Journal of Computer-Aided Molecular Design, **15:** 1065–1083, 2001.
KLUWER/ESCOM

Identifying potential binding modes and explaining partitioning behavior using flexible alignments and multidimensional scaling

Miklos Feher* & Jonathan M. Schmidt
SignalGene Inc., Unit 2, 335 Laird Road, Guelph, Ontario, N1G 487, Canada

Received 14 May 2001; accepted 28 November 2001

Key words: binding mode, binding orientation, clustering, conformations, flexible alignment, multidimensional scaling, partitioning, similarity

Summary

A method is described for the rapid and automatic analysis of flexible molecular alignments using multidimensional scaling and a normalized scoring scheme. A projection scheme was devised to separate orientational and conformational effects. It is shown that the approach can be utilized for the identification of common binding orientations or to the study of differences in partioning behavior. It is suggested that the method can be employed as a novel approach exploring molecular similarity as a dynamic property, so that it includes aspects of motion (by way of mutual orientations), conformations and molecular properties.

Introduction

In our recent publication it was shown [1] how metric and multidimensional scaling can be used to identify and visualize significant groupings of conformations and assign group membership to the conformers. The process is automatic and robust and was shown to separate clusters efficiently. In addition, the simplicity of determining the optimum number of clusters was demonstrated.

Molecular alignment is a useful method in drug discovery for establishing the similarity of molecules, finding pharmacophores and for setting up quantitative structure-activity relationships (for a recent review, see [2]). The methods MultiSEAL [3] and FlexAlign [4] were both shown to find the overlay in which the conformation and orientation of the molecules is most similar to that in the crystal. In both of these alignment methods, the strategy was always to select the top scoring alignment and compare only that alignment with the experimental data. When performing and evaluating alignments this way, the question often arose as to whether there is any additional information in the lower scoring alignment solutions and if so, how to extract it effiently. It became clear that a visual clustering tool, similar to that used for conformational clustering, would help to identify the most important relative orientations of the aligned molecules. It was hoped that this information might also be useful to predict potential alternative binding modes.

To our knowledge, automated clustering of molecular alignments has not been previously described in the literature. In order to perform such tasks, a number of novel steps have been introduced with respect to our previous work [3, 4]. The flexible alignment scores had to be normalized, so that they are independent of the number and types of atoms in the system. A scoring system had to be introduced for cases, where big parts of the overlapped molecules may differ. In addition, the confounding effects of relative position and molecular conformation had to be separated. The method of multidimensional scaling, previously tested for conformer clustering, had to be applied to cluster alignments. Finally, a consistent method had to be devised to test this alignment clustering procedure. These steps are described in detail in the next section.

*To whom correspondence be addressed. E-mail: miklos.feher@signalgene.com

Clustering alignments

The flexible alignment method generally produces a large number of overlays. Clearly, if the aim is to extract information from these, an efficient procedure had to be found to characterize and group these alignments. Because of the obvious similarities between clustering alignments and clustering conformations, the main steps in our recently developed process for conformer clustering [1] will be summarized first.

Our procedure for clustering conformers [1] is based on the application of metric and multidimensional scaling on the distance matrix. The distance matrix is a matrix of Euclidean distances obtained after pairwise rigid-body superposition of conformations. Usually only symmetry-unique atoms are considered in this distance matrix to avoid artificially increased distances that would be generated for symmetry equivalent atoms [1]. In general only distances between heavy atoms are included in this matrix. There is an extra pre-filtering step in which duplicates or too similar conformations are removed from the set. The initial estimates for multidimensional scaling derive from metric scaling. It was demonstrated [1] that the STRESS (acronym for standardized residual sum of squares [5]) is a good figure of merit for the quality of conformer clustering. The scaled coordinates are then clustered using a hierarchical scheme. It was shown [1] that cluster memberships in this case are identical to those obtained when the original distance matrix is clustered, provided the STRESS is below about 0.25. The optimum number of clusters is determined in an iterative process from the visual inspection of cluster boundaries from the multidimensional plot.

This method can be extended to clustering molecular alignments. However, the situation is much more complex compared to clustering single molecules. In the alignment of two molecules, three effects simultaneously conotribute to the r.m.s. distance of the overlays: the orientation of the two molecules, the differences in conformation of the first molecule and the differences in conformation of the second molecule. Clearly, these effects need to be separated in order that the individual contributions can be assessed.

We can obtain a distance matrix for alignments in the same manner as for single conformations, after pairwise rigid-body superposition of the alignment solutions. In other words, each entry in the distance matrix represents the minimized r.m.s. distance between two alignment solutions. In order to separate the conformational and orientational effects a simple projection scheme was developed, which is based on the properties of Euclidean distances. The total r.m.s. distance in the alignment of two molecules arises as a sum of three components: the distance that is dependent only on their orientation (d_{orient}), the conformational distance in molecule ($d_{1,\text{conf}}$) and that in molecule 2 ($d_{2,\text{conf}}$). If these distances are represented as vectors and the number of atoms in the two systems is n_1 and n_2, the total distance for a molecule pair (d_{tot}) is given as the vectorial sum of the three components, weighted by the respective number of atoms. Based upon this, the distance matrix resulting from only the relative orientations of the two molecules is given as follows:

$$\underline{\underline{d}}_{\text{orient}} = \frac{1}{\sqrt{n_1 + n_2}} \times \sqrt{\left|(n_1 + n_2)\underline{\underline{d}}^2_{\text{tot}} - n_1\underline{\underline{d}}^2_{1,\text{conf}} - n_2\underline{\underline{d}}^2_{2,\text{conf}}\right|} \tag{1}$$

It is important to remember that n_1 and n_2 in the above formula pertain to the actual number of atoms included in the clustering process (i.e., whether the clustering is performed on all atoms, selected atoms and/or heavy atoms only).

