# Advantages of the Approximate Similarity Approach in the QSAR Prediction of Ligand Activities for Alzheimer Disease Detection

Irene Luque Ruiz, Manuel Urbano-Cuadrado and Miguel Ángel Gómez-Nieto

*Abstract*— The improvement on the QSAR prediction of the trans-stilbenes affinity for the β-amyloid peptide (employed for detecting the Alzheimer disease) achieved by means of using approximate similarity measurement is presented in this work. A wide spectrum of similarity methods is described, and results obtained by approximate similarity are compared with those obtained by constitutional, fingerprint and descriptor-based similarity. The fact of using similarity corrections by considering distances between the non-isomorphic fragments (the approximate similarity concept) led to accurate QSAR models ($Q^2$ > 0.80). The high predictive ability achieved by simple methods is remarked.

*Index Terms*— Similarity, Non-isomorphic Dissimilarity, Approximate similarity, Drug activity prediction.

## I. INTRODUCTION

Chemists have always employed models aimed at representing complex chemical entities in a simple way: names, molecular weight, graphs, and so on. But the rise of computer science has allowed developing a great amount of methods with aim of transforming molecules into data structures amenable to be processed by computers [1].

Computational chemistry encompasses a series of mathematical methods implemented by computer which show a wide spectrum of applications, namely: reproduction of chemical processes, modeling of structures, prediction of properties, activities and reaction variables, etc. [2]

The description of structures by means of numbers enables the application of statistical methods to establish a mathematical relationship between the description carried out and properties and/or behavior of molecules. Quantitative Structure Activity/Property Relationship (QSAR/QSPR) are

the computational chemistry disciplines which propose different methodologies for molecular description and study their efficiency for in-silico prediction [3].

Several classifications of these methodologies could be carried out depending on the characteristics of the descriptor vector (2D or 3D, local or global, constitutional or geometrical, etc.) and on the mathematical method employed to define the QSAR/QSPR equation (univariate or multivariate, parametric or non parametric, global or feature selection, etc.) [4].

2D similarity approaches can be defined as simple methods since they employ topological measurements derived from molecular graphs, but many times they show inconsistencies for the appropriate representation of QSAR/QSPR predictive spaces [5]. Thus, accuracy achieved is not enough to consider models as predictive tools. In this work we present the advantages resulting from the use of the recently proposed approximate similarity. This measurement refines the chemical information extracted from molecular graphs by considering new chemical pieces of information: the non-isomorphic substructures. The affinity of trans-stilbenes for the β-amyloid peptide [6] has been the chemical frame considered due to its application in the early detection of the Alzheimer disease.

This work has been organized as follows: after the introductory section, a general description of the use of the similarity concept in QSAR/QSPR is given. Section 3 describes different classical similarity approaches, whereas the approximate similarity concept is defined in section 4. Evaluation of the different QSAR models developed is carried out in section 5, and finally, conclusions are given.

## II. SIMILARITY APPROACHES

Similarity measurements are often employed to develop screening methods of chemical databases and to predict molecular properties. The latter application is based on the "similar molecules show similar properties" chemical principle, thus enabling the study of both physico-chemical properties and biochemical behavior.

The development of similarity-based QSAR models consists of a series of common stages, namely: a) first, a data set which represents both the molecular diversity and the property/activity range to be modeled and predicted is selected; b) second, 2D or 3D methods are employed to compute the

isomorphism shown by each pair of the compounds which compose the data set; c) a similarity matrix is built by means of using any of the similarity metrics summarized in literature; and d) multivariate regression techniques and validation strategies are employed to establish the prediction equations and to assess the uncertainty reduction achieved, respectively.

Several 2D and 3D similarity methods have been proposed. The former makes use of graph (called molecular graph) representation of all the data set elements by considering the atoms and bonds which compose a molecule as the nodes and edges, respectively. There are two ways of using a molecular graph. First, after computing one or several descriptors over the molecule, we can build univariate or multivariate models to predict chemical properties/activities. In other way, measures of similarity between the different molecule descriptions of the data set can be carried out by means of considering size and nature of molecules and of the isomorphic fragments [7].

