

# Molecular alignment as a penalized permutation Procrustes problem

Farnaz Heidar Zadeh · Paul W. Ayers

Received: 3 September 2012 / Accepted: 19 November 2012 / Published online: 18 December 2012  
© Springer Science+Business Media New York 2012

**Abstract** Molecular alignment is viewed as a permutation Procrustes problem, where the goal is to find the best assignment of points (or functional groups) in one molecule to the points in another molecule. A penalty function ensures that the optimal alignment respects the underlying connectivity between atoms/points. This method helps reveal why molecular alignment suffers from the curse of dimension.

**Keywords** Molecular similarity · 3D-QSAR · Quantum QSAR · Molecular alignment · Permutation Procrustes problem ·  $k$ -nearest neighbor alignment

## 1 Motivation

Three-dimensional quantitative structure activity relationships (3D-QSAR) are built by computing the similarity between different molecules, then using the precept that similar molecules have similar properties to make predictions. [1–8] In assessing the similarity of two molecules, however, one must first choose the appropriate relative position, orientation, and (sometimes) conformation of the molecules: a molecule will not even be similar to itself if it is misaligned. This leads to the problem of finding the optimal alignment between molecules. There are many approaches in the literature (see Ref. [9]) and references cited therein). We are primarily interested in the quantum QSAR alignment problem, where the molecules are aligned based on quantum mechanical properties. Cf. Refs. [10–16]. However, our approach is valid for any sort of molecular alignment and, indeed, the general problem of structural alignment.

---

F. H. Zadeh · P. W. Ayers (✉)  
Department of Chemistry and Chemical Biology, McMaster University, Hamilton, ON, Canada  
e-mail: ayers@mcmaster.ca

One problem in molecular alignment is that there is no way to quantify what indicates a “good alignment” method. Consider two ambiphilic molecules. (A molecule is ambiphilic if it has both electrophilic and nucleophilic sites.) The activity of these molecules with respect to an electrophile is determined by their nucleophilic sites, so assessing the similarity of the molecules’ nucleophilic activity requires aligning their nucleophilic regions. Similarly, for assessing the activity of the molecules to a nucleophilic reagent, their electrophilic regions should be aligned. For another property (e.g., their solubility), an entirely different alignment protocol might be preferable. It is important, then, to have a flexible alignment method that can align molecules based on many different criteria. In this paper we present a flexible alignment method of this type.

We assume that each molecule,  $M$ , is described by a set of  $N_{\text{pts}}^{(M)}$  points,  $\{\mathbf{q}_i^{(M)}\}_{i=1}^{N_{\text{pts}}^{(M)}}$ .

At each point one has a vector of  $N_{\text{prop}}$  properties,  $\left\{ \left[ p_1^{(M)}, p_2^{(M)}, \dots, p_{N_{\text{prop}}}^{(M)} \right] \right\}_{i=1}^{N_{\text{pts}}^{(M)}}$ . The similarity of the molecules is written as

$$Z^{(M,M')} = \sum_{i=1}^{\min(N_{\text{pts}}^{(M)}, N_{\text{pts}}^{(M')})} w_{ij} \left( \mathbf{p}_i^{(M)} \right)^T \mathbf{L} \mathbf{L}^T \mathbf{p}_j^{(M')} \quad (1)$$

where the point-weights  $\{w_{ij}\}_{i,j=1}^{\min(N_{\text{pts}}^{(M)}, N_{\text{pts}}^{(M')})}$  and the  $N_{\text{prop}} \times N_{\text{prop}}$  lower diagonal matrix  $\mathbf{L}$  are adjusted to construct problem-specific alignment methods. (The point-weights could depend on the properties, but we shall assume that they do not.) In most cases the point-weights will either be  $w_i = 1/N_{\text{pts}}$  or diagonal (e.g. integration) weights  $w_{ij} = w_i \delta_{ij}$ ; in most cases the matrix  $\mathbf{L}$  will be diagonal, but allowing off-diagonal structure can be helpful when specific linear combinations of properties are good predictors of the molecular activity. By suitably choosing  $\mathbf{L}$ , one can tune the relative importance of matching different properties (or linear combinations of properties) of the molecules. By choosing  $w_{ij}$ , one controls the relative importance of different points/atoms in the molecules. Notice that the number of points in the two molecules can be different; that is, not every point in one molecule has to correspond to a point in the other molecule. The alignment problem amounts to finding the optimal correspondence of points.