Prior to clustering, the alignments need to be sorted according to goodness criteria, such as the alignment score or the relative energy. As is the case for single molecule conformers, an optimal pre-filtering step can be used to eliminate duplicates and similar solutions. This is achieved by calculating the smallest matrix element in each row of the distance matrix and removing those solutions that are insufficiently different from previous ones. Subsequent metric and/or multidimensional scaling is used to visualize the results, which in this case include the distribution of orientations and the conformations for both molecules. It is also possible to perform hierarchical clustering of the original or the scaled distance matrices to assign cluster memberships, using the process previously reported for conformers [1]. This can be followed by assigning cluster representatives, i.e., finding the solution closest to the geometric center of the cluster. Because of the nature of the problems described in this paper, the hierarchical clustering step was not undertaken in any of the examples. As multidimensional scaling was more efficient than metric scaling in finding clusters in the data in all cases studied, only the results obtained with multidimensional scaling are shown in this work. However, the results from the two methods were qualitatively similar in all examples.

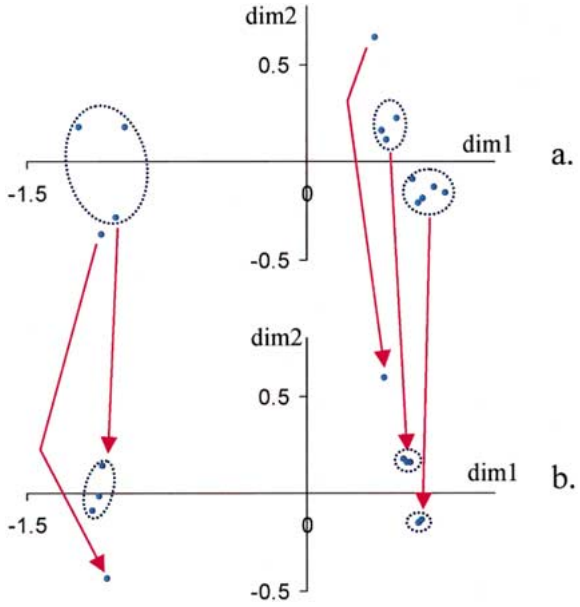

Figure 1. Multidimensional scaling of the best scoring alignments of two 1-fluoro-1,2-dihydroxi-ethane molecules based (a) on the original distance matrix (STRESS 0.07) (b) on the projected distance matrix after removal of conformational effects (STRESS 0.04). The removal of the conformational component from the distance matrix simplifies the identification of orientational families that exhibit the same overlay pattern and only differ in the conformation of one or both of the molecules. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

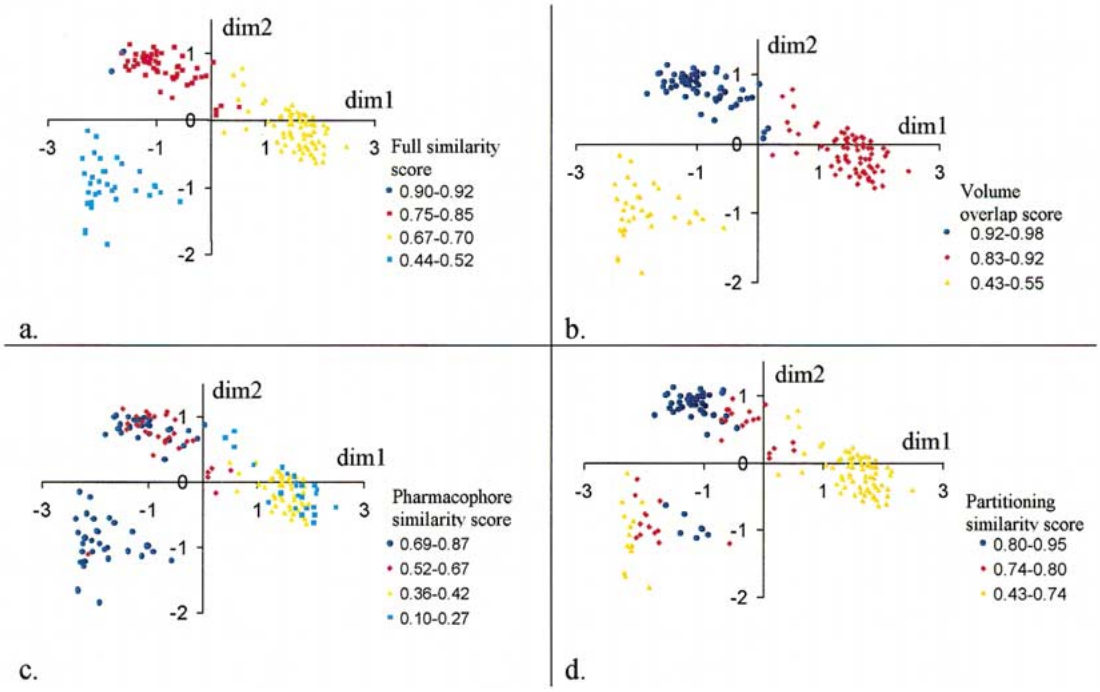

Figure 2. Multidimensional scaling of the orientational component from the flexible alignment solutions of 1-benzylsuccinate and glycyl-1-tyrosine (STRESS: 0.19). The overlay solutions are colored according to the magnitude of the score using (a) the default scoring scheme ($w_s/w_e = 3/1$, where w_s is made up of the volume and aromatic weights, w_e is a composite of hydrogen bond acceptor and donor weights) (b) the steric weights only (c) the donor and acceptor weights only (d) the atomic log *P* increment weights only. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

The projection scheme is demonstrated here with the hypothetical case of aligning two 1-fluoro-1,2-dihydroxy-ethane molecules using the best solutions from a flexible overlay [4]. Those alignments that have the same pattern of orientation (i.e., the same types of atoms are being overlaid) can be identified by visual inspection of the solutions, these are enclosed by dotted circles in Figure 1. The results of multidimensional scaling on the original and the projected distance matrices are shown in Figure 1a and 1b, respectively. As described above, the spread of the datapoints in Figure 1a is due to the mutual orientation and the conformation of the individual molecules, whereas only the mutual orientation of the two systems is responsible for the spread in Figure 1b. Clearly, projecting out conformational effects (i.e., the consideration of only the orientational component of the distance matrix) led to a decreased spread of the datapoints. This can be nicely observed, for example, on alignments on the right hand side in Figure 1 (these correspond to the fluorine atoms being on the same side in the two molecules): the three families in Figure 1a essentially collapse into three solutions in Figure 1b.