Three dimensional similarity methods, based on Comparative Molecular Field Analysis (CoMFA), consider XYZ coordinates of atoms and grids enclosing data set structures. Then, a series of molecular field interactions are computed at each one of the grid points, and similarity between all the molecules are computed by taking into the overlapping between a pattern randomly selected and each data set element in the built grid. Thus, a similarity measurement can be considered as a distance between de calculated descriptors for the XYZ coordinates of each data set element [8].

In spite of the efficiency achieved by 3D methods, some shortcomings related to their high computational cost are involved in this kind of QSAR developments. For instance, optimizations of 3D structures are carried out by complex methods like quantum mechanics, molecular dynamics, molecular mechanics, etc. In other hand, alignment of structures is also required in order to reproduce the 3D space according to key parts for the property/activity. In addition, selection of a representative pattern is also a subjective step.

Regarding 2D methods, optimization and alignment of chemical structures are not involved, thus not requiring high computational resources. In spite of this, similarity measurements have shown some problems in predicting molecular properties even in data sets composed by compounds showing similar structures.

In previous works [9], we have proposed the Approximate Similarity concept. This new similarity measurement, which involves all the characteristics of 2D similarity approaches, considers several pieces of 2D structural information, namely: a) graph information of the data set elements; b) data of common fragments extracted from the isomorphism detection process; c) information of the non-isomorphic fragments obtained in the graph matching, i.e., those substructures which do not compose the isomorphism computed for each pair of the data set elements; and d) invariant-based description of whole graphs, and of isomorphic and non-isomorphic fragments by means of 2D-descriptors which accounts for their chemical nature.

## III. CLASSICAL SIMILARITY APPROACHES

In this work, we expose the application of several 2D similarity methods to the development of QSAR models, thus showing the advantage derived from the use of the approximate similarity.

### A. Representation of the data set elements

The molecules which compose the data set to be modeled are represented by non-directed and non-weighted graphs, known as molecular graphs. A molecular graph $G_A$ consisting of $n_A$ nodes and $e_A$ edges which represent the atoms and bonds of a molecule, respectively.

### B. Constitutional Similarity

Bi-dimensional similarity measurements are obtained after representing the data set elements by means of molecular graphs and isomorphism extraction. Classically, graph matching computes the number of nodes and edges of the two matched molecular graphs and of the common fragments.

Several similarity indices have been proposed with the aim of relating the constitutional description with a similarity measurement normalized within the range [0,1]. Tanimoto and Cosine formulas have been widely employed as similarity indices. These indices are shown in expressions (1) and (2).

$$\text{Tanimoto: } S_{A,B} = \frac{c}{a+b-c} \tag{1}$$

$$\text{Cosine: } S_{A,B} = \frac{c}{\sqrt{a \times b}} \tag{2}$$

where: $a$ and $b$ are the sizes (number of nodes and edges) of the molecular graph $G_A$ and $G_B$, respectively, and $c$ is the size of the molecular graph $G_C$ which represents the isomorphism extracted form the $G_A$ and $G_B$ matching.

The $G_C$ graph can be obtained by different ways. $G_C$ is often considered as a connected graph representing the maximal common subgraph (MCS) between $G_A$ and $G_B$. But other approaches do not show the requirement of fully connected graph, these representing either the set of maximal common edges subgraphs (MCES) or the set of all the maximal common subgraphs (AMCS).

### C. Fingerprint-based Similarity

Other classical approaches to the 2D similarity concept make use of the transformation of the molecules into fingerprints [10]. A fingerprint is a binary array of a preset size which represents structural properties of the molecular graphs. Different kinds of fingerprints have been proposed depending on the structural elements to be represented and on the array size.

In a general way, fingerprint construction consists of a series of steps, namely: a) generation of the molecular graph for each element of the data set; b) obtaining of the subgraphs showing size from 1 to m (often lower than 9) for each graph; c) extraction of preset pattern substructures in some fingerprint kinds; d) assignation of a binary representation and position of each path and pattern presented in the data set; e) and finally, the fingerprint construction.

