Journal of Mathematical Chemistry Vol. 31, No. 3, April 2002 (© 2002)

A general analysis of field-based molecular similarity indices

Gerald M. Maggiora *, James D. Petke and Jordi Mestres

Computer-Aided Drug Discovery, Pharmacia, 301 Henrietta Street, Kalamazoo, MI 49007, USA

Received 3 January 2002

A formalism is presented that incorporates the entirety of all field-based molecular similarity indices of general form $S_{ij} = \Omega_{ij}/h(\Omega_{ii}, \Omega_{jj})$, where the numerator is given by the inner product or "overlap" of field functions F_i and F_j corresponding to the ith and jth molecules, respectively, and the denominator is given by a suitable mean function of the self-similarities Ω_{ii} and Ω_{jj}. This family of similarity indices includes the index initially introduced by Carbó nearly twenty years ago, where $h(\Omega_{ii}, \Omega_{jj})$ is taken to be the geometric mean of Ω_{ii} and Ω_{jj}, and the well-known indices due to Hodgkin and Richards, and Petke, where $h(\Omega_{ii}, \Omega_{jj})$ is taken to be the arithmetic mean and maximum of Ω_{ii} and Ω_{jj}, respectively. Two new indices based upon the harmonic mean and minimum of Ω_{ii} and Ω_{jj} are also defined, and it is demonstrated that the entire set of field-based similarity indices can be generated from a one-parameter family of functions, called generalized means, through proper choice of the parameter value and suitable limiting procedures. Ordering and rigorous bounds for all of the indices are described as well as a number of inter-relationships among the indices. The generalization of field-based similarity indices, coupled with the relationships among indices that have been developed in the present work, place the basic theory of these indices on a more unified and mathematically rigorous footing that provides a foundation for a better understanding of the quantitative aspects of field-based molecular similarity.

1. Introduction

Similarity is a ubiquitous concept that permeates virtually every field of chemistry [1–3]. While there are many ways in which similarity is utilized in chemistry, the focus here is on molecular similarity, or more specifically, on the question of how similar one molecule is to another. Numerous techniques have been developed and applied to this problem [4–63]. The present work is concerned with methodology that is based on similarity indices computed from molecular fields. Early work by Carbó and co-workers [17,18] defined similarity in terms of the electron densities, ρ_i and ρ_j, of the two molecules being compared. The matching function, or *similarity measure*,

* Corresponding author.

0259-9791/02/0400-0251/0

employed is the inner product of ρ_i and ρ_j,

$$\int \rho_i \rho_j \, \mathrm{d}V$$

which measures the "overlap" of the electron densities in terms of their relative positions and orientations. Carbó et al. obtained numerical values of the similarity from a *normalized similarity measure* or *similarity index*

$$\frac{\int \rho_i \rho_j \, \mathrm{d}V}{\sqrt{\int \rho_i^2 \, \mathrm{d}V} \cdot \sqrt{\int \rho_j^2 \, \mathrm{d}V}}$$

which is given as the ratio of the similarity measure to the product of the norms of the respective electron densities, i.e., *self-similarity measures*.

In the Carbó approach, the electron density surrounding each molecule may be viewed as a *steric* field that describes the size and shape of the molecule [19]. Comparing two molecules in this way gives their *steric similarity*. Subsequently, a computationally more tractable approach to steric similarity was outlined by Good and Richards [20], who proposed a steric field composed of atom-centered spherical Gaussian functions to represent individual atomic densities. Interestingly, many of the early applications of field-based similarity that followed Carbó's seminal work dealt solely with *electrostatic similarity* based upon molecular electrostatic potential (MEP) fields [21–29]. Initially, grid-based methods were used. In these methods, the MEP at the kth grid point, $\mathbf{r}_k$, for a given molecule is approximated by the classical point-charge formula

$$\sum_m \frac{q_m}{\mathbf{r}_k - \mathbf{R}_m},$$

where q_m is the partial charge of the mth atom located at $\mathbf{R}_m$. The electrostatic similarity measure is then computed as a sum over grid points of the product of MEPs of the molecules being compared.[1] Currently, a more computationally efficient formulation of this approach is based upon an approximation of $1/(\mathbf{r}_k - \mathbf{R}_m)$ by a linear combination of spherical Gaussian functions, and replacement of the sum over grid points by integration of the product of MEPs over all space [30].

Other molecular fields or pseudo-fields, such as lipophilic potential fields [31–37], have been used in addition to the steric and electrostatic fields noted above. In addition, a number of similarity indices of the same general form as that of the Carbó index have been defined by modifying the manner in which the self-similarities in the de-

[1] A similarity index in which the similarity measure is computed as a sum over grid points or an integral over all space of the product of molecular field functions is known as a *cumulative* similarity index. An alternative approach to field-based similarity employs a *discrete* similarity index, which gives the similarity at individual grid points in space. The overall similarity is then computed as the average of the discrete values. The theory developed in the present paper applies only to cumulative indices in which the similarity measure is given as an integral. For more information on discrete similarity indices, see [14,15,28].

nominator are combined. In the Carbó index, for example, the denominator is given by the *geometric mean* of the two self-similarities, $\sqrt{\int \rho_i^2 \, \mathrm{d}V} \cdot \sqrt{\int \rho_j^2 \, \mathrm{d}V}$. Two additional indices in use today are those defined by Hodgkin and Richards [21] and by Petke [28]. The respective denominators of these indices are based upon the *arithmetic mean*, $\frac{1}{2}(\int \rho_i^2 \, \mathrm{d}V + \int \rho_j^2 \, \mathrm{d}V)$, and *maximum*, $\max(\int \rho_i^2 \, \mathrm{d}V, \int \rho_j^2 \, \mathrm{d}V)$, of the self-similarities of the ith and jth molecules. Carbó et al. [38–41] have also defined more elaborate "quantum" similarity measures and indices, which are defined in terms of integrals involving electron densities and assorted quantum mechanical operators. Alternative approaches used to determine similarity include those based on the distance between molecular field functions [42,43], or more general forms of similarity indices utilizing set based [44] or graph theoretic [45] approaches, which will not be considered further in this work.

Numerous applications of steric and electrostatic field-based similarity have been published in the last several years. Many involve determination of an optimal molecular superposition that maximizes the value of the similarity index as a function of the relative positions and orientations of the molecules being compared. This is often carried out by simultaneously maximizing both steric and electrostatic similarity using a weighted sum of steric and electrostatic similarity indices [46–54]. Other studies involve computing the similarity of a series of overlaid structures for use as descriptors in 3D QSAR studies [55–63]. In both cases, numerical values of similarity indices have been employed as data, without a fundamental understanding of the mathematical properties of the indices used.

The present work describes a general theory of a class of field-based molecular similarity indices that includes those given by Carbó, Hodgkin and Richards, and Petke. Rigorous bounds, inequalities, and other mathematical relationships among the indices are developed. Additionally, a more general form of similarity index, based upon a one-parameter family of functions called generalized means [64], is derived and fully characterized. This work serves to elucidate the fundamental characteristics of field-based similarity indices, and to put them on a mathematically rigorous footing. Moreover, the analysis provided here should also lead to a clearer understanding of the numerical properties of field-based similarity.