This general form of molecular similarity includes many of the prevalent cases in the literature. The venerable Carbó indicator arises when the weights are integration weights and the property is the electron density. [8, 17, 18] Similarity measures based on one (or several) chemical reactivity indicators, e.g., those based on conceptual DFT, [19–23] can be handled similarly. [14, 15, 24–26] A condensed representation of the molecule can be obtained by computing atomic contributions to molecular properties (e.g., the energy or the total molecular charge) and chemical reactivity indicators. [27–34] In this case, one measures molecular similarity by building a mapping between atoms in one molecule and atoms in a different molecule. However, the procedure we are using applies to the alignment of any set of points, not just atomic positions.

In this paper, we propose an algorithm for maximizing the similarity of two molecules, as measured by Eq. (1). To do this, we view finding the optimal alignment of the molecular point sets of molecules  $M$  and  $M'$ ,  $\{\mathbf{p}_i^{(M)}\}_{i=1}^{N_{pts}^{(M)}}$  and  $\{\mathbf{p}_i^{(M')}\}_{i=1}^{N_{pts}^{(M)'}}$ , as a generalized permutation Procrustes problem. Molecular and macromolecular alignment has been viewed as a Procrustes problem before, [35–37] but to our knowledge, this is the first time the link to the permutation Procrustes algorithm has been made and the first time an Procrustes-based algorithm has been proposed for the very general similarity measure in Eq. (1). (In fact, the molecular similarity measure we treat is even more general than this one, as it includes information about the similarity of  $k$ th nearest neighbors. [38–40])

## 2 Penalized permutation Procrustes

Denote the properties of the  $k$  nearest neighbors of point  $i$  as  $\mathbf{p}_{i,0}^{(M)}, \mathbf{p}_{i,1}^{(M)}, \dots, \mathbf{p}_{i,k}^{(M)}$ , where  $\mathbf{p}_{i,0}^{(M)}$  are the properties of the point itself. Construct an  $N_{pts}^{(M)} \times (k + 1)$  matrix containing the (transformed) property vectors,

$$\mathbf{F}^{(M)} \equiv \begin{bmatrix} \mathbf{L}_0^T \mathbf{p}_{1,0}^{(M)} & \mathbf{L}_1^T \mathbf{p}_{1,1}^{(M)} & \dots & \mathbf{L}_k^T \mathbf{p}_{1,k}^{(M)} \\ \mathbf{L}_0^T \mathbf{p}_{2,0}^{(M)} & \mathbf{L}_1^T \mathbf{p}_{2,1}^{(M)} & \dots & \mathbf{L}_k^T \mathbf{p}_{2,k}^{(M)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}_0^T \mathbf{p}_{N_{pts},0}^{(M)} & \mathbf{L}_1^T \mathbf{p}_{N_{pts},1}^{(M)} & \dots & \mathbf{L}_k^T \mathbf{p}_{N_{pts},k}^{(M)} \end{bmatrix} \quad (2)$$

In general, the  $\mathbf{L}_k^T$  will decrease in magnitude as  $k$  increases, because it is less important to ensure that two points/atoms  $k$ th nearest neighbors have similar properties than it is to ensure that the points/atoms themselves are similar. In the case where  $k = 0$ , the below method will optimize the similarity measure in Eq. (1). Alternatively, if we consider the properties of the  $k$ -nearest-neighbor atoms to be properties of the atom itself, the similarity expression in Eq. (1) is recovered by considering,

$$\mathbf{p}_i^{(M)} = \left[ \mathbf{L}_0^T \mathbf{p}_{i,0}^{(M)}, \mathbf{L}_1^T \mathbf{p}_{i,1}^{(M)}, \dots, \mathbf{L}_k^T \mathbf{p}_{i,k}^{(M)} \right]^T \quad (3)$$

