Journal of Mathematical Chemistry, Vol. 39, No. 1, January 2006 (© 2005)
DOI: 10.1007/s10910-005-9007-3

Putting molecular similarity into context: asymmetric indices for field-based similarity measures

Jordi Mestres*

Chemogenomics Laboratory, Research Unit on Biomedical Informatics, Institut Municipal d'Investigació Mèdica and Universitat Pompeu Fabra, 08003 Barcelona (Catalonia), Spain
E-mail: jmestres@imim.es

Gerald M. Maggiora

Department of Pharmacology and Toxicology, University of Arizona College of Pharmacy, P.O. Box 210207, Tucson, AZ 85721-0207, USA

Received 13 September 2004; revised 3 November 2004

Some of the most widely used indices in molecular similarity searching are intrinsically symmetric in nature, meaning that each molecule under comparison contributes equally to the similarity index. For example, the Carbó and the Hodgkin–Richards similarity indices are respectively, related to the geometric and arithmetic averages of the molecular self-similarities. This work introduces the asymmetric forms of an entire family of field-based molecular similarity indices. By incorporating a weighted contribution of each molecule into the similarity index, the newly obtained asymmetric forms allow for measuring and modulating the similarity of one molecule in the context of another and thus have the potential of alleviating the size dependency often observed in chemical similarity searching.

KEY WORDS: asymmetric similarity, molecular fields, similarity searching, virtual screening

1. Introduction

Similarity and dissimilarity are two alternative viewpoints of the same relational concept when comparing a pair of objects. In a clear abuse of nomenclature, the overall comparative concept is usually referred to as similarity or dissimilarity depending on the predominance of common or uncommon features between two objects. The fact that similarities and dissimilarities are both embedded in a single similarity concept enforces the comparison of objects to have a *symmetric character*, which means that object i is as similar to object j as j is to i. In this respect, when attempting to quantify the similarity between two

*Corresponding author.

objects, preservation of this symmetric character has the advantage of reducing the overall comparison to a single value but some interesting information regarding their relative comparison may remain masked. Furthermore, the significance of this global value cannot be unequivocally determined. It is evident that high-similarity values are related to the fact that both objects possess far more features in common than uncommon. However, different cases can be distinguished when low-similarity values are observed. On one hand, object i may be fully comparable to object j and consequently the low-similarity value comes strictly from additional salient features of object j. On the other hand, none of the objects may be fully comparable to the other, low-similarity being then related to the fact that both molecules neatly present diverse features. Thus it seems clear that besides the overall comparison, means to assess the relative comparison between objects can provide relevant new insights to understand their comparability.

On the basis of empirical results from a wide range of domains, Tversky developed a *contrast model* to explain the *asymmetric character* of similarity [1]. In the contrast model, similarity is defined as an increasing function of common features and decreasing functions of distinctive features as

$$S_{ij} = \theta \cdot f(i \cap j) - \alpha \cdot f(i - j) - \beta \cdot f(j - i) \tag{1}$$

in which similarity, S_{ij}, between i and j is defined in terms of the features common to i and j, $i \cap j$, the features that are distinctive to i, $i - j$, and the features that are distinctive to j, $j - i$. The variables θ, α, and β are parameters that determine the relative weights of these three components of similarity and introduce some flexibility to the definition of similarity in whether common or distinctive features will have more influence. The function f measures the salience of a particular set of features. Therefore, according to the contrast model, the asymmetric character of similarity is explained by differential salience and differential weighting of distinctive features. An important aspect is that Tversky recognized also that the salience of features is not fixed but may vary with context. For example, a set of features in object i may be salient in the context of object j but not in the context of object k and thus the parameters θ, α, and β in the respective comparisons will vary depending on the context.

In the same lines, Holman proposed a descriptive model of asymmetric proximity that incorporates similarity and bias [2]. It was later referred to as the *additive similarity and bias model* by Nosofsky [3]. In the additive similarity and bias model, the proximity of object i and j, P_{ij}, is given by

$$P_{ij} = F\left(S_{ij} + r_i + c_j\right), \tag{2}$$

where F is an increasing function, S_{ij} is a symmetric similarity function and r_i (row) and c_j (column) are bias functions on the individual objects.