So, fingerprints can be considered as data structures which do not require great computational costs for their handling and they store greater structural information than that shown by the chemical graph. Nevertheless, some shortcomings are involved in the use of fingerprints. On a hand, if fingerprint size is small, different paths or patterns are located at the same bits, thus giving redundancy and high density fingerprints which produce high similarity values and data inconsistencies. On the other hand, big sizes are often responsible for scattered fingerprints which produce extremely low similarity measurements.

Fingerprint similarity values are most of times obtained by means of Tanimoto index and through the computation of boolean operations between the data set fingerprints. Values for a and b (see expressions 1 and 2) are the number of bits equal to 1 of fingerprints $A$ and $B$, respectively, whereas c represents the number of bits equal to 1 and common to the fingerprints of molecules $A$ and $B$.

### D. Descriptor-based Similarity

Recently, the use of different descriptors computed over molecular graphs has been proposed with the aim of obtaining similarity measurements. So, taking into account the molecular graphs $G_A$ and $G_B$ representing the molecules $A$ and $B$, respectively, a similarity approach based on extracting a descriptor or invariant over graphs $G_A$, $G_B$ and $G_C$ can be proposed. Advantages of this kind of approaches derive from the use of descriptors which account for different structural properties of molecular graphs. Therefore, a given descriptor can be selected in order to explain in a major extent the property/activity we are trying to model and predict.

Descriptor-based similarity measurements can be obtained by any descriptor, for instance, the Cosine index as follows:

$$Cosine: \ S_{A,B} = \frac{td(G_C)}{\sqrt{td(G_A) \times td(G_B)}} \qquad (3)$$

where: $td(G_A)$, $td(G_B)$ and $td(G_C)$ are the values of a given descriptor computed over the graphs of molecules $A$, $B$ and of the extracted isomorphism, respectively.

### IV. APPROXIMATE SIMILARITY APPROACH

The above described similarity approaches employ characteristics of molecular graphs and of isomorphic subgraphs extracted in the matching process for all the pairs of the data set. But these approaches do not consider straightly the fragments which do not compose the extracted isomorphism.

For instance, taking into account the three molecular graphs shown in Fig. 1, classical similarity measurements only consider the characteristics of graphs $G_A$, $G_B$ and $G_C$ and characteristics of the isomorphic subgraph $G_I$. As can be observed, the matching processes $G_A$-$G_B$ and $G_B$-$G_C$ computed the same isomorphism.

Nevertheless, non-isomorphic fragments are also responsible for the properties/activities of the molecules shown in Fig. 1.

Our proposal is to consider the contribution to the similarity

measurement of the non-isomorphic fragments. Thus, we take into account the distances between the subgraphs that do not form the isomorphism $I_{A,B}$. Thus, the structural difference $\Gamma_{A,B}$ (dissimilarity or distance) between two molecular graphs $G_A$ and $G_B$ is calculated as follows:

$$\Gamma_{A,B} = g[td(G_A, I_{A,B}), td(G_B, I_{A,B})] =$$
$$= g[td(NIF_A), td(NIF_B)] \qquad (4)$$

where: $I_{A,B}$ represents the isomorphism extracted from matching of the $G_A$ and $G_B$ molecular graphs; non-isomorphic subgraphs $NIF_A = G_A - I_{A,B}$ and $NIF_B = G_B - I_{A,B}$ correspond to the subgraphs of $G_A$ and $G_B$, respectively, that do not form the isomorphism $I_{A,B}$; $g()$ is a function aimed at obtaining a distance value (e.g. Euclidean, Mahalanobis, etc.) between $td(NIF_A)$ and $td(NIF_B)$; and $td$ is a topological descriptor which describe the non-common subgraphs, namely: Wiener (W), Hyper Wiener (WW) and so on indexes. Contrary to similarity, higher the $\Gamma_{A,B}$ shows, higher the dissimilarity between the molecules $A$ and $B$ is.
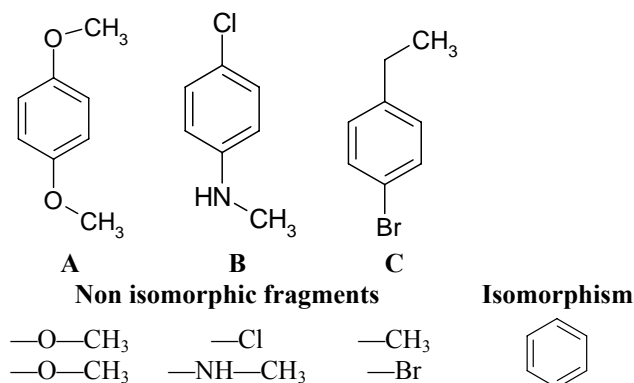