Our goal is to find a permutation of the rows of  $\mathbf{F}^{(M)}$  that maximizes its similarity (equivalently, minimizes its distance) to  $\mathbf{F}^{(M')}$ . Without loss of generality, we assume that  $N_{pts}^{(M)} < N_{pts}^{(M')}$ . Let  $\mathbf{P}$  be a  $N_{pts}^{(M')} \times N_{pts}^{(M)}$  generalized permutation matrix. That is, all the elements of  $\mathbf{P}$  are either zero or one, all the row-sums are either zero (in which case a point/atom in  $M'$  is unassigned) or one, and all the column-sums are either zero (in which case an atom in  $M$  is unassigned) or one. Each nonzero entry,  $p_{kl} = 1$ , in  $\mathbf{P}$  indicates that the  $k$ th point/atom in  $M'$  is assigned to the  $l$ th point/atom in  $M$ . Our problem is then,

$$\underbrace{\min}_{\left\{ p_{ij} \in \{0,1\} \left| \begin{array}{l} \{0, 1\} = \sum_i^{N^{(M')}} p_{ij} \\ \{0, 1\} \geq \sum_{j=1}^{N^{(M)}} p_{ij} \end{array} \right. \right\}} \text{Tr} \left[ \left( \mathbf{F}^{(M')} - \mathbf{P}\mathbf{F}^{(M)} \right)^T \mathbf{W} \left( \mathbf{F}^{(M')} - \mathbf{P}\mathbf{F}^{(M)} \right) \right] \quad (4)$$

$\mathbf{W}$  is the matrix containing the  $w_{ij}$ 's from Eq. (1)

If we require that every point/atom in  $M$  be assigned, then (4) simplifies to

$$\underbrace{\max}_{\left\{ p_{ij} \in \{0,1\} \left| \begin{array}{l} \{0, 1\} = \sum_i^{N^{(M')}} p_{ij} \\ 1 = \sum_{j=1}^{N^{(M)}} p_{ij} \end{array} \right. \right\}} \text{Tr} \left[ \left( \mathbf{F}^{(M')} \right)^T \mathbf{W} \left( \mathbf{P}\mathbf{F}^{(M)} \right) \right] \quad (5)$$

because  $(\mathbf{F}^{(M)})^T \mathbf{P}^T \mathbf{P} \mathbf{F}^{(M)} = (\mathbf{F}^{(M)})^T \mathbf{F}^{(M)}$  does not depend on the choice of  $\mathbf{P}$ . The assumption that all points in  $M$  are assigned can be relaxed later.

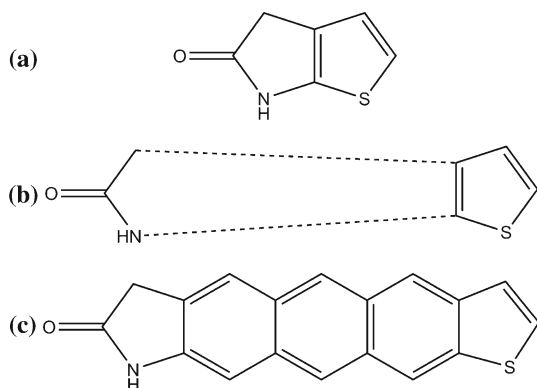
This is a generalized (because  $\mathbf{F}^{(M')}$  and  $\mathbf{F}^{(M)}$  are not necessarily the same size) permutation Procrustes problem. [41, 42] It can be solved by noticing that (5) is linear in  $\mathbf{P}$ . Therefore, the linear programming problem,

$$\underbrace{\max}_{\left\{ 0 \leq p_{ij} \leq 1 \left| \begin{array}{l} 1 \geq \sum_i^{N^{(M')}} p_{ij} \\ 1 = \sum_{j=1}^{N^{(M)}} p_{ij} \end{array} \right. \right\}} \left( \mathbf{F}^{(M')} \right)^T \mathbf{W} \left( \mathbf{P}\mathbf{F}^{(M)} \right) \quad (6)$$

has a solution on the boundary of the allowed region, where  $p_{ij} \in \{0, 1\}$ . Problem (6) is an assignment problem; it can be efficiently solved using methods like the Hungarian algorithm. [43] The optimum solution in Eq. (6) is a solution (albeit a very questionable one) to the molecular alignment problem.