Inspired by Tversky's work, Johannesson proposed more recently the *relative prominence model* [4]. In this model, the experienced directed similarity from objects i to j is proportional to some symmetric similarity measure between i and j, S_{ij}, and the quotient between the prominences for j, j_p, and i, i_p.

$$P_{ij} = S_{ij} \cdot \frac{j_p}{i_p}. \tag{3}$$

The similarity concept has found also wide applicability in chemistry [5], in particular for the searching of chemical databases [6]. Despite its widespread use, few efforts have focused on analyzing the implications that the asymmetric character of molecular similarity may have in this field. Those efforts have mainly focused on deriving and applying an asymmetric form of the Tanimoto coefficient to similarity searches of bit string-based molecular representations [7–9]. However, many molecular similarity studies rely on the use of molecular fields [10–12] and an entire family of similarity indices has been described to assess field-based molecular similarity [13, 14]. This contribution complements the analysis presented earlier on a family of field-based molecular similarity indices [14] by introducing the generalized asymmetric forms derived for the most commonly used similarity indices within this family, namely, the Carbó, Hodgkin–Richards, and Petke indices, as well as two other related indices.

2. Symmetric field-based similarity indices

In a recent work [14], an analysis of a family of field-based molecular similarity indices was presented. The indices of this family take the general form

$$S_{ij} = \frac{\Omega_{ij}}{\Delta_{ij}}, \tag{4}$$

where the similarity measure, Ω_{ij}, is given by the inner product

$$\Omega_{ij} = \int F_i(r) F_j(r) \mathrm{d}r \tag{5}$$

and F_i and F_j are field functions for the ith and jth molecules, respectively. The denominator, Δ_{ij}, is made up of a specific combination of self-similarities, Ω_{ii} and Ω_{jj}, and acts as a normalizing factor that defines a particular similarity index. The self-similarities always satisfy the relation $\Omega_{ii} > 0$ and thus the bounds of S_{ij} are determined by the nature of the field functions. For non-negative field functions, as is the case for steric fields, $S_{ij} \in [0, 1]$, the two limit values of 0 and 1 reflecting complete dissimilarity and identity, respectively. When field functions can take negative values, as is the case for electrostatic potentials, $S_{ij} \in [-1, 1]$, the limit value of -1 reflecting the situation of perfect complementarity [14].

The most commonly used similarity indices within this family are given in equations (6)–(8) and are usually referred to as the Carbó (C_{ij}) [15], Hodgkin–Richards (H_{ij}) [16], and Petke (P_{ij}) [17] indices, respectively,

$$C_{ij} = \frac{\Omega_{ij}}{(\Omega_{ii} \cdot \Omega_{jj})^{1/2}}, \tag{6}$$

$$H_{ij} = \frac{\Omega_{ij}}{\frac{1}{2}(\Omega_{ii} + \Omega_{jj})}, \tag{7}$$

$$P_{ij} = \frac{\Omega_{ij}}{\max(\Omega_{ii}, \Omega_{jj})}. \tag{8}$$

Two additional indices related to the Hodgkin–Richards (H^*_{ij}) and Petke (P^*_{ij}) indices were also defined in a recent publication [14] as

$$H^*_{ij} = \frac{\Omega_{ij}}{\left(\frac{1}{2}(\Omega^{-1}_{ii} + \Omega^{-1}_{jj})\right)^{-1}}, \tag{9}$$

$$P^*_{ij} = \frac{\Omega_{ij}}{\min(\Omega_{ii}, \Omega_{jj})}. \tag{10}$$

The denominator in each of the five indices corresponds to a particular form of average [18]. Specifically, the denominator in equations (6), (7), and (9) corresponds, respectively, to the geometric, arithmetic, and harmonic averages of the self-similarities Ω_{ii} and Ω_{jj}.