2. Field-based similarity indices

2.1. Overview of similarity indices

The general form taken by molecular field-based *similarity indices* considered in the present analysis is the ratio

$$S_{ij} = \frac{\Omega_{ij}}{\Delta_{ij}}, \tag{1}$$

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

$$\Omega_{ij} = \int F_i(\mathbf{r}) \cdot F_j(\mathbf{r})\, \mathrm{d}^3\mathbf{r} \tag{2}$$

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 the *self-similarities*, Ω_{ii} and Ω_{jj}, and defines a particular similarity index. The self-similarities, which are just the squared norms of the field functions, always satisfy the relation $\Omega_{ii} > 0$, and therefore the sign of S_{ij} is determined by that of Ω_{ij}. For the material presented in sections 2.1–2.5, it is assumed that the field functions are non-negative, as is the case for steric fields, and thus that $\Omega_{ij} \geqslant 0$. The case in which Ω_{ij} may take on negative values, as may occur, for example, in MEP fields, is considered in section 2.6.

Three of the indices most commonly used in field-based similarity calculations are those developed by Carbó et al. [17], Hodgkin and Richards [21], and Petke [28], and are given respectively by

$$C_{ij} = \frac{\Omega_{ij}}{(\Omega_{ii} \cdot \Omega_{jj})^{1/2}}, \tag{3}$$

$$H_{ij} = \frac{\Omega_{ij}}{\frac{1}{2}(\Omega_{ii} + \Omega_{jj})}, \tag{4}$$

$$P_{ij} = \frac{\Omega_{ij}}{\max(\Omega_{ii}, \Omega_{jj})}, \tag{5}$$

where the denominators correspond, respectively, to the *geometric mean*, the *arithmetic mean*, and the *maximum* of the self-similarities Ω_{ii} and Ω_{jj}. All three indices are bounded by zero and unity (see section 2.4 for further discussion), i.e.

$$0 \leqslant C_{ij}, H_{ij}, P_{ij} \leqslant 1. \tag{6}$$

As will be shown in section 3, a whole family of field-based similarity indices can be defined simply by modifying the denominator appropriately. Along this line, a new index based upon the *harmonic mean* of Ω_{ii} and Ω_{jj} is given by

$$H^*_{ij} = \frac{\Omega_{ij}}{\dfrac{1}{\frac{1}{2}(1/\Omega_{ii} + 1/\Omega_{jj})}}. \tag{7}$$

Moreover, by choosing Δ_{ij} as $\min(\Omega_{ii}, \Omega_{jj})$, an alternative to the usual Petke similarity index P_{ij} is obtained, i.e.,

$$P^*_{ij} = \frac{\Omega_{ij}}{\min(\Omega_{ii}, \Omega_{jj})}, \tag{8}$$

which plays an important role in the subsequent analysis. From equations (5) and (8) it is clear that

$$P_{ij} \leqslant P^*_{ij}. \tag{9}$$

As will be discussed in detail in section 2.4, both H_{ij}^* and P_{ij}^* are bounded from below by zero but are unbounded from above.

Additionally, it should be noted that all of the above indices are symmetric since both the numerator and denominator of equation (1) are symmetric, i.e., $\Omega_{ij} = \Omega_{ji}$ and $\Delta_{ij} = \Delta_{ji}$. The former follows from the symmetry of the inner product and the latter from the symmetry of the appropriate mean.

2.2. *Ordering of similarity indices*

Without loss of generality it can be assumed that

$$\Omega_{ii} \geqslant \Omega_{jj}. \tag{10}$$

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

$$\Omega_{jj} = \mu \cdot \Omega_{ii}, \tag{11}$$

where $0 < \mu \leqslant 1$.

A *fundamental* ordering of the different similarity indices is obtained when the value of the numerator, Ω_{ij}, in all of the indices is the same. In such cases, the ordering is determined solely by the terms in the denominator. Substituting equation (11) into the denominators of the set of similarity indices P_{ij}^*, H_{ij}^*, C_{ij}, H_{ij} and P_{ij} yields

$$P_{ij}^* \Longleftrightarrow \min(\Omega_{ii}, \Omega_{jj}) = \Omega_{jj} = \mu \cdot \Omega_{ii}, \tag{12}$$

$$H_{ij}^* \Longleftrightarrow \frac{1}{\frac{1}{2}(\Omega_{ii}^{-1} + \Omega_{jj}^{-1})} = \frac{2\mu}{1+\mu} \cdot \Omega_{ii}, \tag{13}$$

$$C_{ij} \Longleftrightarrow (\Omega_{ii} \cdot \Omega_{jj})^{1/2} = \sqrt{\mu} \cdot \Omega_{ii}, \tag{14}$$

$$H_{ij} \Longleftrightarrow \frac{1}{2}(\Omega_{ii} + \Omega_{jj}) = \frac{1+\mu}{2} \cdot \Omega_{ii}, \tag{15}$$

$$P_{ij} \Longleftrightarrow \max(\Omega_{ii}, \Omega_{jj}) = 1 \cdot \Omega_{ii}. \tag{16}$$

Since $0 < \mu \leqslant 1$, the following inequality holds:

$$\mu \leqslant \frac{2\mu}{1+\mu} \leqslant \sqrt{\mu} \leqslant \frac{1+\mu}{2} \leqslant 1, \tag{17}$$

which is clear from figure 1.

Multiplying all terms by Ω_{ii} yields for the respective denominators

$$\mu \cdot \Omega_{ii} \leqslant \frac{2\mu}{1+\mu} \cdot \Omega_{ii} \leqslant \sqrt{\mu} \cdot \Omega_{ii} \leqslant \frac{1+\mu}{2} \cdot \Omega_{ii} \leqslant 1 \cdot \Omega_{ii}, \tag{18}$$

so that the following inequality holds for the five similarity indices

$$P_{ij} \leqslant H_{ij} \leqslant C_{ij} \leqslant H_{ij}^* \leqslant P_{ij}^*. \tag{19}$$

Equality obtains when $\Omega_{ii} = \Omega_{jj} = \Omega_{ij}$.

2.3. Relationships among similarity indices

Using equation (11), a variety of relationships among the different similarity indices can be derived as functions of μ and specific similarity indices: $S_{ij} = f(\mu, X_{ij})$, where S_{ij} and $X_{ij} = P_{ij}, H_{ij}, C_{ij}, H^*_{ij}$, or P^*_{ij}. Table 1 summarizes these relationships with respect to $P_{ij}, H_{ij}, C_{ij}, H^*_{ij}$ and P^*_{ij}, five indices which have special significance in terms of the generalized means formalism described in section 3.

Equations (u)–(y) of table 1 can be rearranged as ratios with respect to P^*_{ij}, i.e.,

$$\frac{P_{ij}}{P^*_{ij}} : \frac{H_{ij}}{P^*_{ij}} : \frac{C_{ij}}{P^*_{ij}} : \frac{H^*_{ij}}{P^*_{ij}} : \frac{P^*_{ij}}{P^*_{ij}} = \mu : \frac{2\mu}{1+\mu} : \sqrt{\mu} : \frac{1+\mu}{2} : 1. \tag{20}$$

Interestingly, these ratios yield functions of μ that are identical to those given in equation (17) and depicted in figure 1. Thus, the behavior of these ratios is clearly illustrated in the figure. It should also be noted that since the ratios are *independent* of Ω_{ij} they depend only on the self-similarities.