Why questionable? Recall that permuting the rows of Eq. (2) corresponds to rearranging the points/atoms in the molecule. The permutation Procrustes algorithm in Eq. (6) does this without respecting the connectivity of the atoms (the proximity of the points). For a pair of molecules like the ones shown in Fig. 1, the alignment will “rip apart” the smaller molecule.

To avoid this problem, we need to penalize the Procrustes analysis so that molecular conformations that are high in energy are eliminated. This can be done in two ways. The simplest way is to define the distance matrix  $\mathbf{D}^{(M)}$ , with elements



**Fig. 1** In a non-penalized permutation Procrustes method, atoms in molecule (a) will be aligned with the most similar atoms in molecule (c) according to cartoon (b). This favorable matching of atom types does not respect the connectivity of the atoms in molecule (a). With a penalized permutation Procrustes method, once the magnitude of the penalty is large enough, molecule (a) will eventually decide to align either with the *right-hand-side* or the *left-hand-side* of molecule (c)

$$d_{ij}^{(M)} = \left| \mathbf{q}_i^{(M)} - \mathbf{q}_j^{(M)} \right|. \quad (7)$$

We then minimize the expression in Eq. (4), subject to a penalty factor that grows quadratically when the distance matrices of  $M$  and  $M'$  are dissimilar,

$$\min_{\left\{ \begin{array}{l} p_{ij} \in \{0,1\} \\ \{0,1\} = \sum_j^{N_{\text{pts}}(M')} p_{ij} \\ 1 = \sum_{j=1}^{N_{\text{pts}}(M)} p_{ij} \end{array} \right\}} \left( \mathbf{F}^{(M')} - \mathbf{P}\mathbf{F}^{(M)} \right)^T \mathbf{W} \left( \mathbf{F}^{(M')} - \mathbf{P}\mathbf{F}^{(M)} \right) + \Omega \left| \mathbf{D}^{(M')} - \mathbf{P}\mathbf{D}^{(M)}\mathbf{P}^T \right|^2 \quad (8)$$

This ensures that points/atoms that are close to each other in molecule  $M$  are not assigned to atoms that are far apart in molecule  $M'$ . The number  $\Omega > 0$  controls the importance of the distance constraint; should be increased until a chemically reasonable molecular alignment is achieved. In Fig. 1, as  $\Omega$  increases, the method will be eventually be forced to choose between aligning either the right-hand-ends or the left-hand ends of the molecules. The particular alignment chosen will depend on the choice of  $\mathbf{L}_k$  and  $\mathbf{W}$ , and will therefore be different depending on the target property.

A better solution to problem (8) could be obtained if we did not force every atom in  $M$  to be assigned to an atom in  $M'$ . However, without this assumption problem (8) is quartic (not quadratic) in  $\mathbf{P}$  (because  $\mathbf{P}\mathbf{D}^{(M)}\mathbf{P}^T\mathbf{P}\mathbf{D}^{(M)}\mathbf{P}^T = \mathbf{P}(\mathbf{D}^{(M)})^2\mathbf{P}^T$  only if all the atoms/points in  $M$  are assigned).