Two additional *relative comparability* indices can be defined, namely,

$$R_{i,j} = \frac{\Omega_{ij}}{\Omega_{ii}}, \tag{11}$$

$$R_{j,i} = \frac{\Omega_{ij}}{\Omega_{jj}}, \tag{12}$$

where $R_{i,j}$ is the relative comparability of the ith molecule to the jth molecule and $R_{j,i}$ is the relative comparability of the jth molecule to the ith molecule. Assuming Ω_{ij} is essentially the overlap between the ith and the jth molecules, if the field functions are non-negative definite and the self-similarities are representative of some measure of the size of the molecules, then $R_{i,j}$ and $R_{j,i}$ become the fraction of the ith molecule that is similar to the jth molecule and the fraction of the jth molecule that is similar to the ith molecule, respectively, being thus a reflection of the relative prominence of each molecule with respect to another. In Tversky's terms, $R_{i,j}$ would be the similarity of the ith molecule

in the context of the jth molecule, whereas $R_{j,i}$ would be the similarity of the jth molecule in the context of the ith molecule.

The Carbó, Hodgkin–Richards, and Petke indices and their related forms, as given in equations (6)–(10), can now be expressed as

$$C_{ij} = \left(R_{i,j} \cdot R_{j,i}\right)^{1/2}, \tag{13}$$

$$H_{ij} = \left[\tfrac{1}{2}\left(R_{i,j}^{-1} + R_{j,i}^{-1}\right)\right]^{-1}, \tag{14}$$

$$P_{ij} = R_{i,j}, \tag{15}$$

$$H_{ij}^{*} = \tfrac{1}{2}\left(R_{i,j} + R_{j,i}\right), \tag{16}$$

$$P_{ij}^{*} = R_{j,i}. \tag{17}$$

Interestingly, equations (13), (14), and (16) reveal that the original forms of the Carbó, Hodgkin–Richards, and the related Hodgkin–Richards indices correspond to the geometric, harmonic, and arithmetic averages of the relative comparability indices as defined in equations (11) and (12). On this basis, equations (13), (14), and (16) will be referred to as the *symmetric forms* of the Carbó, Hodgkin–Richards, and related Hodgkin–Richards similarity indices, respectively.

Without loss of generality, it can be assumed that

$$\Omega_{ii} \geqslant \Omega_{jj}. \tag{18}$$

The self-similarities of the ith and jth molecules are then related by the scale factor μ,

$$\Omega_{jj} = \mu \cdot \Omega_{ii}, \tag{19}$$

where $0 < \mu \leqslant 1$, and thus

$$R_{i,j} = \mu \cdot R_{j,i}. \tag{20}$$

Accordingly, the similarity indices given in equations (13)–(17) can be rewritten as

$$C_{ij} = \mu^{-1/2} \cdot R_{i,j}, \tag{21}$$

$$H_{ij} = \frac{2}{\mu + 1} \cdot R_{i,j}, \tag{22}$$

$$H_{ij}^{*} = \frac{\mu + 1}{2\mu} \cdot R_{i,j}, \tag{23}$$

$$P_{ij}^{*} = \mu^{-1} \cdot R_{i,j} \tag{24}$$

with P_{ij} being given directly by equation (15).

3. Asymmetric field-based similarity indices

The definition of the relative comparability indices in equations (11) and (12) insinuates already the asymmetric character of similarity. For any two molecules for which $\Omega_{ii} \neq \Omega_{jj}$, then $R_{i,j} \neq R_{j,i}$, which reveals that in general the relative comparabilities of molecules are asymmetric. Taking this concept further, the symmetric forms provided in equations (13), (14), and (16) can be generalized by differentially weighting the contribution of each molecule to the similarity index as

$$\breve{C}_{ij} = R_{i,j}^{1-w} \cdot R_{j,i}^{w}, \tag{25}$$

$$\breve{H}_{ij} = \left[(1 - w) \cdot R_{i,j}^{-1} + w \cdot R_{j,i}^{-1} \right]^{-1}, \tag{26}$$

