Costantino, A.D. Corso, G. Rastelli, J.M. Petrash, U. Mura, Eur. J. Med. Chem. **36**, 697–703 (2001)
14. G. Bruno, L. Costantino, C. Curinga, R. Maccari, F. Monforte, F. Nicolò, R. Ottanà, M.G. Vigorita, Bioorg. Med. Chem. **10**, 1077–1084 (2002)
15. F.D. Settimo, G. Primofiore, A.D. Settimo, C.L. Motta, F. Simorini, E. Novellino, G. Greco, A. Lavecchia, E. Boldrini, J. Med. Chem. **46**, 1419–1428 (2003)
16. I. Nicolaou, C. Zika, V.J. Demopoulos, J. Med. Chem. **47**, 2706–2709 (2004)
17. A. Pau, B. Asproni, G. Boatto, G.E. Grella, P.D. Caprariis, L. Costantino, G.A. Pinna, Eur. J. Pharm. Sci. **21**, 545–552 (2004)
18. European Chemicals Bureau. Toxicology and Chemicals Substances. http://ecb.jrc.it/
19. (a) H. Kubinyi, Drug Discov Today **2**(11), 457–467 (1997) (b) H. Kubinyi, Drug Discov Today **2**(12), 538–546 (1997)
20. H. Van de Waterbeemd (ed.), Structure-Property Correlations in Drug Research (Academic Press, Austin, TX, 1996)
21. K. Ko, Y. Won, Bioorg. Med. Chem. **13**, 1445–1452 (2005)
22. K. Ko, H. Won, Y. Won, Bioorg. Med. Chem. **14**, 3090–3097 (2006)
23. R.P. Sheridan, P. Hunt, J.C. Culberson, J. Chem. Inf. Comput. Sci. **46**, 180–192 (2006)
24. R.A. Lewis, J. Med. Chem. **48**, 1638–1648 (2005)
25. M. Urbano Cuadrado, I. Luque Ruiz, M.A. Gómez-Nieto, A New Quantitative Structure-Property Relationship Based on Topological Distances on Non-isomorphic Subgraphs, in Lectures Series on Computer and Computational Sciences: Advances in Computational Methods in Sciences and Engineering (Brill Academic Publisher, 2005), pp. 135–138
26. M. Urbano Cuadrado, I. Luque Ruiz, M.A. Gómez-Nieto, J. Chem. Inf. Model. **46**(4), 1678–1686 (2006)
27. M. Urbano Cuadrado, I. Luque Ruiz, M.A. Gómez-Nieto, J. Chem. Inf. Model. **46**(5), 2022–2029 (2006)
28. T. Negoro, M. Murata, S. Ueda, B. Fujitani, Y. Ono, A. Kuromiya, M. Komiya, K. Suzuki, J.I. Matsumoto, J. Med. Chem. **41**, 4118–4129 (1998)

29. Chemaxon Ltd. http://www.chemaxon.com/marvin
30. G. Cerruela García, I. Luque Ruiz, M.A. Gómez-Nieto, J. Chem. Inf. Comput. Sci. **44**, 30–41 (2004)
31. S. Wold, M. Sjostrom, L. Eriksson, Chemom. Intell. Lab. Syst. **58**, 109–130 (2001)
32. S. Wold, L. Eriksson, in statistical Validation of QSAR Results. In Chemometrics Methods in Molecular Design. ed. by H. van de Waterbeemd (VCH, Weinheim, 1995), pp. 309–318
33. A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. **20**, 269–276 (2002)
34. MATLAB, Version 6.5 (R13). Mathworks Inc., 2003; www.mathworks.com