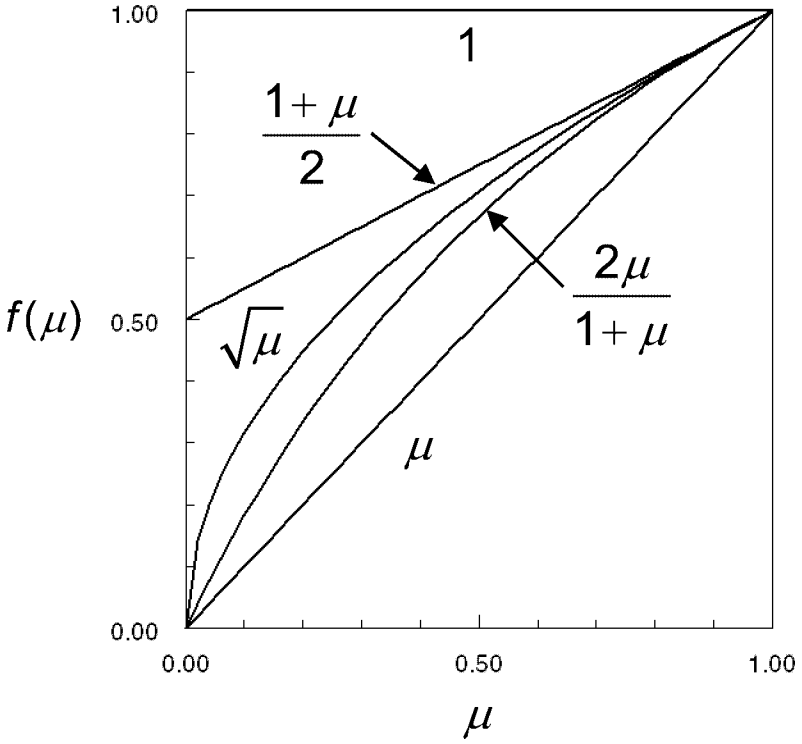

Figure 1. Ratios of the various simlarity indices with respect to P^*_{ij}, as given by equation (20).

Table 1
Relationships among similarity indices.

$S_{ij} = f(\mu, P_{ij})$	$S_{ij} = f(\mu, H_{ij})$	$S_{ij} = f(\mu, C_{ij})$	$S_{ij} = f(\mu, H^*_{ij})$	$S_{ij} = f(\mu, P^*_{ij})$
(a) $P_{ij} = 1 \cdot P_{ij}$	(f) $P_{ij} = \frac{1+\mu}{2} \cdot H_{ij}$	(k) $P_{ij} = \sqrt{\mu} \cdot C_{ij}$	(p) $P_{ij} = \frac{2\mu}{1+\mu} \cdot H^*_{ij}$	(u) $P_{ij} = \mu \cdot P^*_{ij}$
(b) $H_{ij} = \frac{2}{1+\mu} \cdot P_{ij}$	(g) $H_{ij} = 1 \cdot H_{ij}$	(l) $H_{ij} = \frac{2\sqrt{\mu}}{1+\mu} \cdot C_{ij}$	(q) $H_{ij} = \left(\frac{2\sqrt{\mu}}{1+\mu}\right)^2 \cdot H^*_{ij}$	(v) $H_{ij} = \frac{2\mu}{1+\mu} \cdot P^*_{ij}$
(c) $C_{ij} = \frac{1}{\sqrt{\mu}} \cdot P_{ij}$	(h) $C_{ij} = \frac{1+\mu}{2\sqrt{\mu}} \cdot H_{ij}$	(m) $C_{ij} = 1 \cdot C_{ij}$	(r) $C_{ij} = \frac{2\sqrt{\mu}}{1+\mu} \cdot H^*_{ij}$	(w) $C_{ij} = \sqrt{\mu} \cdot P^*_{ij}$
(d) $H^*_{ij} = \frac{1+\mu}{2\mu} \cdot P_{ij}$	(i) $H^*_{ij} = \left(\frac{1+\mu}{2\sqrt{\mu}}\right)^2 \cdot H_{ij}$	(n) $H^*_{ij} = \frac{1+\mu}{2\sqrt{\mu}} \cdot C_{ij}$	(s) $H^*_{ij} = 1 \cdot H^*_{ij}$	(x) $H^*_{ij} = \frac{1+\mu}{2} \cdot P^*_{ij}$
(e) $P^*_{ij} = \frac{1}{\mu} \cdot P_{ij}$	(j) $P^*_{ij} = \frac{1+\mu}{2\mu} \cdot H_{ij}$	(o) $P^*_{ij} = \frac{1}{\sqrt{\mu}} \cdot C_{ij}$	(t) $P^*_{ij} = \frac{2}{1+\mu} \cdot H^*_{ij}$	(y) $P^*_{ij} = 1 \cdot P^*_{ij}$

2.4. Bounds on similarity indices

Although the inequalities presented in equation (19) establish an ordering of the similarity indices, rigorous bounds have not been determined for all of these quantities. To do this, the Schwartz inequality [65], which can be written in terms of the similarity measure and self-similarities as

$$|\Omega_{ij}|^2 \leqslant \Omega_{ii} \cdot \Omega_{jj}, \tag{21}$$

is used. Recalling that $\Omega_{ij} \geqslant 0$ and taking the square root of both sides of equation (21) yields, upon rearranging, the upper bound for the Carbó index

$$C_{ij} = \frac{\Omega_{ij}}{\sqrt{\Omega_{ii}} \cdot \sqrt{\Omega_{jj}}} \leqslant 1. \tag{22}$$

Thus, it follows that

$$0 \leqslant C_{ij} \leqslant 1, \tag{23}$$

where the lower bound occurs when $\Omega_{ij} = 0$. Combining the latter expression with equation (19) leads to the overall relationship

$$0 \leqslant P_{ij} \leqslant H_{ij} \leqslant C_{ij} \leqslant 1. \tag{24}$$

As noted earlier, both H^*_{ij} and P^*_{ij} are bounded from below by zero and are unbounded from above. To prove that H^*_{ij} is unbounded from above, it can be shown from equations (l) and (n) in table 1 that

$$H_{ij} \cdot H^*_{ij} = C^2_{ij} \leqslant 1. \tag{25}$$

Rearranging the terms in equation (25) gives

$$H^*_{ij} = C_{ij} \cdot \left(\frac{C_{ij}}{H_{ij}}\right). \tag{26}$$

Upon substitution of equation (h) from table 1, equation (26) may be rewritten as

$$H^*_{ij} = C_{ij} \cdot \left(\frac{1+\mu}{2\sqrt{\mu}}\right), \tag{27}$$

which, using the expression given in equation (23), leads to the inequality

$$H^*_{ij} \leqslant \frac{1+\mu}{2\sqrt{\mu}}. \tag{28}$$

Taking the limit as $\mu \to 0$ yields

$$\lim_{\mu\to 0} H^*_{ij} \leqslant \lim_{\mu\to 0} \left(\frac{1+\mu}{2\sqrt{\mu}}\right) \to \infty, \tag{29}$$

which shows that H_{ij}^* is indeed unbounded. Since $H_{ij}^* \leqslant P_{ij}^*$, as shown in equation (19), it follows that P_{ij}^* is also unbounded, i.e.,

$$P_{ij}^* \leqslant \infty, \tag{30}$$

so that

$$0 \leqslant H_{ij}^* \leqslant P_{ij}^* \leqslant \infty. \tag{31}$$

A more general treatment of the bounds of similarity indices of the form given by equation (1) is presented in section 3.