$$\breve{H}_{ij}^{*} = (1 - w) \cdot R_{i,j} + w \cdot R_{j,i}, \tag{27}$$

where $w \in [0, 1]$ is a weighting factor of the relative comparability between the ith and jth molecules. In the limit case of $w = 0$, all indices converge and are equal to $R_{i,j}$. This corresponds to the case where all of the weight in the relative comparison of the two molecules being given to the ith molecule. In the other limit case of $w = 1$, all indices converge now to $R_{j,i}$, reflecting the situation where all of the weight is now given to the jth molecule. When $w = 1/2$, the original symmetric forms of the similarity indices are recovered. For this reason, equations (25)–(27) will be referred to as the *asymmetric forms* of the similarity indices.

Combining equations (25)–(27) with equations (13), (14), and (16), respectively, the following relationships between the symmetric and asymmetric forms of the Carbó and Hodgkin–Richards indices can be deduced

$$\breve{C}_{ij} = C_{ij} \cdot R_{j,i}^{w-1/2} \cdot R_{i,j}^{-(w-1/2)}, \tag{28}$$

$$\breve{H}_{ij} = \left[H_{ij}^{-1} + (w - 1/2) \cdot R_{j,i}^{-1} - (w - 1/2) \cdot R_{i,j}^{-1} \right]^{-1}, \tag{29}$$

$$\breve{H}^{*}_{ij} = H^{*}_{ij} + (w - 1/2) \cdot R_{j,i} - (w - 1/2) \cdot R_{i,j}. \tag{30}$$

At this stage, it is remarkable to realize that the asymmetric forms of the similarity indices defined by equations (28)–(30) are reminiscent of the asymmetric models of similarity proposed in other contexts [1–4]. In this respect, the asymmetric Carbó index, equation (28), takes the form of the relative prominence model proposed by Johannesson [4], equation (3), from which $j_p = R_{j,i}^{w-1/2}$ and $i_p = R_{i,j}^{w-1/2}$ reflect the relative prominences of j and i. Alternatively, the asymmetric related Hodgkin–Richards index, equation (30), is consistent with both the original contrast model proposed by Tversky [1], equation (1), in which $\theta = 1$, $\alpha = w - 1/2$, $\beta = 1/2 - w$, $f(i - j) = R_{i,j}$, and $f(j - i) = R_{j,i}$, and the proximity model proposed by Holman [2], equation (2), in which $r_i = (1/2 - w) \cdot R_{i,j}$ and $c_j = (w - 1/2) \cdot R_{j,i}$.

Considering equations (18)–(20) the asymmetric forms of the similarity indices are given by

$$\breve{C}_{ij} = \mu^{-w} \cdot R_{i,j}, \tag{31}$$

$$\breve{H}_{ij} = [1 + w \cdot (\mu - 1)]^{-1} \cdot R_{i,j}, \tag{32}$$

$$\breve{H}^{*}_{ij} = \left[1 + w \cdot (\mu^{-1} - 1)\right] \cdot R_{i,j}, \tag{33}$$

which, when combined with equations (21)–(23), yield the simplified relationships between the symmetric and asymmetric forms

$$\breve{C}_{ij} = C_{ij} \cdot \mu^{-(w-1/2)}, \tag{34}$$

$$\breve{H}_{ij} = H_{ij} \cdot \frac{\mu + 1}{2 \cdot [1 + w \cdot (\mu - 1)]}, \tag{35}$$

$$\breve{H}^{*}_{ij} = H^{*}_{ij} \cdot \frac{2 \cdot \left[1 + w \cdot (\mu^{-1} - 1)\right]}{\mu^{-1} + 1}. \tag{36}$$

4. Relationships among asymmetric similarity indices

It was demonstrated in a previous work [14] that the symmetric forms of the five similarity indices follow the ordering

$$P_{ij} \leqslant H_{ij} \leqslant C_{ij} \leqslant H^{*}_{ij} \leqslant P^{*}_{ij}. \tag{37}$$