Normalized scoring schemes

An important aspect in the characterization of molecular alignments is the quality (or score) of the overlay. The flexible alignment procedure produces relative scores that successfully classify the alignments of a given molecule pair [4]. Although these scores are highly characteristic of the quality of the overlays, their numeric value is dependent on the types and number of atoms in the system. Clearly, a normalized scheme would permit the direct comparison of similarities between different molecule pairs. Such normalized schemes are introduced in this section.

The scoring function in the flexible alignment method has been described in detail [4]. In summary, the overlap scores can be expressed in the following way, given three properties defined by atomic property weights, u, v, w:

$$F = \sum_{i=1}^{n}\sum_{j=1}^{n'} \frac{u_i u_j' + v_i v_j' + w_i w_j'}{nn'} \times \left(\frac{a^2}{2\pi(r_i^2 + r_j'^2)}\right)^{3/2} \exp\left[-\frac{a^2}{2}\frac{|x_i - x_j'|^2}{r_i^2 + r_j'^2}\right], \quad (2)$$

where n and n' are the number of atoms in the two overlapped molecules, $x_1, \ldots, x_n$ denote the 3D positions of the atoms, r is the van der Waals radius. In this formula, a determines the breadth of the density, e.g., if $a = 2$ then approximately 90% of the density will be contained within the van der Waals radius. The calculated score depends on the number and type of atoms and their weights, as well as other parameters. Property weights have been defined for a number of different properties, such as the presence or absence of atoms, aromatic rings, hydrophobic atoms, H-bond acceptors and donors, as well as surface exposure, the value of atomic log P contributions, charge and molar refractivity contributions. The optimum weights required to reproduce experimental alignments have been reported [4] and these weights were used in this work in most examples.

In order to normalize the score, self-similarity needs to be defined. It is the overlap score of a molecule in a given conformation with itself in exactly the same position. Two approaches were tested to define a normalized alignment score. One is a simple normalization:

$$s_{\text{norm}} = \frac{1/F_{1,2}}{\sqrt{1/F_1 \cdot 1/F_2}} \quad (3)$$

Here the self-similarity scores are denoted as F_1 and F_2 and the overlap score of the two molecules as $F_{1,2}$. The reciprocals arise as the FlexAlign scores are inversely proportional to the amount of overlap. This normalization is analogous to that introduced for SEAL overlays [6]. Alternatively, we can introduce a Tanimoto-like similarity function for the two molecules, which essentially gives the ratio of the number of common 'features', divided by total the number of features:

$$s_{\text{sim}} = \frac{1/F_{1,2}}{1/F_1 + 1/F_2 - 1/F_{1,2}} \quad (4)$$

It is easy to see that this quantity is also equal to 1 in case of identity ($F_1 = F_2 = F_{1,2}$) and is equal to 0 in case of no overlap ($F_{1,2} \to \infty$). Although both normalized scores can be applied to compare the results from different overlays, the Tanimoto-like score was chosen because it is directly related to the number of common features. Hence this will be used exclusively in the rest of the work and will be simply referred to as the similarity score.

The flexible overlay of estradiol and raloxifene [4] illustrates the problems caused by groups with very different spatial occupancy. Raloxifene has a long

sidechain, which has no equivalent in estradiol, hence the position of this sidechain remained undetermined in the overlay. Nevertheless, the two molecules bind similarly to the estrogen receptor and it is beneficial to introduce a scoring scheme that only considers common elements.

$$s_{\mathrm{inc},1} = \frac{1/F_{1,2}}{1/F_1} \quad \text{and} \quad s_{\mathrm{inc},2} = \frac{1/F_{1,2}}{1/F_2} \tag{5}$$

These are essentially the ratios of common features to the total number of features in each set. In particular, the greater of these scores can be used as a measure to describe the inclusion of one set in the other. We shall denote this as the 'includedness' score. Again it is easy to see that this number is 1 if one set is fully included in the other (say $F_2 = F_{1,2}$) and is equal to 0 in case of no overlap ($F_{1,2} \to \infty$).

It must be added that the value of s_{inc} does not strictly lie between 0 and 1 in all cases. Under extreme circumstances it is possible that the overlap score of a molecule ($F_{1,2}$) with a similar but not identical other molecule is somewhat greater than its self alignment score. This arises as the overlap function decreases exponentially with distance and in some rare cases better 'interactions' may be possible between the two dissimilar molecules than between the identical ones. The difference, however, is small and the includedness score in all tested cases has been below 1.01. This small numerical error has very little impact on the usefulness of the scheme.

Alignment-based definition of similarity

A normalized scoring scheme, such as the Tamimoto-like score defined above, allows the similarity of two molecules be defined using their flexible alignment. Furthermore, the use of different property weights in the scoring function, coupled to the fact that all alignments can be studied simultaneously via multidimensional scaling, opens up the possibility of comparing molecules in different positions and conformations using different properties. The default properties (volume, presence of aromatic atoms, hydrogen bond donors and acceptors) and weighting scheme (a ratio of weights between the steric and electrostatic contributions of 3:1) have been optimized to predict the experimental binding orientation of the ligands [4]. By including specific terms in the overlap function, we can define different aspects of similarity between two molecules. For example, by including only pharmaphoric measures (e.g., hydrogen bond acceptors and donors) as atomic weights in the overlap function (F, see above), we can define pharmacophore similarity. This score is a measure of similarity between the two molecules based exclusively on their hydrogen bonding properties. In a similar manner, we can define hydrophobic similarity by only including weights for hydrophobic contributions. Bulk similarity and surface similarity would express the extent of the overlap of van der Waals volumes and the accessible surfaces, respectively. By applying selected property weights, such as the atomic log P increments, we can evaluate the similarity of two molecules in terms of the steric distribution of functional groups contributing to octanol-water partitioning. More sophisticated similarity measures can also be devised as linear combinations of such parameter weights.

A given set of alignments can be rescored using different similarity functions (i.e., using different parameters and weights). In order to gain some insight into the distribution of alignment solutions with any of these scoring schemes, they can be displayed as color coding of the alignments in the multidimensional plot. An important feature of this similarity scoring system is that it can be applied to rescore any set of alignments, even if these were not produced using FlexAlign.