**Fig. 1**. Isomorphic and non-isomorphic fragments for three examples of molecules

### A. Correction of the structural similarity: the Approximate Similarity

With the aim of defining a new similarity measurement which takes into accounts both the classical similarity and the non-isomorphic distance, the Approximate Similarity (AS) is defined as follows:

$$AS_{A,B} = f(S_{A,B}, \Gamma_{A,B}, w_\Gamma) \qquad (5)$$

where: $S_{A,B}$ is a classical similarity measurement (constitutional, fingerprint-based or descriptor-based); $\Gamma_{A,B}$ is the dissimilarity defined in equation (4), and $w_\Gamma$ is a weighting factor which adjusts the distance contribution in the approximate similarity calculation.

Thus, chemical similarity achieved by the AS approach will be more accurate due to the consideration of the difference between the non-common substructures of the matched molecules, which most of times are responsible for their properties/activities.
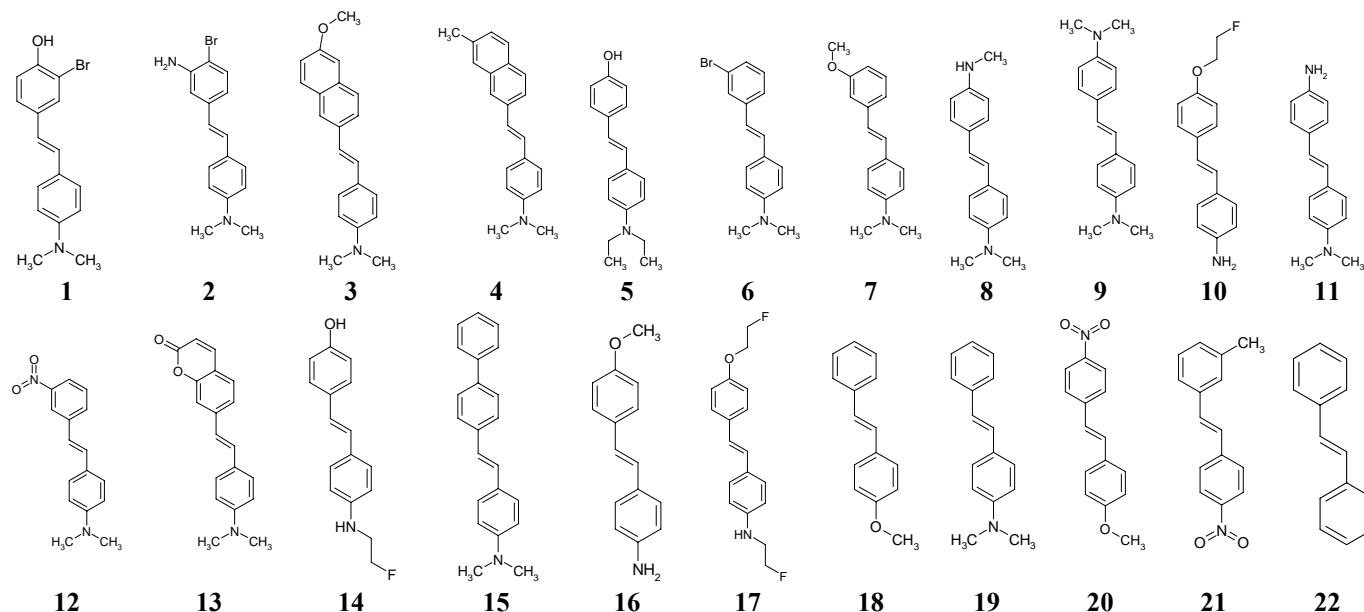
**Fig. 2.** Molecular graphs for the data set of 22 trans-stilbene