Taking into consideration equations (15), (24), and (31)–(33), the following relationships can be derived

$$1 \leqslant [1 + w \cdot (\mu - 1)]^{-1} \leqslant \mu^{-w} \leqslant 1 + w \cdot (\mu^{-1} - 1) \leqslant \mu^{-1} \tag{38}$$

from which it can be deduced that, for a given w, the inequality defined in equation (37) also holds for the asymmetric forms of the similarity indices,

$$P_{ij} \leqslant \breve{H}_{ij} \leqslant \breve{C}_{ij} \leqslant \breve{H}^{*}_{ij} \leqslant P^{*}_{ij}. \tag{39}$$

The following analysis will focus on the relationship between the Carbó and the Hodgkin–Richards indices, since they are the most widely used indices in field-based molecular similarity. Taking the ratio of the original symmetric forms given in equations (21) and (22) yields

$$\frac{H_{ij}}{C_{ij}} = \frac{2\mu^{1/2}}{\mu + 1}. \tag{40}$$

The ratio goes to unity as $\mu \to 1$, that is both indices become equal. However, as $\mu \to 0$ the ratio approaches zero, that is the difference in value between the two indices becomes increasingly large.

Taking equations (31) and (32), the relationship between the asymmetric forms of these two indices is given by

$$\frac{\breve{H}_{ij}}{\breve{C}_{ij}} = \frac{\mu^{w}}{1 + w \cdot (\mu - 1)}. \tag{41}$$

The trend followed by this ratio upon varying the values of μ and w is illustrated in figure 1. Also indicated in figure 1 with a dotted line is the relationship that would arise from using the original symmetric forms as derived in equation (40). The general trend observed is that the larger the difference in size between the two molecules under comparison (that is, the smaller the μ), the larger the difference between the values obtained for these two similarity indices (that is, the smaller the value of their ratio). For $w = 0$ and $w = 1$, the two limit cases of asymmetry, the similarity indices converge to $\breve{C}_{ij} = \breve{H}_{ij} = R_{i,j}$ and $\breve{C}_{ij} = \breve{H}_{ij} = R_{j,i}$, respectively, and thus the ratio approaches unity. For $\mu \in (0, 1)$, the relationship defined in equation (41) reaches a minimum when

$$w = \frac{1}{\ln \mu} - \frac{1}{\mu - 1}. \tag{42}$$

Taking the limit as $\mu \to 1$, the weighting factor w approaches the value of 0.5, that is, as the fields of the molecules become more comparable, they tend to contribute equally to the similarity index and thus the asymmetric forms converge to the symmetric forms. Taking the limit as $\mu \to 0$, then w approaches unity, that is the similarity indices converge to $\breve{C}_{ij} = \breve{H}_{ij} = R_{j,i}$. It can also be noticed that this minimum deviates increasingly from the symmetric position ($w = 0.5$) as the difference between the self-similarities of the two molecules becomes more significant.