2.5. *Similarity indices as means of other similarity indices*

An examination of the relationships given in table 1 suggests that other relationships may also exist among the various similarity indices. For example, consider the geometric mean of P_{ij} and P_{ij}^*, which with equation (e) from table 1 may be written as

$$\left(P_{ij} \cdot P_{ij}^*\right)^{1/2} = \left(P_{ij} \cdot \frac{1}{\mu} P_{ij}\right)^{1/2}, \tag{32}$$

or

$$\left(P_{ij} \cdot P_{ij}^*\right)^{1/2} = \frac{1}{\sqrt{\mu}} \cdot P_{ij}. \tag{33}$$

Substituting equation (c) from table 1 into equation (33) yields the desired relation

$$\left(P_{ij} \cdot P_{ij}^*\right)^{1/2} = C_{ij}. \tag{34}$$

Now consider the arithmetic mean. Again employing equation (e) from table 1 yields

$$\frac{1}{2}\left(P_{ij} + P_{ij}^*\right) = \frac{1}{2}\left(P_{ij} + \frac{1}{\mu} P_{ij}\right), \tag{35}$$

or

$$\frac{1}{2}\left(P_{ij} + P_{ij}^*\right) = \left(\frac{1+\mu}{2\mu}\right) \cdot P_{ij}, \tag{36}$$

which upon substitution of equation (d) from table 1 gives

$$\frac{1}{2}\left(P_{ij} + P_{ij}^*\right) = H_{ij}^*. \tag{37}$$

Lastly, consider the harmonic mean of P_{ij} and P_{ij}^*, which with equation (e) from table 1 may be expressed as

$$\frac{1}{\frac{1}{2}(1/P_{ij} + 1/P_{ij}^*)} = \frac{1}{\frac{1}{2}(1/P_{ij} + \mu/P_{ij})}, \tag{38}$$

or

$$\frac{1}{\frac{1}{2}(1/P_{ij} + 1/P^*_{ij})} = \left(\frac{2}{1+\mu}\right) \cdot P_{ij}. \tag{39}$$

Upon substitution of equation (b) from table 1 the desired relation

$$\frac{1}{\frac{1}{2}(1/P_{ij} + 1/P^*_{ij})} = H_{ij} \tag{40}$$

is obtained. It is interesting to note that, unlike the relationships listed in table 1, the "means" relations given by equations (34), (37), and (40) are independent of μ. Moreover, the relations illustrated in this section suggest the possibility of other relations among these similarity indices, which may also be of interest such as the one shown in equation (25) relating the product of H_{ij} and H^*_{ij} to C_{ij}. Further developments on this topic will be covered in section 3.2.

2.6. Considerations for the case of negative similarity

In order to avoid confusion in the development of relationships for bounds and inequalities, the previous discussion was restricted to the case in which the value of Ω_{ij}, and thus the values of all of the similarity indices discussed up to now are greater than zero. In this section the case where $\Omega_{ij} < 0$ is treated specifically. Relationships presented in the previous sections that do not hold are identified and alternative relationships are provided where appropriate. In this connection, it should be noted that the "pairwise" relationships given in table 1 are valid for both positive and negative similarity.

In sections 2.1 and 2.2, the inequalities given by equations (6), (9), and (19) are invalid when $\Omega_{ij} < 0$. However, for the case of negative similarity, applying the same arguments used to derive equation (19) yield an expression *with the order of terms reversed*,

$$P_{ij} \geqslant H_{ij} \geqslant C_{ij} \geqslant H^*_{ij} \geqslant P^*_{ij}. \tag{41}$$

The upper and lower bounds given previously in section 2.4 are also invalid, but rigorous bounds for negative similarity may be obtained. Using the Schwartz inequality, equation (21), we may write

$$|\Omega_{ij}| \leqslant \sqrt{\Omega_{ii}} \cdot \sqrt{\Omega_{jj}}, \tag{42}$$

which upon rearranging terms gives

$$\frac{|\Omega_{ij}|}{\sqrt{\Omega_{ii}} \cdot \sqrt{\Omega_{jj}}} \leqslant 1. \tag{43}$$

Since $\Omega_{ij} < 0$, $|\Omega_{ij}| = -\Omega_{ij}$, so that

$$\frac{-\Omega_{ij}}{\sqrt{\Omega_{ii}} \cdot \sqrt{\Omega_{jj}}} \leqslant 1, \tag{44}$$

which, in turn, leads to the lower bound for the Carbó index

$$C_{ij} = \frac{\Omega_{ij}}{\sqrt{\Omega_{ii}} \cdot \sqrt{\Omega_{jj}}} \geqslant -1. \tag{45}$$

Combining this result with the inequality given by equation (41) yields the relation

$$-1 \leqslant C_{ij} \leqslant H_{ij} \leqslant P_{ij} \leqslant 0. \tag{46}$$

The procedure used to determine the lower bound for H^*_{ij} when $\Omega_{ij} < 0$ is identical to that employed in section 2.4. Combining equation (27) with equation (46) leads to the inequality

$$H^*_{ij} \geqslant -\left(\frac{1+\mu}{2\sqrt{\mu}}\right). \tag{47}$$

Taking the limit of equation (47) as $\mu \to 0$ gives

$$\lim_{\mu\to 0} H^*_{ij} \geqslant \lim_{\mu\to 0}\left(-\frac{1+\mu}{2\sqrt{\mu}}\right) \to -\infty, \tag{48}$$

showing that H^*_{ij} is unbounded from below. Therefore, since according to equation (41) $P^*_{ij} \leqslant H^*_{ij}$, it follows that

$$-\infty \leqslant P^*_{ij} \leqslant H^*_{ij} \leqslant 0. \tag{49}$$

Importantly, the expression for the geometric mean of P_{ij} and P^*_{ij} derived from equations (32)–(34) does not apply in the case of negative similarity. However, equation (34) may be replaced by the expression

$$\left(P_{ij} \cdot P^*_{ij}\right)^{1/2} = |C_{ij}|, \tag{50}$$

which holds for both positive and negative similarity. The other expressions for the arithmetic and harmonic means of P_{ij} and P^*_{ij} given by equations (37) and (40), respectively, are valid for both positive and negative similarity.