It is reasonable, but not essential, to symmetrize this problem by writing

$$\left\{ \begin{array}{l} \min \\ p_{ij} \in \{0, 1\} \\ q_{ij} \in \{0, 1\} \end{array} \left| \begin{array}{l} \{0, 1\} = \sum_{i=1}^{N_{\text{pts}}^{(M')}} p_{ij} \\ \{0, 1\} = \sum_{j=1}^{N_{\text{pts}}^{(M)}} q_{ij} \\ 1 = \sum_{j=1}^{N_{\text{pts}}^{(M)}} p_{ij} \\ 1 = \sum_{i=1}^{N_{\text{pts}}^{(M')}} q_{ij} \end{array} \right. \right. \left( \begin{array}{l} (\mathbf{F}^{(M')} - \mathbf{PF}^{(M)})^T \mathbf{W} (\mathbf{F}^{(M')} - \mathbf{PF}^{(M)}) \\ + (\mathbf{F}^{(M')} \mathbf{Q}^T - \mathbf{F}^{(M)})^T \mathbf{W} (\mathbf{F}^{(M')} \mathbf{Q}^T - \mathbf{F}^{(M)}) \\ + \Omega \left( |\mathbf{D}^{(M')} - \mathbf{PD}^{(M)} \mathbf{P}^T|^2 + |\mathbf{QD}^{(M')} \mathbf{Q}^T - \mathbf{D}^{(M)}|^2 \right) \end{array} \right) \quad (9)$$

For simplicity, we will consider only problem (8), but all our comments will apply to problem (9) also. We call (8) and (9) penalized permutation Procrustes problems.

Instead of problem (8), we can write the problem as the maximization of the expectation value of a positive-definite quadratic form using 0–1 vectors (PosDef-0-1-Max), [44]

$$\left\{ \begin{array}{l} \max \\ p_{ij} \in \{0, 1\} \end{array} \left| \begin{array}{l} \{0, 1\} = \sum_{i=1}^{N_{\text{pts}}^{(M')}} p_{ij} \\ 1 = \sum_{j=1}^{N_{\text{pts}}^{(M)}} p_{ij} \end{array} \right. \right. (\mathbf{F}^{(M')})^T \mathbf{W} (\mathbf{PF}^{(M)}) - \Omega |\mathbf{D}^{(M')} - \mathbf{PD}^{(M)} \mathbf{P}^T|^2 \quad (10)$$

(This is equivalent to problem (8) only because we assumed that all the points/atoms in  $M$  are assigned.) This is a standard 0–1 integer programming problem that is, unfortunately, NP-hard. Such problems are often solved using branch-and-bound approaches. The following algorithm is of that type:

Step 1. Initialization. Solve the problem for  $\Omega = 0$ . This is a generalized permutation Procrustes problem, so the Hungarian algorithm (or a related technique) can be applied. It gives an initial permutation matrix.

Step 2. Increase  $\Omega$  unless the present alignment is judged to be satisfactory (in the sense that the penalty term in problem (10) has a sufficiently small value).

Step 3. For  $\Omega > 0$ , minimize over all relaxed pseudo-permutation matrices, i.e., all  $\tilde{\mathbf{P}}$  with

$$0 \leq \tilde{p}_{ij} \leq 1 \quad 1 \geq \sum_{i=1}^{N_{\text{pts}}^{(M')}} \tilde{p}_{ij} \quad 1 \geq \sum_{j=1}^{N_{\text{pts}}^{(M)}} \tilde{p}_{ij} \quad (11)$$

This is a conventional optimization problem with inequality constraints. This gives a lower bound on the true function value because of the first constraint above.

Step 4. Look at the columns of the permutation matrix and select the one that is closest to a unit vector. We branch at this point by setting that column equal to the corresponding unit vector.

There are two cases:

- An upper bound on the cost-function for this state is obtained by forming a permutation matrix by (a) seeing whether each additional column in the matrix is closer to the zero vector or a unit vector and (b) choosing this “nearest pseudopermutation matrix”. This gives an upper-bound to the cost. Minimizing with this unit-vector constraint gives a lower-bound to the cost.
- Row  $k$  has a zero in position  $l$ . Add this constraint and then minimize over the relaxed pseudo-permutation matrices. This gives a lower bound. If the lower-bound is greater than the upper-bound, this possibility can be neglected.

Step 5. Take the lower-cost choice from step 4 and “branch” again by finding a possible unit vector. Keep doing this “branching” until one finds a pseudo-permutation matrix, with  $p_{ij} \in \{0, 1\}$ . At that stage, all of the alternatives with a lower bound greater than this solution can be discarded. The lowest-cost alternative is then examined by

Step 6. Go back to step 2. (In this method, one needs to gradually increase  $\Omega$  until “sensible” results are obtained.)