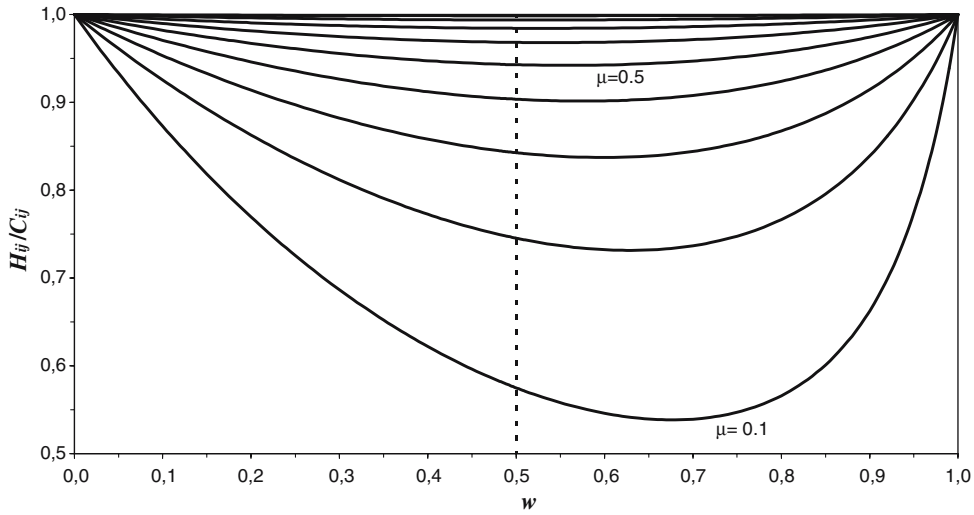

Figure 1. Relationship between the asymmetric forms of the Carbó and Hodgkin–Richards similarity indices, equation (41), upon varying the weighting factor (w) for different self-similarity ratios (μ).

5. Generalized asymmetric similarity index

A generalized similarity index was recently introduced [14] to define this family of similarity indices not as a discrete set of five indices but as a continuous set of indices of the general form

$$S_{ij}(\lambda) = \frac{\Omega_{ij}}{\Delta_{ij}(\lambda)}, \tag{43}$$

where the self-similarity normalization term in the denominator, Δ_{ij}, corresponds to a particular form of average [18] defined as

$$\Delta_{ij}(\lambda) = \left(\frac{\Omega_{ii}^{\lambda} + \Omega_{jj}^{\lambda}}{2} \right)^{1/\lambda}, \tag{44}$$

and the parameter λ can take values in the range of $-\infty \leqslant \lambda \leqslant \infty$. It can be demonstrated [14] that for the particular cases of $\lambda = 1$ and $\lambda = -1$, Δ_{ij} corresponds, respectively, to the denominator in the Hodgkin–Richards, equation (7), and the related Hodgkin–Richards, equation (9), similarity indices. As the limit of λ tends to 0, Δ_{ij} achieves the expression found in the Carbó index, equation (6). Finally, taking the limits as λ approaches ∞ and $-\infty$, one can recover, respectively, the Petke, equation (8), and related Petke, equation (10), similarity indices.

Arranging the generalized similarity index in equation (43) in terms of the two relative comparability indices defined in equations (11) and (12) gives

$$S_{ij}(\lambda) = \left[\tfrac{1}{2}\left(R_{i,j}^{-\lambda} + R_{j,i}^{-\lambda}\right)\right]^{-1/\lambda}. \tag{45}$$

By considering equations (18)–(20), equation (45) can be further simplified to

$$S_{ij}(\lambda) = \left(\frac{\mu^{\lambda}+1}{2}\right)^{-1/\lambda} \cdot R_{i,j} \tag{46}$$

from which the expressions for the five particular similarity indices given in equations (15) and (21)–(24) can be deduced.

Equation (45) will be referred to as the *symmetric form* of the generalized similarity index. The incorporation of a weighted contribution, w, of each molecule results in

$$\breve{S}_{ij}(\lambda) = \left[(1-w)\cdot R_{i,j}^{-\lambda} + w \cdot R_{j,i}^{-\lambda}\right]^{-1/\lambda}, \tag{47}$$

which will be referred to as the *asymmetric form* of the generalized similarity index. By taking into consideration of equations (18)–(20), equation (47) leads to

$$\breve{S}_{ij}(\lambda) = \left[1 + w \cdot (\mu^{\lambda} - 1)\right]^{-1/\lambda} \cdot R_{i,j} \tag{48}$$

from which the asymmetric forms of the Carbó and Hodgkin–Richards similarity indices given in equations (31)–(33) can be deduced.

Finally, by combining equations (46) and (48), the relationship between the symmetric and asymmetric forms of the generalized similarity index can be obtained

$$\breve{S}_{ij}(\lambda) = S_{ij}(\lambda) \cdot \left(\frac{2\cdot\left[1 + w\cdot(\mu^{\lambda}-1)\right]}{\mu^{\lambda}+1}\right)^{-1/\lambda}. \tag{49}$$

6. Summary and conclusions

Means to quantify molecular similarity are widely and regularly applied to assess the similarity of one molecule to another. These measures are traditionally of an intrinsic symmetric nature, implying that relevant information related to the salient features present in each molecule when compared to another may be lost in the process [5, 8]. Therefore, additional means with an inherent asymmetric character are required to extract the individual contributions of the molecules under comparison and allow measuring and modulating the similarity of one molecule in the context of another.