**Table I**. SMILE structures, and experimental, predicted and reference affinities for the data set

| | Molecules | pK$_i$ | | |
|---|---|---|---|---|
| | | Exp. | Pred. | Ref. |
| 1 | CN(C)C1=CC=C(\C=C\C2=CC(Br)=C(O)C=C2)C=C1 | 8.70 | 8.46 | 8.56 |
| 2 | CN(C)C1=CC=C(\C=C\C2=CC=C(Br)C(N)=C2)C=C1 | 8.55 | 8.21 | 8.13 |
| 3 | COC1=CC2=CC=C(\C=C\C3=CC=C(C=C3)N(C)C)C=C2C=C1 | 8.15 | 8.14 | 8.23 |
| 4 | CN(C)C1=CC=C(\C=C\C2=CC=C3C=CC(C)=CC3=C2)C=C1 | 8.14 | 7.91 | 8.09 |
| 5 | CCN(CC)C1=CC=C(\C=C\C2=CC=C(O)C=C2)C=C1 | 8.00 | 8.05 | 8.04 |
| 6 | CN(C)C1=CC=C(\C=C\C2=CC=CC(Br)=C2)C=C1 | 7.90 | 8.52 | 7.74 |
| 7 | COC1=CC(=CC=C1)\C=C\C2=CC=C(C=C2)N(C)C | 7.87 | 8.07 | 7.60 |
| 8 | CNC1=CC=C(C=C1)\C=C\C2=CC=C(C=C2)N(C)C | 7.82 | 7.70 | 7.82 |
| 9 | CN(C)C1=CC=C(\C=C\C2=CC=C(C=C2)N(C)C)C=C1 | 7.73 | 7.83 | 7.99 |
| 10 | NC1=CC=C(C=C1)\C=C\C2=CC=C(OCCF)C=C2 | 7.59 | 7.29 | 7.42 |
| 11 | CN(C)C1=CC=C(\C=C\C2=CC=C(N)C=C2)C=C1 | 7.58 | 7.56 | 7.80 |
| 12 | CN(C)C1=CC=C(\C=C\C2=CC=CC(=C2)N(=O)=O)C=C1 | 7.56 | 7.29 | 7.65 |
| 13 | CN(C)C1=CC=C(\C=C\C2=CC=C3C=CC(=O)OC3=C2)C=C1 | 7.55 | 7.72 | 7.93 |
| 14 | OC1=CC=C(C=C1)\C=C\C2=CC=C(NCCF)C=C2 | 7.52 | 7.47 | 7.65 |
| 15 | CN(C)C1=CC=C(\C=C\C2=CC=C(C=C2)C3=CC=CC=C3)C=C1 | 7.47 | 7.76 | 7.48 |
| 16 | COC1=CC=C(C=C1)\C=C\C2=CC=C(N)C=C2 | 7.44 | 7.74 | 7.68 |
| 17 | FCCNC1=CC=C(C=C1)\C=C\C2=CC=C(OCCF)C=C2 | 7.41 | 7.44 | 7.21 |
| 18 | COC1=CC=C(C=C1)\C=C\C2=CC=CC=C2 | 7.36 | 7.08 | 7.08 |
| 19 | CN(C)C1=CC=C(\C=C\C2=CC=CC=C2)C=C1 | 7.35 | 7.27 | 7.31 |
| 20 | COC1=CC=C(C=C1)\C=C\C2=CC=C(C=C2)N(=O)=O | 7.33 | 7.45 | 7.31 |
| 21 | CC1=CC=CC(\C=C\C2=CC=C(C=C2)N(=O)=O)=C1 | 6.82 | 6.82 | 6.94 |
| 22 | C1=CC=C(C=C1)\C=C\C2=CC=CC=C2 | 6.24 | 6.36 | 6.38 |

## V. MATERIAL AND METHODS

In this work, the efficiency of approximate similarity methods was tested for the prediction of binding affinity of trans-stilbene derivatives to the β-amyloid (Aβ) peptide. Aβ accumulation in the brain is a key symptom for the development of Alzheimer's disease (AD), which destroys the part of the nervous system responsible for storing memories. The fact of detecting Aβ accumulations by means of non invasive spectroscopic techniques is pursued by the scientific community in order to detect the disease in its early development.