3. Generalization of Δ_{ij}

3.1. Generalized and specific similarity indices

As noted in section 2, it is the form of the self-similarity normalization terms in the denominator, Δ_{ij}, that distinguish the various similarity indices from one another. In the following it is shown that a generalization of Δ_{ij}, which is based upon a *generalized means* function [64],

$$\Delta_{ij}(\lambda) = \left(\frac{\Omega^\lambda_{ii} + \Omega^\lambda_{jj}}{2}\right)^{1/\lambda} \tag{51}$$

can be used to define a *one-parameter family* of similarity indices,

$$S_{ij}(\lambda) = \frac{\Omega_{ij}}{\Delta_{ij}(\lambda)}. \tag{52}$$

Moreover, through appropriate choices of the parameter λ and suitable limiting procedures, expressions for the denominators used in the five similarity indices discussed in section 2 may be obtained. The easiest to derive is the denominator of Hodgkin–Richards index H_{ij}, which can be obtained simply by setting $\lambda = 1$,

$$\Delta_{ij}(1) = \frac{\Omega_{ii} + \Omega_{jj}}{2}. \tag{53}$$

Similarly, taking $\lambda = -1$ yields the denominator of the similarity index H^*_{ij},

$$\Delta_{ij}(-1) = \frac{1}{\frac{1}{2}(\Omega^{-1}_{ii} + \Omega^{-1}_{jj})}. \tag{54}$$

More sophisticated limiting processes are needed to obtain the remaining denominators. For example, to generate the "Carbó denominator" requires taking the limit as $\lambda \to 0$ of equation (51). To accomplish this, we first take the natural logarithm of both sides of equation (51) and then the limit as $\lambda \to 0$

$$\Delta_{ij}(0) = \lim_{\lambda\to 0} \Delta_{ij}(\lambda), \tag{55}$$

$$\lim_{\lambda\to 0} \ln \Delta_{ij}(\lambda) = \lim_{\lambda\to 0} \frac{\ln(\Omega^{\lambda}_{ii} + \Omega^{\lambda}_{jj}) - \ln 2}{\lambda}. \tag{56}$$

Using L'Hospital's rule yields

$$\lim_{\lambda\to 0} \ln \Delta_{ij}(\lambda) = \lim_{\lambda\to 0} \frac{\Omega^{\lambda}_{ii} \ln \Omega_{ii} + \Omega^{\lambda}_{jj} \ln \Omega_{jj}}{\Omega^{\lambda}_{ii} + \Omega^{\lambda}_{jj}} \tag{57}$$

which gives upon taking the limit

$$\lim_{\lambda\to 0} \ln \Delta_{ij}(\lambda) = \frac{\ln \Omega_{ii} + \ln \Omega_{jj}}{2} = \frac{\ln(\Omega_{ii} \cdot \Omega_{jj})}{2} = \ln(\Omega_{ii} \cdot \Omega_{jj})^{1/2} \tag{58}$$

finally yielding the desired relationship

$$\lim_{\lambda\to 0} \Delta_{ij}(\lambda) = (\Omega_{ii} \cdot \Omega_{jj})^{1/2}. \tag{59}$$

Determining the two denominators employed in the similarity indices P_{ij} and P^*_{ij} involves taking the limits as $\lambda \to \infty$ and $\lambda \to -\infty$, respectively. Thus, for P_{ij}

$$\Delta_{ij}(\infty) = \lim_{\lambda\to\infty} \Delta_{ij}(\lambda), \tag{60}$$

or

$$\Delta_{ij}(\infty) = \lim_{\lambda\to\infty} \left(\frac{\Omega^{\lambda}_{ii} + \Omega^{\lambda}_{jj}}{2} \right)^{1/\lambda}. \tag{61}$$

Since it can be assumed without loss of generality that $\Omega_{ii} \geqslant \Omega_{jj}$, equation (11) can be substituted into the above equation yielding

$$\Delta_{ij}(\infty) = \lim_{\lambda\to\infty} \left(\frac{\Omega_{ii}^{\lambda} + \mu^{\lambda} \cdot \Omega_{ii}^{\lambda}}{2} \right)^{1/\lambda} \tag{62}$$

which upon rearrangement gives

$$\Delta_{ij}(\infty) = \lim_{\lambda\to\infty} \Omega_{ii} \cdot \left(\frac{1 + \mu^{\lambda}}{2} \right)^{1/\lambda}. \tag{63}$$

In taking the limit the term $[(1 + \mu^{\lambda})/2]^{1/\lambda}$ goes to unity (recall that $0 < \mu \leqslant 1$) so that

$$\Delta_{ij}(\infty) = \Omega_{ii} = \max(\Omega_{ii}, \Omega_{jj}), \tag{64}$$

which is what was to be proved.

To determine the denominator of P_{ij}^{*} a procedure similar to that used for P_{ij} is employed, but the limit of equation (51) as $\lambda \to -\infty$ is taken. Hence,

$$\Delta_{ij}(-\infty) = \lim_{\lambda\to-\infty} \Delta_{ij}(\lambda) \tag{65}$$

so that

$$\Delta_{ij}(-\infty) = \lim_{\lambda\to-\infty} \left(\frac{\Omega_{ii}^{\lambda} + \Omega_{jj}^{\lambda}}{2} \right)^{1/\lambda}. \tag{66}$$

As before, substituting equation (11) into the above equation yields

$$\Delta_{ij}(-\infty) = \lim_{\lambda\to-\infty} \left(\frac{\mu^{-\lambda} \cdot \Omega_{jj}^{\lambda} + \Omega_{jj}^{\lambda}}{2} \right)^{1/\lambda}. \tag{67}$$

Rearranging terms gives

$$\Delta_{ij}(-\infty) = \lim_{\lambda\to-\infty} \Omega_{jj} \cdot \left(\frac{\mu^{-\lambda} + 1}{2} \right)^{1/\lambda}. \tag{68}$$

Upon taking the limit, we obtain the desired result

$$\Delta_{ij}(-\infty) = \Omega_{jj} = \min(\Omega_{ii}, \Omega_{jj}). \tag{69}$$

To summarize, equations (51) and (52) have been used to define a *generalized similarity index*, $S_{ij}(\lambda)$, which is characterized by the parameter λ, where $-\infty \leqslant \lambda \leqslant \infty$. It was shown that the specific similarity indices H_{ij} and H_{ij}^{*} were derivable from the generalized index by choosing λ as $+1$ and -1, respectively. Furthermore, the indices C_{ij}, P_{ij} and P_{ij}^{*} were shown to have special significance in that they correspond to limiting cases for $S_{ij}(\lambda)$, as λ approaches 0, ∞, and $-\infty$, respectively. These relationships lead naturally to an alternate notation for $\lambda \neq 0$, in which *conjugate pairs* of generalized indices, $\{S_{ij}(\omega), S_{ij}(-\omega)\}$, where $\omega > 0$, are defined. Thus, $\{H_{ij}, H_{ij}^{*}\}$ and $\{P_{ij}, P_{ij}^{*}\}$ are conjugate pairs of indices with $\omega = 1$ and $\omega = \infty$, respectively, while C_{ij} may be considered to be a *self-conjugate* index, $C_{ij} = C_{ij}^{*}$, since $\omega = 0$ in this case.

3.2. Relationships among generalized similarity indices

In section 2.5, expressions for the geometric, arithmetic, and harmonic means of P_{ij} and P^*_{ij} were derived in terms of other specific similarity indices. In this section, the possibility of finding similar relationships for $\{S_{ij}(\omega), S_{ij}(-\omega)\}$ is examined. In the case of the geometric mean, consider the product

$$S_{ij}(\omega) \cdot S_{ij}(-\omega) = \frac{\Omega_{ij}}{((\Omega_{ii}^{\omega} + \Omega_{jj}^{\omega})/2)^{1/\omega}} \cdot \frac{\Omega_{ij}}{((\Omega_{ii}^{-\omega} + \Omega_{jj}^{-\omega})/2)^{-1/\omega}}, \tag{70}$$

which by direct manipulation becomes

$$S_{ij}(\omega) \cdot S_{ij}(-\omega) = \frac{\Omega_{ij}^2}{((\Omega_{ii}^{\omega} + \Omega_{jj}^{\omega})/(\Omega_{ii}^{-\omega} + \Omega_{jj}^{-\omega}))^{1/\omega}}, \tag{71}$$

and reduces to

$$S_{ij}(\omega) \cdot S_{ij}(-\omega) = \frac{\Omega_{ij}^2}{\Omega_{ii} \cdot \Omega_{jj}} = C_{ij}^2. \tag{72}$$

Taking the positive square root of both sides of equation (72) gives a general relation for the geometric mean

$$\left[S_{ij}(\omega) \cdot S_{ij}(-\omega)\right]^{1/2} = |C_{ij}|, \tag{73}$$

that is independent of the parameter μ and is valid for all values of Ω_{ij}. This demonstrates the significance of the Carbó formula as a mean value in the family of generalized similarity indices.