The present work introduces a generalized asymmetric similarity index covering both the symmetric and asymmetric forms of a family of field-based similarity indices that include some of the most widely used indices, namely, Carbó, Hodgkin–Richards, and Petke indices, as well as two other related indices. The newly obtained asymmetric forms of the similarity indices incorporate a weighted contribution of each molecule into the similarity index in a manner consistent with some of the asymmetric models derived in other contexts [1–4]. An open question at this stage is the most appropriate form of the weighting parameter introduced in the various asymmetric similarity indices. This would be the focus of future research in this direction.

Acknowledgment

This research is supported by a grant from the Instituto de Salud Carlos III (Ministerio de Sanidad y Consumo), research project reference number 02/3051.

References

[1] A. Tversky, Features of similarity, Psychol. Rev. 84 (1977) 327–352.
[2] E.W. Holman, Monotonic models for asymmetric proximities, J. Math. Psychol. 20 (1979) 1–15.
[3] R.M. Nosofsky, Stimulus bias, asymmetric similarity, and classification, Cogn. Psychol. 23 (1991) 91–140.
[4] M. Johannesson, Modelling asymmetric similarity with prominence, Br. J. Math. Stat. Psychol. 53 (2000) 121–139.
[5] G.M. Maggiora and V. Shanmugasundaram, Molecular Similarity Measures, in *Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery*, ed. J. Bajorath, Methods Mol Biol., vol. 275, pp. 1–50.
[6] P. Willett, Chemical similarity searching, J. Chem. Inf. Comput. Sci. 38 (1998) 983–996.
[7] J. Bradshaw, Introduction to Tversky similarity measure. Presented at the 11th Annual Daylight MUG Meeting, Laguna Beach (CA), February 1997. http://www.daylight.com/meetings/mug97/Bradshaw/MUG97/tv_tversky.html.
[8] G.M. Maggiora, J. Mestres, T.R. Hagadone and M.S. Lajiness, Asymmetric similarity and molecular diversity, Presented at the 213th National Meeting of the American Chemical Society, San Francisco (CA), April 1997.
[9] J.D. Holliday, N. Salim, M. Whittle and P. Willett, Analysis and display of the size dependence of chemical similarity coefficients, J. Chem. Inf. Comput. Sci. 43 (2003) 819–828.
[10] R.D. Cramer III, D.E. Patterson and J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[11] G. Klebe, U. Abraham and T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem. 37 (1994) 4130–4146.
[12] C. Lemmen and T. Lengauer, Computational methods for the structural alignment of molecules, J. Comput. Aided Mol. Design 14 (2000) 215–232.
[13] R. Carbó, E. Besalú, L. Amat and X. Fradera, On quantum molecular similarity measures (QMSM) and indices (QMSI), J. Math. Chem. 19 (1996) 47–56.

[14] G.M. Maggiora, J.D. Petke and J. Mestres, A general analysis of field-based molecular similarity indices, J. Math. Chem. 31 (2002) 251–270.
[15] R. Carbó, L. Leyda and M. Arnau, How similar is a molecule to another? An electron density measure of similarity between two molecular structures, Intl. J. Quantum Chem. 17 (1980) 1185–1189.
[16] E.E. Hodgkin and W.G. Richards, Molecular similarity based on electrostatic potential and electric field, Intl. J. Quantum Chem. Quantum Biol. Symp. 14 (1987) 105–110.
[17] J.D. Petke, Cumulative and discrete similarity analysis of electrostatic potentials and fields, J. Comput. Chem. 14 (1993) 928–933.
[18] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic – Theory and Applications* (Prentice-Hall, Upper Saddle River, NJ, 1995) Chapter 3.