The synthesis of specific ligands of Aβ which act as imaging factors to reveal the presence of Aβ has been widely studied. The trans-stilbene series (22 compounds) studied in this work shows a wide range of affinity for the Aβ plaques consisting of Aβ$_{1-40}$ aggregates in the brain of AD people. Thus, there is a great interest in developing computer tools for the aid of development of trans-stilbene derivatives. A

fast and efficient QSAR model based on 2D similarity calculations and distance corrections of similarity, respectively, could provide a useful predictive tool.

Fig. 2 and Table I show, respectively, the 2D structures and the SMILE representations of the 22 stilbene compounds. *MarvinScketch* was employed as builder for the 2D structures, whereas the fingerprints were generated by *generfp* of *JChem* [11]. As can be observed, the structure of the compound 22 is a substructure of the rest of stilbene molecules.

The affinity values are also shown in Table I, expressed as $pK_i = -log(K_i)$, and the statistical characterization of the experimental data set affinity is as follows: *N*: 22, *mean*: 7.64, *min*: 6.24, *max*: 8.70, *standard deviation*: 0.52.

The isomorphism computation for all the pairs of compounds was carried out by using an algorithm which extracts the maximum common substructure (MCS). This program was developed by authors [12]. Wiener index (eq. 6) was employed to calculate descriptor-based similarities.

$$W = \sum_{i}^{N} \sum_{\substack{j \\ i<j}}^{N} d_{ij} \qquad (6)$$

The Wiener computation was also modified (W*) in order to consider the weighted distance matrix. In this matrix, each elements *(i, j)* corresponds to the minimal pathway length between the graph nodes (atoms) *i* and *j* computed by considering interatomic distances relative to the *C-C* bond.

The Wiener index has been demonstrated to be useful for predicting molecular properties since three decades ago [7]. This index allows describing the molecular size and volume, which have been widely related to properties/activities of molecules. In addition, W* provides information of the atoms and bonds that compose the molecule, thus refining the chemical information shown by this kind of graph invariants.

## VI. EVALUATION OF THE DIFFERENT SIMILARITY APPROACHES

Partial Least Squares Regression (PLSR) was employed as multivariate regression technique [13]. PLSR reduced original similarity spaces by considering variances of predictors and properties/activities. In addition, PLSR permitted the use of symmetric matrices —other regression techniques, e.g. Multiple Linear Regression (MLR), require systems with more objects than predictors—. Leave one out was the strategy employed to validate the quality of equations. Different statistical parameters were studied, namely: coefficient of determination ($Q^2$), standard error in cross-validation (*SECV*), and slope and intercept of the predicted vs. experimental plot. All these parameters are referred to predictions. Study of anomalies was also carried out.

QSAR community considers that meaningful models are obtained when $Q^2 > 0.50$ and, on the other hand, when the *SECV* value is much lower than the standard deviation of the

data set. In addition, slope and bias close to 1 and 0, respectively, also confirm the models predictive capacity.

Table II shows the models built by the different similarity approaches studied in this work. As can be observed in Table II, classical similarity matrices did not give good predictive models since all the indices employed led to poor values for all the statistical parameters. In a similar way, descriptor-based similarities did not show predictive ability in spite of using the modified Wiener index (W*). $Q^2$ values were much lower than 0.5 and *SECV* was similar to the standard deviation of the affinities, thus not obtaining uncertainty reduction.

Regarding models derived from use of fingerprints, acceptable $Q^2$ values were obtained, namely: 0.58 and 0.56 for the Tanimoto and Cosine indices, respectively. The slightly higher predictive ability achieved for the Tanimoto index could be due to the reduction of redundancies observed in the location of fragments in both high density and scattered fingerprints. Nevertheless, these models only can be employed for screening tools (separation of low, medium and high affinity values). Greater $Q^2$ values are required to achieve robust QSAR models.