Rescoring alignments using different scoring schemes can be demonstrated by the previously studied example [4] of the flexible alignment of 1-benzylsuccinate and glycyl-1-tyrosine. The flexible alignment generated 1000 solutions, of which the 161 most different were selected (these differed from all other by an r.m.s. distance of at least 0.5 Å). However, instead of simply studying the best scoring alignment in relation to the experimental structure, the solutions after the multidimensional scaling of the projected distance matrix can be plotted, as shown in Figure 2. The points in Figure 2a are colored according to the magnitude of the similarity score that had been optimized previously to reproduce the X-ray alignment [4]. There are two high quality solutions (shown as dark blue dots in Figure 2a). The corresponding overlays differ mainly in their conformations but have almost identical relative orientations, as shown by their proximity in the projected plot. We can conclude from the plot that there are essentially three other types of alignments and these differ greatly in quality. The solutions in the first group (pink squares in Figure 2a) have somewhat similar relative orientation of the ligands as the two best scoring alignments.

The other two sets of solutions (shown in orange and blue) are of considerably worse quality. When all the overlays are rescored using only the similarity of molecular volumes (see Figure 2b), the two overlays, shown as dark blue dots in Figure 2a, are no longer greatly different from those with similar orientation. This is not surprising, as the high scores with the original scoring function arose as a result of simultaneous overlay of both steric and pharmacophore features. When only pharmacophore similarity is scored in the overlays (see Figure 2c), different quality solutions are not separated into clear clusters and the relative quality of alignments in the second and third groups changes. Finally, scoring the same alignment solutions using atomic log *P* increments (see Figure 2d) should help to choose orientations of the two molecules in which the spatial distribution of groups contributing to octanol-water partitioning is most similar.

Computational details

All modeling work, including the flexible alignments, was performed using the Molecular Operating Environment (MOE) suite of programs [7]. The MOE implementation of the MMFF94 [8] force field was used in all energy calculations. In order to emulate the effects of water as a solvent, a continuum solvation model [9, 10] was employed. Molecular geometries were optimized using the truncated Newton optimizer preceded by two steps of steepest descent within MOE [7]. The minimization was terminated when the RMS gradient fell below 0.001. Conformational searches were carried out with the Random Incremental Pulse Rearch (RIPS) method [11]. The octanol-water partition coefficients at given pH values (log *D*) were evaluated using the ACD log*D* suite [12].

The calculation of the distance matrix, the separation of the conformational and orientational contributions and the pre-filtering steps were implemented in the SVL programming language of MOE [7]. The distance matrix was calculated from the pairwise Euclidean r.m.s. distance between the same atoms of two alignments after their optimal rigid body superposition (i.e., by minimizing the weighted least squares error function). Symmetry equivalent atoms were identified using the previously described procedure [1] and eliminated from the calculation of the distance matrix. The pre-filtering step was similar to that described for conformers [1] except that the selection criterion was based on the full or the projected r.m.s. distance between alignments.

It is implied in the text that the proximity of two points in a multidimensional plot indicates that the corresponding alignments are similar. Strictly speaking, such a conclusion could only be drawn from metric scaling, since the distances in a multidimensional plot only correlate with the rank order of the original distances. However, metric scaling was performed in each of these cases and it was ensured that the plots from metric scaling are similar to the multidimensional ones.

Metric and multidimensional scaling, as well as hierarchical clustering, were performed using the NAG statistical add-in for Excel [13]. This tool has recently been shown [14] to produce results identical to the NAG libraries and is free of the errors that have been documented for Excel itself [15]. The only current limitation to the program is the size of the Excel spreadsheet (256 columns), which makes pre-filtering necessary in most cases.

Calculations were performed on a 800 MHz Intel Pentium III processor running Windows NT. It has been shown that an alignment similar to that in the crystal is usually produced in the first 1000 alignments (although it is typically whitin the first 100). The most time-consuming process in the evaluation of alignments is the calculation of the distance matrix, where the CPU time depends on the number of atoms in the two molecules and the total number of alignments. For example, in the case of 1-benzylsuccinate and glycyl-1-tyrosine, it took about 90 s to produce the distance matrix for 1000 alignments. If only 200 selected overlays are considered, the distance matrix is generated in only 5 s. Scaling and visualization rarely require more than a few seconds of CPU time. This makes the methods ideally suited for the rapid evaluation of molecular alignments.

Results and discussions

The methods described above provide a powerful means to investigate binding property differences between molecules. Flexible alignments with an optimized scoring scheme can be applied to predict the binding orientation in good agreement with experimental data [4]. A logical extension of this idea would be to expect a set of overlays to predict different possible binding orientations and conformations of a ligand pair. There is an important caveat to this: the same