Problem (10) does not allow for conformational flexibility of molecules. An alternative approach is to assign the points/atoms in molecule  $M$  to the positions of the corresponding atoms in molecule  $M'$ . One then computes the energy of that conformation. In this case, assignments that correspond to low-energy motions of molecule  $M$  (obviously including translations and rotations, but also including torsions) will be considered favorable. This corresponds to the problem,

$$\min_{\left\{ \begin{array}{l} p_{ij} \in \{0,1\} \\ \{0, 1\} = \sum_i^{N_{\text{pts}}^{(M)}} p_{ij} \\ \{0, 1\} = \sum_{j=1}^{N_{\text{pts}}^{(M)}} p_{ij} \end{array} \right.} \left( \begin{array}{l} \left( \mathbf{F}^{(M')} - \mathbf{P}\mathbf{F}^{(M)} \right)^T \mathbf{W} \left( \mathbf{F}^{(M')} - \mathbf{P}\mathbf{F}^{(M)} \right) \\ + \exp \left( -\beta \left( E^{(M)} \left( \mathbf{q}_i^{(M)} \right) - E^{(M)} \left( \mathbf{P}^T \mathbf{q}_i^{(M')} \right) \right) \right) \end{array} \right) \quad (12)$$

The Boltzmann-factor dependence on the energy difference just one of many reasonable forms. Problem (12) is also a 0–1 integer program, but it is no longer quadratic. Since there is no longer any benefit to assuming that all atoms in  $M$  are assigned to atoms in  $M'$ , we relax that assumption. The preceding algorithm, which does not exploit the special properties of a quadratic 0–1 programming problem, is applicable to problem (12) also.

### 3 Summary

We introduce the penalized permutation Procrustes problem as an approach to the molecular alignment problem. In this approach, we maximize the similarity of a point (or atom) and its  $k$ -nearest neighbors ( $k \geq 0$ ) in molecule  $M$  to a point (or atom) in molecule  $M'$  by treating this problem as an assignment problem. The bare assignment problem is a (generalized) permutation Procrustes problem; it can be efficiently solved using linear programming techniques.

Unfortunately, the bare assignment problem does not preserve the proximity of points/atoms: points that are close together in molecule  $M$  might be far apart in molecule  $M'$ . To avoid this problem, we add a penalty function that forces points (or atoms) that are close together in  $M$  to be assigned to points that are close together in  $M'$ . (By using the energy, instead of the distance, to assess the closeness, conformational flexibility in the molecules can be included.) The resulting penalized permutation Procrustes problem is a 0–1 integer programming problem; it is unfortunately NP-hard, though we suspect that greedy algorithms may work well for it. (One reason that we include the  $k$ -nearest neighbor information in our proposed similarity measures is that we believe this will increase the likelihood that greedy algorithms will perform well. This is especially true if the distance to the nearest neighbors is included in the property vector.)

Finally, we believe that this approach helps elucidate why molecular alignment is so difficult. Even though we believe that our formulation is relatively efficient, the problem is still NP-hard. This observation explains why alignment-free 3D-QSAR methods are popular.