For the arithmetic and harmonic means of $S_{ij}(\omega)$ and $S_{ij}(-\omega)$, no simple expression may be obtained that is independent of the parameter μ. Previously, it was shown that the arithmetic mean of P_{ij} and P^*_{ij} was equal to H^*_{ij}. To derive an analogous expression for $S_{ij}(\omega)$ and $S_{ij}(-\omega)$, we begin by substituting equation (11) into the expression for $S_{ij}(\omega)$ to give

$$S_{ij}(\omega) = \frac{\Omega_{ij}}{\Omega_{ii}} \cdot \left(\frac{2}{1+\mu^{\omega}}\right)^{1/\omega}, \tag{74}$$

or,

$$S_{ij}(\omega) = P_{ij} \cdot \left(\frac{2}{1+\mu^{\omega}}\right)^{1/\omega}. \tag{75}$$

A similar substitution of equation (11) into the formula for $S_{ij}(-\omega)$ yields

$$S_{ij}(-\omega) = \frac{\Omega_{ij}}{\Omega_{ii}} \cdot \left(\frac{1+\mu^{\omega}}{2\mu^{\omega}}\right)^{1/\omega}, \tag{76}$$

or

$$S_{ij}(-\omega) = P^*_{ij} \cdot \left(\frac{1+\mu^{\omega}}{2}\right)^{1/\omega}. \tag{77}$$

Solving equations (75) and (77) for P_{ij} and P^*_{ij}, respectively, and substituting into equation (37) gives

$$\frac{1}{2}\left[\left(\frac{1+\mu^{\omega}}{2}\right)^{1/\omega}\cdot S_{ij}(\omega)+\left(\frac{2}{1+\mu^{\omega}}\right)^{1/\omega}\cdot S_{ij}(-\omega)\right]=H^*_{ij}. \tag{78}$$

Thus, for the general case, the simple arithmetic mean has been replaced by the mean of weighted indices. Similarly for the harmonic mean, by utilizing equations (75) and (77) along with equation (40), the following weighted mean expression is obtained.

$$\frac{1}{\frac{1}{2}\left[\frac{1}{((1+\mu^{\omega})/2)^{1/\omega}\cdot S_{ij}(\omega)}+\frac{1}{(2/(1+\mu^{\omega}))^{1/\omega}\cdot S_{ij}(-\omega)}\right]}=H_{ij}. \tag{79}$$

3.3. Bounds and ordering of generalized similarity indices

The upper and lower bounds as well as the ordering of $S_{ij}(\omega)$ for various values of ω are now investigated. From equations (51) and (52), it is clear that $S_{ij}(\omega)=0$ if $\Omega_{ij}=0$, and that $S_{ij}(\omega)=1$ when $\Omega_{ij}=\Omega_{ii}=\Omega_{jj}>0$. In addition, for $\Omega_{ij}>0$, it is easy to show that $S_{ij}(\omega_2)<S_{ij}(\omega_1)$ if $\omega_2>\omega_1$, which holds provided

$$\left(\frac{\Omega_{ii}^{\omega_1}+\Omega_{jj}^{\omega_1}}{2}\right)^{1/\omega_1}\leqslant\left(\frac{\Omega_{ii}^{\omega_2}+\Omega_{jj}^{\omega_2}}{2}\right)^{1/\omega_2}. \tag{80}$$

That this is valid is seen by substituting equation (11) into equation (80), giving

$$\left(\frac{1+\mu^{\omega_1}}{2}\right)^{1/\omega_1}\leqslant\left(\frac{1+\mu^{\omega_2}}{2}\right)^{1/\omega_2}, \tag{81}$$

which in turn may be verified by inspection. Considering these results along with the upper bound for the Carbó index, equation (22), and recalling that $C_{ij}=\lim_{\omega\to 0}S_{ij}(\omega)$, establishes that

$$0\leqslant S_{ij}(\omega_2)\leqslant S_{ij}(\omega_1)\leqslant 1. \tag{82}$$

For the case in which $\Omega_{ij}<0$, a similar argument leads to the expression

$$-1\leqslant S_{ij}(\omega_1)\leqslant S_{ij}(\omega_2)\leqslant 0. \tag{83}$$

The latter two equations are consistent with the bounds obtained earlier for P_{ij} and H_{ij} in equations (24) and (46).

For $S_{ij}(-\omega)$,

$$S_{ij}(-\omega)=\frac{\Omega_{ij}}{((\Omega_{ii}^{-\omega}+\Omega_{jj}^{-\omega})/2)^{-1/\omega}}, \tag{84}$$

or

$$S_{ij}(-\omega)=\frac{\Omega_{ij}}{(2\Omega_{ii}^{\omega}\cdot\Omega_{jj}^{\omega}/(\Omega_{ii}^{\omega}+\Omega_{jj}^{\omega}))^{1/\omega}}, \tag{85}$$

which, using equation (11), may be converted to

$$S_{ij}(-\omega) = \frac{\Omega_{ij}}{\Omega_{ii} \cdot (2\mu^{\omega}/(1+\mu^{\omega}))^{1/\omega}}. \tag{86}$$

Clearly, $S_{ij}(-\omega) = 0$ if $\Omega_{ij} = 0$ and $\lim_{\mu\to 0} S_{ij}(-\omega) \to +\infty$ or $-\infty$, as $\Omega_{ij} > 0$ or $\Omega_{ij} < 0$, respectively. Moreover, for $\omega_2 > \omega_1 > 0$, the following inequality holds for $0 < \mu \leqslant 1$:

$$\left(\frac{2\mu^{\omega_2}}{1+\mu^{\omega_2}}\right)^{1/\omega_2} \leqslant \left(\frac{2\mu^{\omega_1}}{1+\mu^{\omega_1}}\right)^{1/\omega_1}. \tag{87}$$

Thus, for $\Omega_{ij} > 0$,

$$0 \leqslant S_{ij}(-\omega_1) \leqslant S_{ij}(-\omega_2) \leqslant \infty, \tag{88}$$

and, for $\Omega_{ij} < 0$,

$$-\infty \leqslant S_{ij}(-\omega_2) \leqslant S_{ij}(-\omega_1) \leqslant 0. \tag{89}$$

4. Summary and conclusions

A general formalism has been presented that incorporates the entirety of all field-based similarity indices of the form $S_{ij} = \Omega_{ij}/\Delta_{ij}$. The numerator represents the inner-product (i.e., overlap) of the field functions, F_i and F_j, for the ith and jth molecules, respectively, and the denominator is given by a suitable form of mean of the self-similarities Ω_{ii} and Ω_{jj}. More specifically, the denominators of the well-known Carbó (C_{ij}), Hodgkin and Richards (H_{ij}), and Petke (P_{ij}) similarity indices are obtained from the respective geometric mean, arithmetic mean, and maximum of the self-similarities. Two new indices, H^*_{ij} and P^*_{ij}, whose denominators are the harmonic mean and minimum of the self-similarities, respectively, were also defined. For a fixed value of Ω_{ij}, it was found that the set of similarity indices $\{P_{ij}, H_{ij}, C_{ij}, H^*_{ij}, P^*_{ij}\}$ is fundamentally ordered according to the inequality $P_{ij} \leqslant H_{ij} \leqslant C_{ij} \leqslant H^*_{ij} \leqslant P^*_{ij}$ when $\Omega_{ij} \geqslant 0$, while the reverse ordering holds when $\Omega_{ij} \leqslant 0$. Moreover, by applying the Schwartz inequality, rigorous bounds were established for all of the indices. These bounds, which hold for any value of Ω_{ij}, are given by $-1 \leqslant P_{ij}, H_{ij}, C_{ij} \leqslant 1$ and $-\infty \leqslant H^*_{ij}, P^*_{ij} \leqslant \infty$.