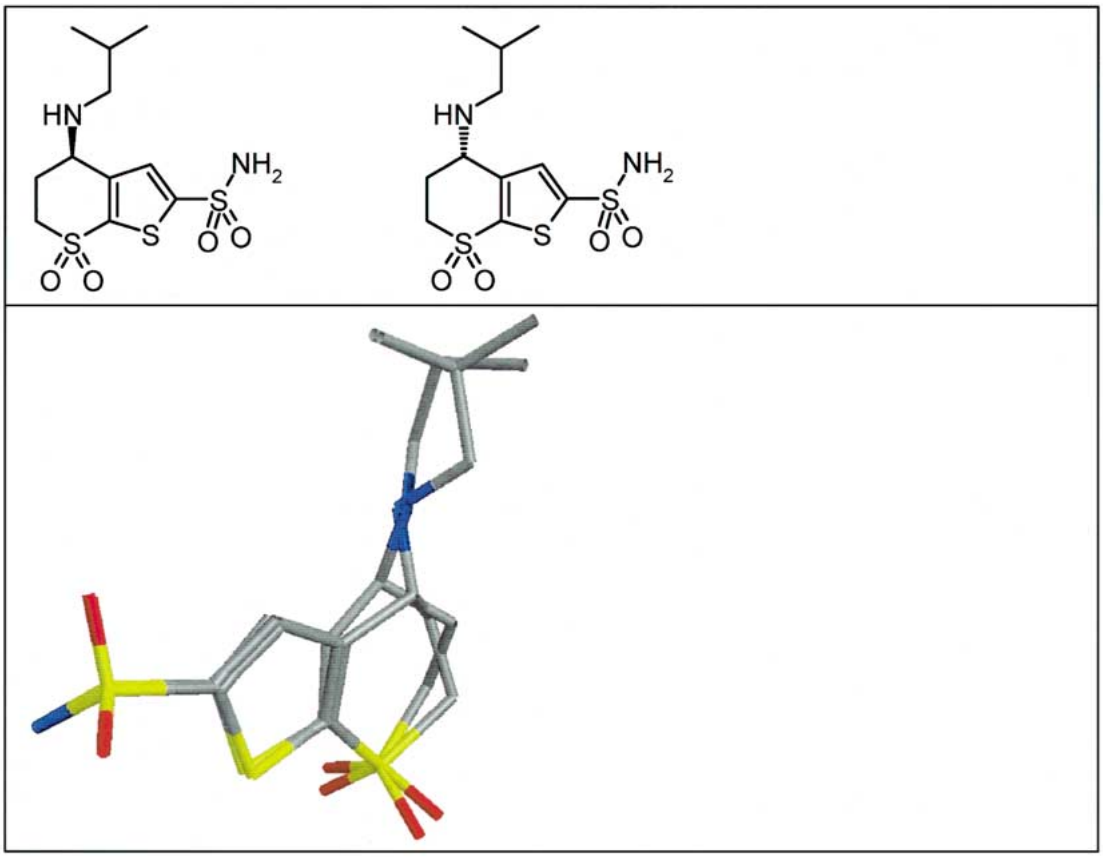

Figure 3. The best scoring overlay from the flexible alignment of two enantiomers of a sulfonamide (structural formulae are also shown).

functional groups of the receptor should be involved in the binding of both species, as the overlays would be unable to extract information on the process otherwise. The first four examples in this section will show how this kind of information can be extracted from the alignments using multidimensional scaling. It is important to mention that this process is aimed at identifying 'potential' binding modes, independently of whether these modes have been or can be identified experimentally. As these potential modes indicate the possibillity of different binding orientations from the ligand structure only, steric or other factors in the receptor structure may preclude the existence of some of these modes. On the other hand, if there is a possibility of an alternative arrangement, the described method should identify it.

Another possible application is the study of transport processes. Most molecules interact differently with aqueous and a non-aqueous media. As a result, transport from one medium to the other may invoke a conformational change. The difference in the partitioning among molecules between the two media can often be explained by differences in the extent that they can undergo such conformational changes [16]. In the context of molecular alignments, such differences are expected to manifest themselves in the internal energy of the best alignment solutions. The last two examples in this section were selected to demonstrate how flexible alignment and scaling can be applied to detect differences in blood-brain barrier penetration between molecules, which are known to bind in the same manner to their targets.

The binding orientation of two enantiomers: The flexible alignment of carboanhydrase ligands

As a first example, two enantiomers of a sulfonamide (see Figure 3) were examined, both of which are ligands of carbonic anhydrase. The two enantiomers both bind to the enzyme but differ by a hundredfold in affinities [17]. The X-ray structure of these molecules shows that they bind in a similar fashion [17] and their greatly different affinities are explained by the fact

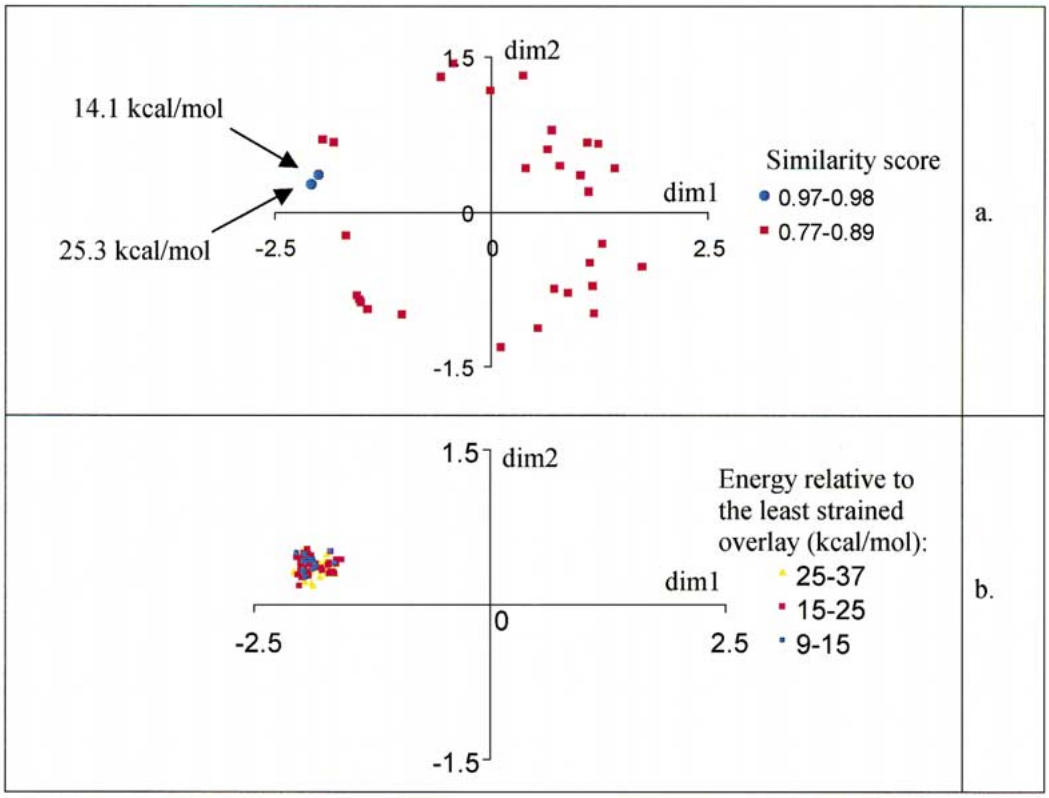

Figure 4. Plot of the coordinates after multidimensional scaling of the orientational component for two enantiomer sulphonamides, the structures of which are shown in Figure 3a (a) the 32 most different alignment solutions (total rmsd >0.8 Å from all higher quality solutions, STRESS: 0.23). There are only two high quality solutions and they appear close to each other in the plot, indicating that they represent similar molecular orientations and differ only in their conformations. The relative energies of the two best solutions are also indicated. (b) the solutions with a normalized similarity score over 0.95 (94 solutions altogether, STRESS: 0.14). This plot is shown on the same scale as the one before to demonstrate the orientational similarity of the obtained overlay solutions. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

that one of the enantiomers assumes a highly strained conformation on binding. Our aim was to determine how the difference between the enantiomers can be observed using flexible alignments and multidimensional scaling.