Taking into account the above commented results, 2D similarity approaches show shortcomings for the development of QSAR models. In these cases, predictive spaces were symmetric similarity matrices built by means of isomorphism measurements which only consider characteristics of the molecular graphs and of the computed isomorphic fragment. Thus, the development of a 2D similarity measurement which employs non-isomorphic contributions should be attempted. The similarity correction by considering non-isomorphic information was carried out as expression (7).

$$AS_{A,B} = S_{A,B} - \left[ \frac{W_{MCS}^*}{W_A^*} - \frac{W_{MCS}^*}{W_B^*} \right] -$$
$$- \left[ \frac{(W_{NIF_A}^* - W_{NIF_A}) - (W_{NIF_B}^* - W_{NIF_B})}{(W_A^* - W_A) - (W_B^* - W_B)} \right] \qquad (7)$$

We can observe three terms in equation (7). First, one of the classical similarity $S_{A,B}$ above described is employed. The second component is a measurement of the isomorphic fragment weighted contribution for each pair of data set elements with regard to the matched molecules.

This contribution was obtained by using modified Wiener index (W*), which is computed over the weighted distance matrix of graphs. Finally, the third term considers the non-isomorphic fragments (*NIF*) contribution to the similarity correction. With this aim, Wiener and modified Wiener indices (W and W*) were computed over the normal and weighted distance matrices, respectively.

Table II shows the results obtained by the approximate similarity approach of equation (7). Similarity measurement considered was that derived from fingerprints by using the Cosine index.

**Table II**. Experimental results for different similarity-based approaches

| Similarity | | | Prediction Model | | | | |
|---|---|---|---|---|---|---|---|
| Model | Index | Descriptor | Slope | Intercept | $Q^2$ | SECV | Outlier |
| **Classical** | *Tanimoto* | *None* | 0.62 | 2.91 | 0.12 | 0.51 | |
| | *Cosine* | *None* | 0.60 | 3.05 | 0.10 | 0.51 | |
| **Fingerprint-based** | *Tanimoto* | *None* | 0.91 | 0.66 | 0.58 | 0.34 | |
| | *Cosine* | *None* | 0.90 | 0.72 | 0.56 | 0.35 | |
| **Descriptor-based** | *Tanimoto* | *W* | 0.83 | 1.29 | 0.29 | 0.45 | |
| | *Cosine* | *W* | 0.60 | 3.04 | 0.10 | 0.51 | |
| | *Tanimoto* | *W\** | 0.82 | 1.33 | 0.29 | 0.45 | |
| | *Cosine* | *W\** | 0.60 | 3.04 | 0.10 | 0.51 | |
| **Approximate** | *Cosine* | *W, W\** | 1.01 | -0.13 | 0.89 | 0.17 | 1 |

As can be observed, one outlier was detected when a study of descriptor-activity outliers was carried out. On this purpose, the *T* parameter was computed as the *residual/SECV* ratio, and $T_{cut-off}$ was set to 2.5. Without considering the detected anomalous compound, excellent accuracy and precision were achieved ($Q^2 = 0.89$ and $SECV = 0.17$), and slope and intercept extremely close to 1 and 0, respectively.

The outlier corresponded to stilbene derivative 6. As can be observed in Fig. 2 and Table I, the difference between molecules 6 and 7 is at the *meta* substitution of one of the aromatic ring, namely: *bromide* and $-OCH_3$ for molecules 6 and 7, respectively. In this case, the approximate similarity detected the difference of size and electronegativity for both substituents, but this structural difference did not involve great differences between the affinities [14]. Similar interpretation may be given when molecules 1 or 2 (in both cases bromide is present) are compared with 6: very few structural changes are involved in high affinity differences.

When the outlier was considered, statistical parameters were as follows, namely: $Q^2 = 0.80$, $SECV = 0.23$, *slope* = 0.92 and *intercept* = 0.66. Other approximate approaches were tested, but better results were not obtained.