Using the general relation $\Omega_{jj} = \mu \cdot \Omega_{ii}$, where $0 < \mu \leqslant 1$, a complete set of equations was derived that give relationships between all pairs of indices as a function of μ. In addition, it was shown that H_{ij}, $|C_{ij}|$, and H^*_{ij} are equal to the respective harmonic, geometric, and arithmetic means of P_{ij} and P^*_{ij}.

Finally, it was demonstrated that the entire set of field-based similarity indices can be generated from a one-parameter family of functions, called a generalized mean, through proper choice of the parameter value and suitable limiting procedures. Conjugate pairs of generalized similarity indices, $\{S_{ij}(\omega), S_{ij}(-\omega)\}$, were defined in terms of a positive parameter ω, and rigorous bounds, ordering, and inter-index relationships were derived. Furthermore, it was shown that $\{H_{ij}, H^*_{ij}\}$ and $\{P_{ij}, P^*_{ij}\}$ are conjugate pairs of

indices with ω equal to 1 and ∞, respectively, while C_{ij} is a self-conjugate index, i.e. $C_{ij} = C_{ij}^*$.

The generalization of field-based similarity indices, coupled with the relationships among indices that have been developed in the present work, place the basic theory of these indices on a more unified and mathematically rigorous footing. This establishes a foundation for a better understanding of the qualitative and quantitative aspects of field-based molecular similarity, and provides a means for a numerical characterization of the relative performance of various field-based similarity indices. Such knowledge may be useful in interpreting the results of applications such as QSAR studies in which numerical values of similarity indices are used as data.

References

[1] P. Willett, *Similarity and Clustering in Chemical Information Systems* (Wiley, New York, 1987).

[2] M.A. Johnson and G.M. Maggiora (eds.), *Concepts and Applications of Molecular Similarity* (Wiley, New York, 1990).

[3] P.M. Dean (ed.), *Molecular Similarity in Drug Design* (Chapman-Hall, London, 1995).

[4] A.C. Good and W.G. Richards, Explicit calculation of 3D molecular similarity, Perspect. Drug Discov. Design 9–11 (1998) 321–338.

[5] V.J. Gillet, D.J. Wild, P. Willett and J. Bradshaw, Similarity and dissimilarity methods for processing chemical structure databases, Comp. J. 41 (1998) 547–558.

[6] P.E. Bowen-Jenkins, D.L. Cooper and W.G. Richards, Ab initio computation of molecular similarity, J. Phys. Chem. 89 (1985) 2195–2197.

[7] E.E. Hodgkin and W.G. Richards, A semi-empirical method for calculating molecular similarity, J. Chem. Soc., Chem. Commun. (1986) 1342–1344.

[8] D.L. Cooper and N.L. Allan, A novel approach to molecular similarity, J. Comput.-Aided Mol. Design 3 (1989) 253–259.

[9] A.Y. Meyer and W.G. Richards, Similarity of molecular shape, J. Comput.-Aided Mol. Design 5 (1991) 427–439.

[10] R.B. Hermann and D.K. Herron, OVID and SUPER: Two overlap programs for drug design, J. Comput.-Aided Mol. Design 5 (1991) 511–524.

[11] J. Cioslowski and E.D. Fleischmann, Assessing molecular similarity from results of ab initio electronic structure calculations, J. Am. Chem. Soc. 113 (1991) 64–67.

[12] C. Amovilli and R. McWeeny, Shape and similarity: Two aspects of molecular recognition, J. Mol. Struct. (Theochem) 227 (1991) 1–9.

[13] A.M. Richard, Quantitative comparison of molecular electrostatic potentials for structure-activity studies, J. Comput. Chem. 12 (1991) 959–969.

[14] A.C. Good, The calculation of molecular similarity: Alternative formulas, data manipulation and graphical display, J. Mol. Graph. 10 (1992) 144–151.

[15] C.A. Reynolds, C. Burt and W.G. Richards, A linear molecular similarity index, Quant. Struct.-Act. Relat. 11 (1992) 34–35.

[16] C. Lee and S. Smithline, An approach to molecular similarity using density functional theory, J. Phys. Chem. 98 (1994) 1135–1138.

[17] R. Carbó, L. Leyda and M. Arnau, How similar is a molecule to another? An electron density measure of similarity between two molecular structures, Int. J. Quantum Chem. 17 (1980) 1185–1189.

[18] R. Carbó and L. Domingo, LCAO-MO similarity measures and taxonomy, Int. J. Quantum Chem. 32 (1987) 517–545.