For the flexible alignment of the two ligands 1000 overlays were produced and the 32 highest scoring ones that differed from all other ones by at least 0.8 Å in total r.m.s. distance were selected for further analysis. Next, the projection was performed according to Equation 1 so that the distance matrix contains only the orientational component. The result of multidimensional scaling on the projected distance matrix is shown in Figure 4a. There are two high quality solutions, shown as blue dots in this plot. Although the full r.m.s. distance of the corresponding alignments is 1 Å, their projected r.m.s. distance is a mere 0.3 Å and hence the two overlays are almost concurrent in Figure 4a. This arises because the major difference between the two alignments is in the conformation of the two molecules that is removed by the projection. The internal energy of these solutions, relative to the lowest energy conformations is also displayed in Figure 4a. These alignments have substantial internal strains, as one enantiomer cannot assume a low energy conformation required for binding. This conclusion is in agreement with that derived from the X-ray structure [17].

In order to ensure that the detected strain energy does not arise as an artifact of the prefiltering step, the distance matrix of all alignments with similarity scores above 0.95 was also projected and scaled. The resulting plot is shown in Figure 4b. The applied scale was deliberately chosen to be the same as in Figure 4a. The center of mass of the scaled coordinates was translated from the origin, so that the coordinates of the

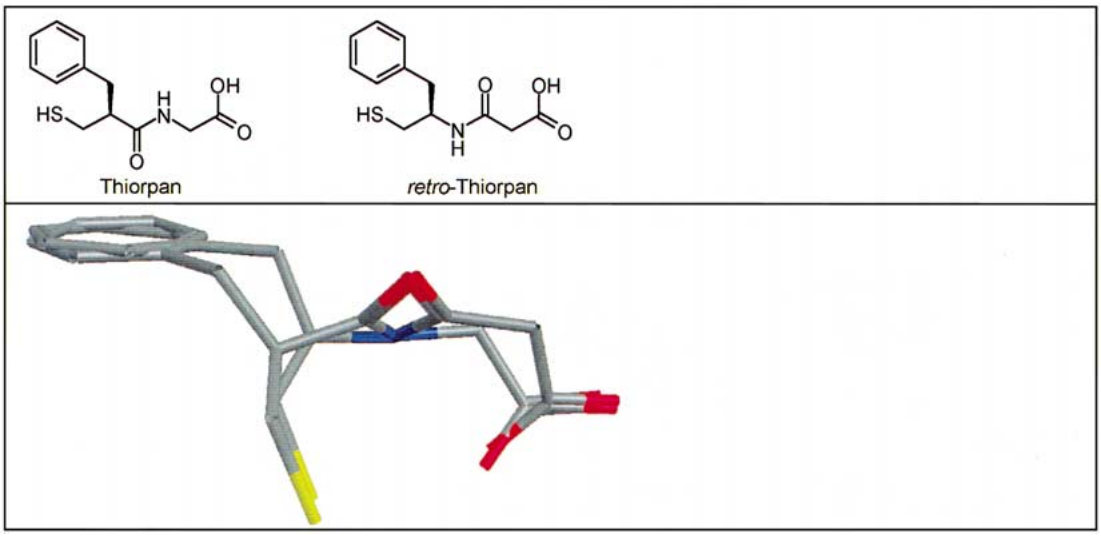

Figure 5. Results from the flexible overlay of the two thermolysin inhibitors, thiorpan and *retro*-thiorpan, showing the best scoring solution, which is also similar to the X-ray alignment.

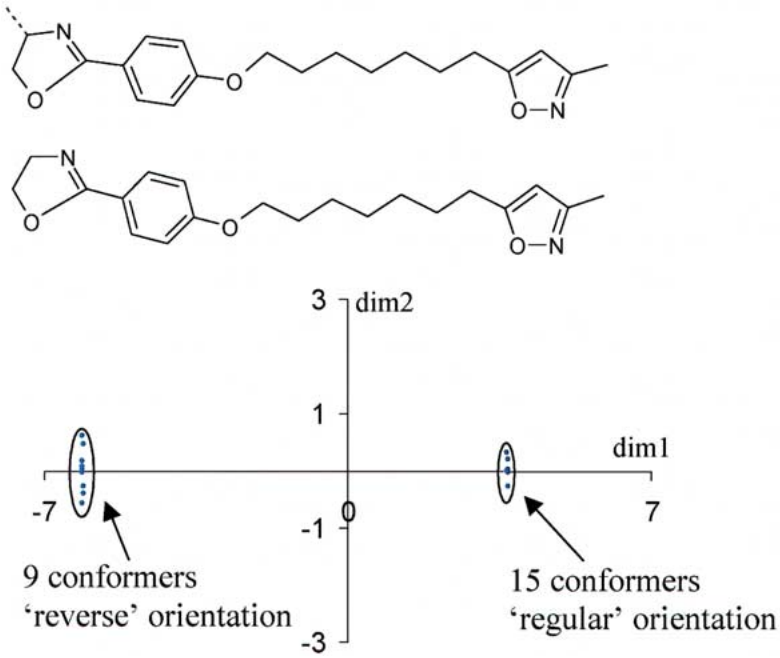

Figure 7. Results from a semirigid alignment of two ligands of the viral coat protein of the human rhinovirus, displaying the plot of the coordinates from the multidimensional scaling of the projected distance matrix (STRESS 0.0002). The two clusters, corresponding to the 'regular' and 'reverse' alignments are clearly separated in the plot. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

two annotated points in Figure 4b remain identical in Figure 4b. The comparison of the spread of points between Figure 4a and 4b indicates that all the best scoring solutions lie close to the two good solutions in Figure 4a and hence they all represent highly similar alignment solutions. Most solutions have high strain energies. There are three overlays with relative stain energies of 9–10 kcal/mol, although these have scores of 0.95–0.96 and hence are less favorable than the best overlays. As the Ki's of low nanomolar inhibitors correspond to a binding free energy of around −11 to −12 kcal/mol, the calculated strain energies are substantial and might indeed explain a drop of activity by two orders of magnitude.