## VII. CONCLUSIONS

In this work, the improvement on the predictive power of similarity-based QSAR models has been achieved by means of the use of a new chemical dimension of the information extracted from molecular graphs. In this way, distances between the subgraphs which do not compose the isomorphism extracted from molecular matching has been considered in correction of classical and invariant-based methods.

Prediction of the affinity of a trans-stilbene series for the β-amyloid (Aβ) peptide was successful in the cases in which the traditional methods fail. Therefore, the consideration of non-isomorphic atoms and bonds led to predictive spaces characterized by high and low $Q^2$ and *SECV* values, respectively. In addition, anomalous behavior shown by compounds was only detected by approximate similarity approaches, thus indicating the richer chemical information modeled.

It is interesting to remark that the simplicity of 2D QSAR methods was maintained since molecular graphs were employed for generating approximate similarity measurements. And as the most important step, quality of prediction was drastically improved, thus concluding that accurate models were obtained in spite of considering only 2D representations.

## REFERENCES

[1] Rouvray, D.H.; Balaban, A.T. Chemical Applications of Graph Theory. Applications of Graph Theory. Wilson, R.J.; Beineke, L.W. (Eds.). Academic Press. 1979, 177-221.

[2] Randic, M. Topological Indices. In Encyclopedia of Computational Chemistry; Scheleyer, P.v.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer, H.F., Schreiner, P.R., Eds.; . John Wiley & Sons; Chichester, 1998; pp. 3018-3032.

[3] Greco, G.; Novellino, E.; Martin, Y.C. Approaches to Three-Dimensional Quantitative Structure-Activity Relationships. In Reviews in Computational Chemistry, Lipkowitz, K. B., Boyd, D. B. (Eds.). Wiley-VCH. New York. 1997.

[4] Ivanciuc, O.; Balaban, A.T. The Graph Description of Chemical Structures. In Topological Indices and Related Descriptors in QSAR and QSPR. Devillers, J., Balaban, A. T. (Eds.). Gordon and Breach Science Publishers. The Netherlands. 1999, 59-167.

[5] Willett, P. Chemical Similarity Searching. J. Chem. Inf. Comput. Sci. 1998, 38, 983-996.

[6] Chen, X. QSAR and primary docking studies of trans-stilbene (TSB) series of imaging agents for b-amyloid plaques. Journal of Molecular Structure: THEOCHEM 2006, 763, 83–89.

[7] Lučić, B.; Lukovits, I.; Nikolić, S.; Trinajstić, N. Distance-Related Indexes in the Quantitative Structure-Property Relationship Modeling, J. Chem. Inf. Comput. Sci. 2001, 41, 527-535.

[8] Kubinyi, H. Comparative Molecular Field Analysis (CoMFA). In Encyclopedia of Computational Chemistry. Schleyer, P.v.R. (Ed. In Chief). Wiley. New York. 1998.

[9] Urbano Cuadrado, M.; Luque Ruiz, I.; Gómez-Nieto, M.A. (a) Refinement and Use of the Approximate Similarity in QSAR Models for benzodiazepine Receptor Ligands. J. Chem. Inf. Model. 2006, 46, 2022-2029. (b) A Steroids QSAR Approach Based on Approximate Similarities Measurements. J. Chem. Inf. Model. 2006, 46, 1678-1686.

[10] Schuffenhauer, A.; Gillet, V. J.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures: Analysis of the BIOSTER Database Using Two-Dimensional Fingerprints and Molecular Field Descriptors. J. Chem. Inf. Comput. Sci. 2000; 40(2); 295-307.

[11] ChemAxon Kft. http://www.chemaxon.com/. (Last acceded February 2007).

[12] Cerruela García, G., Luque Ruiz, I., Gómez-Nieto, M.A. Step-by-Step Calculation of All Maximum Common Substructures through a Constraint Satisfaction Based Algorithm. J. Chem. Inf. Comput. Sci. 2004, 44, 30-41.

[13] Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics, Chemom. Intell. Lab. Syst. 2001, 58, 109-130.

[14] Maggiora, G. F. On Outliers and Activity Cliffss - Why QSAR Often Disappoints. J. Chem. Inf. Model. 2006, 46, 1535-1535.