[19] P.G. Mezey, *Shape in Chemistry: An Introduction to Molecular Shape and Topology* (VCH Publishers, New York, 1993) chapter 2.
[20] A.C. Good and W.G. Richards, Rapid evaluation of shape similarity using Gaussian functions, J. Chem. Inf. Comput. Sci. 33 (1993) 112–116.
[21] E.E. Hodgkin and W.G. Richards, Molecular similarity based on electrostatic potential and electric field, Int. J. Quantum Chem. Quantum Biol. Symp. 14 (1987) 105–110.
[22] W.G. Richards and E.E. Hodgkin, Molecular similarity, Chem. Br. 24 (1988) 1141–1144.
[23] C. Burt, W.G. Richards and P. Huxley, The application of molecular similarity calculations, J. Comput. Chem. 11 (1990) 1139–1146.
[24] C. Burt and W.G. Richards, Molecular similarity: The introduction of flexible fitting, J. Comput.-Aided Mol. Design 4 (1990) 231–238.
[25] F. Manaut, F. Sanz, J. Jose and M. Milesi, Automatic search for maximum similarity between molecular electrostatic potential distributions, J. Comput.-Aided Mol. Design 5 (1991) 371–380.
[26] F. Sanz, F. Manaut, J.A. Sanchez and E. Lozoya, Maximum electrostatic similarity between biomolecules optimizing both relative positions and conformations, J. Mol. Struct. (Theochem) 230 (1991) 437–446.
[27] F. Sanz, F. Manaut, J. Rodríguez, E. Lozoya and E. López-de-Briñas, MEPSIM: A computational package for analysis and comparison of molecular electrostatic potentials, J. Comput.-Aided Mol. Design 7 (1993) 337–347.
[28] J.D. Petke, Cumulative and discrete similarity analysis of electrostatic potentials and fields, J. Comput. Chem. 14 (1993) 928–933.
[29] D.J. Wild and P. Willett, Similarity searching in files of three-dimensional chemical structures. Alignment of molecular electrostatic potential fields with a genetic algorithm, J. Chem. Inf. Comput. Sci. 36 (1996) 159–167.
[30] A.C. Good, E.E. Hodgkin and W.G. Richards, Utilization of Gaussian functions for the rapid evaluation of molecular similarity, J. Chem. Inf. Comput. Sci. 32 (1992) 188–191.
[31] F. Croizet, M.H. Langlois, J.P. Dubost, P. Braquet, E. Audry, Ph. Dallet and J.C. Colleter, Lipophilicity force field profile: An expressive visualization of the lipophilicity molecular potential gradient, J. Mol. Graph. 8 (1990) 153–155.
[32] W. Heiden, G. Moeckel and J. Brickmann, A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces, J. Comput.-Aided Mol. Design 7 (1993) 503–514.
[33] P. Gaillard, P. Carrupt, B. Testa and A. Boudon, Molecular lipophilicity potential, a tool in 3D QSAR: Method and applications, J. Comput.-Aided Mol. Design 8 (1994) 83–96.
[34] R.G.A. Bone and H.O. Villar, Discriminating D1 and D2 agonists with a hydrophobic similarity index, J. Mol. Graph. 13 (1995) 201–208.
[35] I. Rozas, Q. Du and G.A. Arteca, Interrelation between electrostatic and lipophilicity potentials on molecular surfaces, J. Mol. Graph. 13 (1995) 98–108.
[36] G.A. Arteca and D.R. Van Allen, Three-dimensional lipophilicity characterization of molecular pores and channel-like cavities, J. Mol. Graph. 14 (1996) 235–242.
[37] Q. Du, G.A. Arteca and P.G. Mezey, Heuristic lipophilicity potential for computer-aided rational drug design, J. Comput.-Aided Mol. Design 11 (1997) 503–515.
[38] D. Robert and R. Carbó-Dorca, A formal comparison between molecular quantum similarity measures and indices, J. Chem. Inf. Comput. Sci. 38 (1998) 469–475.
[39] R. Carbó, B. Calabuig, E. Besalú and A. Martinez, Triple density res: A general connection between theoretical calculations and experimental results, Mol. Engineering 2 (1992) 43–64.
[40] E. Besalú, R. Carbó, J. Mestres and M. Solà, Foundations and recent developments in molecular quantum similarity, Topics Curr. Chem. 173 (1995) 31–62.
[41] 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.

[42] M. Petitjean, Geometric molecular similarity from volume-based distance minimization: Application to saxitoxin and tetrodotoxin, J. Comput. Chem. 16 (1995) 80–90.
[43] M. Petitjean, Three-dimensional pattern recognition from molecular distance minimization, J. Chem. Inf. Comput. Sci. 36 (1996) 1038–1049.
[44] A. Tversky, Features of similarity, Psychol. Rev. 84 (1977) 327–352.
[45] M.I. Skvortsova, I.I. Baskin, I.V. Stankevich, V.A. Palyulin and N.S. Zefirov, Molecular similarity. 1. Analytical description of the set of graph similarity measures, J. Chem. Inf. Comput. Sci. 38 (1998) 785–790.
[46] S.K. Kearsley and G.M. Smith, An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap, Tetrahedron Comput. Methodol. 3 (1990) 615–633.
[47] J. Mestres, D.C. Rohrer and G.M. Maggiora, MIMIC: A molecular-field matching program. Exploiting applicability of molecular similarity approaches, J. Comput. Chem. 18 (1997) 934–954.
[48] J. Mestres, D.C. Rohrer and G.M. Maggiora, A molecular field-based similarity approach to pharmacophoric pattern recognition, J. Mol. Graph. Model. 15 (1997) 114–121.
[49] J. Mestres, D.C. Rohrer and G.M. Maggiora, A molecular field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors, J. Comput.-Aided Mol. Design 13 (1999) 79–93.
[50] J. Mestres, D.C. Rohrer and G.M. Maggiora, A molecular field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors. 2. The relationship between alignment solutions obtained from conformationally rigid and flexible matching, J. Comput.-Aided Mol. Design, in press.
[51] G. Klebe, T. Mietzner and F. Weber, Different approaches toward an automatic structural alignment of drug molecules: Applications to sterol mimics, thrombin and thermolysin inhibitors, J. Comput.-Aided Mol. Design 8 (1994) 751–778.
[52] G. Klebe, T. Mietzner and F. Weber, Methodological developments and strategies for a fast flexible superposition of drug-size molecules, J. Comput.-Aided Mol. Design 13 (1999) 35–49.
[53] M.F. Parretti, R.T. Kroemer, J.H. Rothman and W.G. Richards, Alignment of molecules by the Monte Carlo optimization of molecular similarity indices, J. Comput. Chem. 18 (1997) 1344–1353.
[54] C. Lemmen, C. Hiller and T. Lengauer, RigFit: A new approach to superimposing ligand molecules, J. Comput.-Aided Mol. Design 12 (1998) 491–502.
[55] A.C. Good, In *Molecular Similarity in Drug Design*, ed. P.M. Dean (Chapman-Hall, London, 1995) p. 24.
[56] A.C. Good, S.-S. So and W.G. Richards, Structure-activity relationships from molecular similarity matrices, J. Med. Chem. 36 (1993) 433–438.
[57] A.C. Good, S.J. Peterson and W.G. Richards, QSARs from similarity matrices. Technique validation and application in the comparison of different similarity evaluation methods, J. Med. Chem. 36 (1993) 2929–2937.
[58] R. Benigni, M. Cotta-Ramusino, F. Georgi and G. Gallo, Molecular similarity matrices and quantitative structure-activity relationships: A case study with methodological implications, J. Med. Chem. 38 (1995) 629–635.
[59] D.C. Horwell, W. Howson, M. Higginbottom, D. Naylor, G.S. Ratcliffe and S. Williams, Quantitative structure-activity relationships (QSARs) of N-terminus fragments of NK1 tachykinin antagonists: A comparison of classical QSARs and three-dimensional QSARs from similarity matrices, J. Med. Chem. 38 (1995) 4454–4462.
[60] C.A. Montanari, M.S. Tute, A.E. Beezer and J.C. Mitchell, Determination of receptor-bound drug conformations by QSAR using flexible fitting to derive a molecular similarity index, J. Comput.-Aided Mol. Design 10 (1996) 67–73.
[61] X. Fradera, L. Amat, E. Besalú and R. Carbó-Dorca, Application of molecular quantum similarity to QSAR, Quant. Struct.-Act. Relat. 16 (1997) 25–32.

[62] M. Lobato, L. Amat, E. Besalú and R. Carbó-Dorca, Structure-activity relationships of a steroid family using quantum similarity measures and topological quantum similarity indices, Quant. Struct.-Act. Relat. 16 (1997) 465–472.
[63] S.S. So and M. Karplus, Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations, J. Med. Chem. 40 (1997) 4347–4359.
[64] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic – Theory and Applications* (Prentice-Hall, Upper Saddle River, NJ, 1995) chapter 3.
[65] P. Dennery and A. Krzywicki, *Mathematics for Physicists* (Harper & Row, New York, 1967) p. 182.