Although the above explanation is solely based on flexible alignments, it is identical to the conclusions based on the X-ray structure [17]. A major benefit of the described approach is, however, that it enables the conclusions to be drawn within seconds, provided flexible overlays are available.

Similarity in the binding of retroamides: The flexible alignment of two thermolysin inhibitors

The thermolysin inhibitors thiorpan and retrothiopan (see Figure 5) differ only in the directionality of their peptide bonds. Despite this, X-ray crystallography shows that they bind in an analogous way [18]. Based on previous experience with the flexible alignment program [4] it was expected that the top scoring solution with the optimized parameterization would match the X-ray alignment. Additionally, we also wanted to test whether other potential relative binding orientations were indicated by the rest of the alignments. As the best way to characterize the solutions in this case is considering only the overlay of the pharmacophoric atoms (donors, acceptors, charge centers and aromatic centers), only these were included in the calculation of the distance matrix. This kind of alignment is essentially equivalent to conventional pharmacophore searches.

In this example, the best 250 solutions from the 1000 produced by the flexible alignment were investigated. The top scoring solution from the flexible overlays, shown in Figure 5, corresponds to the experimental alignment [18]. In addition, rapid insight into the behavior of the other alignments can be gained using multidimensional scaling. The construction of the distance matrix and multidimensional scaling took about 5 s CPU time on an 800 MHz PC. The resulting plot is displayed in Figure 6. The plot is color-coded using the full similarity score in Figure 6a and the pharmacophoric component of the score in Figure 6b. In these plots three groups of alignments can be identified. The majority of the overlays are contained in a tight group of 208 solutions, in all of which the pharmacophoric groups are paired in the same manner as in the best scoring solution. The difference among these 208 solutions is the actual conformations of the molecules, as well as the quality of the matches between the corresponding features in the two molecules. The solutions have relative strain energies of 0–25 kcal/mol compared to the lowest energy solution and the distance between the corresponding centers in the alignment is less than 1.2 Å. It can be seen from Figure 6a that the alignments on the top part of this cluster are of somewhat lower quality, while Figure 6b indicates that the good steric match between the molecules compensates for the weaker pharmacophoric overlap. The weaker pharmacophoric overlap in all of these cases arises as a result of the imperfect pairing of the carboxylic moieties.

As shown in Figure 6, there are two more groups of solutions among the top 250. Both of these clusters contain fewer solutions (36 and 6, respectively) and the clusters are more elongated. As examination of the individual solutions reveals, the common feature between these two groups is that all alignments miss exactly one pharmacophore contact, whereas the ones in the third group simultaneously have another pair mismatched (such as an amide oxygen overlayed on one of the carboxylic oxygens). Based on the similarity score the best quality solution in the second goup ranks 165th, while the best in the third group ranks 205th. The range in the full similarity scores among the first 250 solutions is rather narrow (0.90–0.97) because favorable steric match can compensate for the lost contacts. Obviously, the range in the pharmacophore similarity score is wider (0.68–0.99).

Two principal conclusions can be drawn from this example. First, flexible alignments can reproduce the analogous binding mode of the two molecules despite the different directionality of their peptide bonds. Second, multidimensional scaling coupled with the projection scheme successfully separated the alignments according to the overlay of the corresponding functional groups. The multidimensional plot indicates that there is no conceivable high quality overlay of the two molecules that has a different pairing of functional groups. Therefore, the experimental binding mode is likely to be the only expected binding orientation in this example.