**Acknowledgments** FH acknowledges an Ontario Graduate Scholarship. PWA acknowledges support from NSERC.

### References

1. M. Arakawa, K. Hasegawa, K. Funatsu, The recent trend in QSAR modeling—variable selection and 3D-QSAR methods. *Curr. Comput. Aided Drug Des.* **3**, 254–262 (2007)
2. P. Bultinck, X. Girones, R. Carbó-Dorca, *Molecular quantum similarity: theory and applications* (2005)
3. E. Besalú, X. Girones, L. Amat, R. Carbó-Dorca, Molecular quantum similarity and the fundamentals of QSAR. *Acc. Chem. Res.* **35**, 289–295 (2002)
4. R. Carbó-Dorca, L. Amat, E. Besalú, X. Girones, D. Robert, Quantum mechanical origin of QSAR: theory and applications. *J. Mol. Struct. Theochem* **504**, 181–228 (2000)
5. R. Carbó-Dorca, L. Amat, E. Besalú, M. Lobato, in *Quantum similarity*, ed. by R. CarboDorca, P.G. Mezey. *Advances in Molecular Similarity* pp. 1–42 (1998)
6. R. Carbó-Dorca, E. Besalú, A general survey of molecular quantum similarity. *J. Mol. Struct. Theochem* **451**, 11–23 (1998)
7. R. Carbó, E. Besalú, L. Amat, X. Fradera, On quantum molecular similarity measures (QMSM) and indices (QMSI). *J. Math. Chem.* **19**, 47–56 (1996)
8. R. Carbó, L. Leyda, M. Arnau, How similar is a molecule to another—an electron-density measure of similarity between 2 molecular-structures. *Int. J. Quantum Chem.* **17**, 1185–1189 (1980)
9. C. Lemmen, T. Lengauer, Computational methods for the structural alignment of molecules. *J. Comput. Aided Mol. Des.* **14**, 215–232 (2000)
10. X. Girones, D. Robert, R. Carbo-Dorca, TGSA: a molecular superposition program based on topological considerations. *J. Comput. Chem.* **22**, 255–263 (2001)



11. P. Bultinck, T. Kuppens, X. Girone, R. Carbo-Dorca, Quantum similarity superposition algorithm (QSSA): a consistent scheme for molecular alignment and molecular similarity based on quantum chemistry. *J. Chem. Inf. Comput. Sci.* **43**, 1143–1150 (2003)
12. P. Bultinck, R. Carbo-Dorca, C. Van Alsenoy, Quality of approximate electron densities and internal consistency of molecular alignment algorithms in molecular quantum similarity. *J. Chem. Inf. Comput. Sci.* **43**, 1208–1217 (2003)
13. X. Girones, R. Carbo-Dorca, TGSA-flex: extending the capabilities of the topo-geometrical superposition algorithm to handle flexible molecules. *J. Comput. Chem.* **25**, 153–159 (2004)
14. S. Van Damme, W. Langenaeker, P. Bultinck, Prediction of blood-brain partitioning: a model based on ab initio calculated quantum chemical descriptors. *J. Mol. Graph. Model.* **26**, 1223–1236 (2008)
15. R. Dolezal, S. Van Damme, P. Bultinck, K. Waisser, QSAR analysis of salicylamide isosteres with the use of quantum chemical molecular descriptors. *Eur. J. Med. Chem.* **44**, 869–876 (2009)
16. F. Heidar Zadeh, P.W. Ayers (submitted)
17. R. Carbó, B. Calabuig, Molecular quantum similarity measures and n-dimensional representation of quantum objects. 1. Theoretical foundations. *Int. J. Quantum Chem.* **42**, 1681–1693 (1992)
18. R. Carbó, B. Calabuig, Molecular quantum similarity measures and n-dimensional representation of quantum objects. 2. Practical applications. *Int. J. Quantum Chem.* **42**, 1695–1709 (1992)
19. P.A. Johnson, L.J. Bartolotti, P.W. Ayers, T. Fievez, P. Geerlings, Charge density and chemical reactivity: a unified view from conceptual DFT, in *Modern Charge Density Analysis*, ed. by C. Gatti, P. Macchi (Springer, New York, 2012), pp. 715–764
20. P.W. Ayers, J.S.M. Anderson, L.J. Bartolotti, Perturbative perspectives on the chemical reaction prediction problem. *Int. J. Quantum Chem.* **101**, 520–534 (2005)
21. P. Geerlings, F. De Proft, W. Langenaeker, Conceptual density functional theory. *Chem. Rev.* **103**, 1793–1873 (2003)
22. S.B. Liu, Conceptual density functional theory and some recent developments. *Acta Physico-Chimica Sinica* **25**, 590–600 (2009)
23. J.L. Gazquez, Perspectives on the density functional theory of chemical reactivity. *J. Mexican Chem. Soc.* **52**, 3–10 (2008)
24. P. Bultinck, R. Carbó-Dorca, Molecular quantum similarity using conceptual DFT descriptors. *J. Chem. Sci.* **117**, 425–435 (2005)
25. S. Van Damme, P. Bultinck, 3D QSAR based on conceptual DFT molecular fields: antituberculosic activity. *J. Mol. Struct. Theochem* **943**, 83–89 (2010)
26. S. Van Damme, P. Bultinck, Conceptual DFT properties-based 3D QSAR: analysis of inhibitors of the nicotine metabolizing CYP2A6 enzyme. *J. Comput. Chem.* **30**, 1749–1757 (2009)
27. W.T. Yang, W.J. Mortier, The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *J. Am. Chem. Soc.* **108**, 5708–5711 (1986)
28. R.F.W. Bader, *Atoms in Molecules: A Quantum Theory* (Clarendon, Oxford, 1990)
29. R.F.W. Bader, Atoms in molecules. *Acc. Chem. Res.* **18**, 9–15 (1985)
30. L. Cohen, Local kinetic energy in quantum mechanics. *J. Chem. Phys.* **70**, 788–789 (1979)
31. P. Bultinck, S. Fias, C.V. Alsenoy, P.W. Ayers, R. Carbó-Dorca, Critical thoughts on computing atom condensed Fukui functions. *J. Chem. Phys.* **127**, 034102 (2007)
32. W. Tiznado, E. Chamorro, R. Contreras, P. Fuentealba, Comparison among four different ways to condense the Fukui function. *J. Phys. Chem. A* **109**, 3220–3224 (2005)
33. P. Fuentealba, P. Perez, R. Contreras, On the condensed Fukui function. *J. Chem. Phys.* **113**, 2544–2551 (2000)
34. P.W. Ayers, R.C. Morrison, R.K. Roy, Variational principles for describing chemical reactions: condensed reactivity indices. *J. Chem. Phys.* **116**, 8731–8744 (2002)
35. D.G. Oblinsky, B.M.B. VanSchouwen, H.L. Gordon, S.M. Rothstein, Procrustean rotation in concert with principal component analysis of molecular dynamics trajectories: quantifying global and local differences between conformational samples. *J. Chem. Phys.* **131** (2009)
36. P.M. Kroonenberg, W.J. Dunn, J.J.F. Commandeur, Consensus molecular alignment based on generalized procrustes analysis. *J. Chem. Inf. Comput. Sci.* **43**, 2025–2032 (2003)
37. J.J.F. Commandeur, P.M. Kroonenberg, W.J. Dunn, A dedicated generalized Procrustes algorithm for consensus molecular alignment. *J. Chemom.* **18**, U32–U37 (2004)
38. A. Tropsha, Best practices for QSAR model development. *Valid. Exploitation Mol. Inform.* **29**, 476–488 (2010)

39. B. Hoffman, S.J. Cho, W.F. Zheng, S. Wyrick, D.E. Nichols, R.B. Mailman, A. Tropsha, Quantitative structure-activity relationship modeling of dopamine D-1 antagonists using comparative molecular field analysis, genetic algorithms-partial least-squares, and K nearest neighbor methods. *J. Med. Chem.* **42**, 3217–3226 (1999)
40. W.F. Zheng, A. Tropsha, Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci.* **40**, 185–194 (2000)
41. J.C. Gower, G.B. Dijksterhuis, *Procrustes Problems* (Oxford UP, Oxford, 2004)
42. J.C. Gower, Multivariate analysis: ordination, multidimensional scaling, and allied topics, in *Statistics*, ed. by E.H. Lloyd (Wiley, Chichester, 1984), pp. 727–781
43. H.W. Kuhn, The Hungarian method for the assignment problem. *Naval Res. Logist. Q.* **2**, 83–97 (1955)
44. P. Gritzmann, V. Klee, On the 0–1-maximization of positive definite quadratic forms, in *Proceedings*, ed. by Operations Research (Berlin, Springer, 1988), pp. 222–227