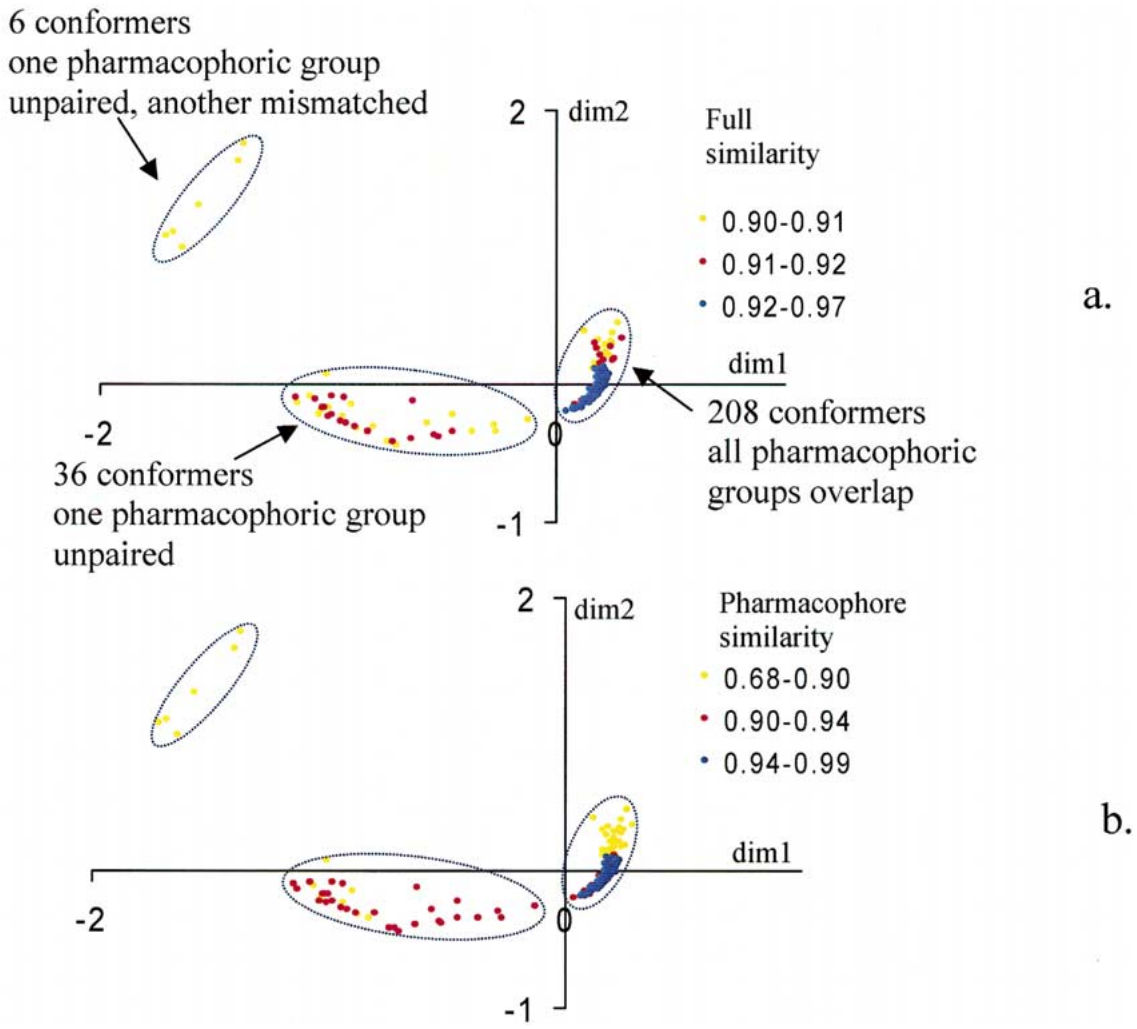

Figure 6. Plot of the coordinates from multidimensional scaling (STRESS 0.066) of the projected pharmacophore distance matrix of the 250 best solutions. The majority of the alignments have all pharmacophoric groups aligned and differ only in the conformation of the rest of the molecule. Some of the lower quality solutions miss one interaction. The plots are colored with (a) the full similarity score (b) the pharmacophore similarity score. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

Establishing binding orientations from semirigid alignments: The binding modes of two antiviral compounds

Rather surprising binding modes have been reported for two viral coat protein ligands of the human rhinovirus, the structures of which are shown in Figure 7. Although the two molecules are highly similar, X-ray crystallography shows that they bind in a reversed direction [19]. Our interest was to determine whether flexible alignment followed by the multidimentional scaling of the solutions would predict this unusual binding orientation.

Molecular overlays were generated using the default scoring system. Because of the highly flexible nature of the molecules, it was decided that only a part of the available conformational space would be searched, using semirigid alignments. In this process, the torsional angles were not varied initially. Instead, a random pre-alignment of the two molecules was performed, followed by the minimization of the objective function containing similarity and internal energy [4]. The termination criterion was 1000 failures to generate a new alignment in a row. This process resulted in 14 different solutions, of which only nine had similarity scores greater than 0.8. Of these, eight had the regular

and one the 'reverse' orientation. Hence, the two potential binding modes could indeed be predicted using flexible alignments.

The alternative binding orientation has been explained in the literature in terms of the symmetry of the electrostatic potential surfaces of the two molecules [20]. In order to reproduce such effects, overlays were generated in which charge was applied to mimic the effect of electrostatic potential and volume replaced steric similarity. The relative weight for the two properties was 1:8, based on previous optimization of SEAL overlays [6] with similarity functions analogous to the ones in this work [4]. The flexible overlays generated a total of 24 alignments that differed in the relative orientation of the four heterocyclic rings.

The analysis process (projection, distance matrix generation, multidimensional scaling and plotting) took less than 1 s of CPU time on an 800 MHz PC. Two groups of solutions can be identified in the resulting plot (see Figure 7), corresponding to the 'regular' and the 'reverse' binding modes. The similarity scores of the two molecules in their best 'regular' and 'reverse' orientations are 0.98 and 0.84. Although the number of alignments in this case is small, the example still demonstrates the ease with which the possible binding orientations can be recognized with the method. These results also support the explanation that the existence of the second binding orientation is due to the similarity and symmetry of electrostatic potentials [20].

The prediction of possible binding modes from flexible overlays: The alignment of the estrogen-α receptor ligands raloxifene and estradiol

The flexible alignment of estradiol and raloxifene was previously performed [4] and was found to correctly reproduce the experimental X-ray alignments [21]. It has been shown above how flexible alignments can be used to identify potential binding modes. Our aim was to test, whether flexible alignments would predict single or multiple binding orientations for this ligand pair.

Although only a single binding mode was observed under the condition of crystallization [21], there is considerable evidence for the existence of an alternative binding orientation. Two different binding orientations for estradiol have been postulated from docking experiments to the estrogen-α receptor (with the two hydroxyl ends of the molecule exchanged) [22]. There is also evidence for alternative binding orientations of estadiol in case of other related receptors [23–25]. Similarly, two possible binding modes of raloxifene were predicted by docking experiments to the estrogen-α receptor [26]. Thus, it is intriguing to determine whether flexible alignments would predict the second binding mode.

Two issues had to be dealt with when comparing the flexible overlays of estradiol and raloxifene with experimental X-ray structures. One was that raloxifene has a long sidechain that has no equivalent in estradiol. It was concluded [4] that the conformation of the sidechain cannot be decided from the overlays and hence this sidechain was simply excluded from the fits. Even in this case, the includedness score (see Equation 5) can be applied to account for the similarity of the two molecules. The other issue was that although flexible alignments attempt to overlay both hydroxyl groups, the X-ray structure indicates that oxygens in one pair are far apart in the binding pocket (4.7 Å from each other). Good agreement (rmsd: 0.65 Å) with the actual orientation in the crystal can be obtained by simply excluding this pair from the overlays [4]. However, excluding certain atoms requires prior knowledge of the experimental structure and our aim was to utilize flexible alignments in the absence of such structure. If this oxygen is included in the fit, the agreement with the experimental structure is still acceptable (rmsd: 1.4 Å). Hence no groups on the main skeleton of the two molecules were excluded from the overlays in this work.

In the flexible alignment of raloxifene and estradiol 1000 alignments were produced to get a representative set of overlays. On studying the results it becomes immediately clear that at least within the 15 top-scoring solutions two possible orientations are present. These are shown in Figure 8 (without the sidechain of raloxifene). In order to find out about the rest of the alignments, the distance matrix was calculated for the 146 most different overlays. The selection was based on the projected distance matrix, with the solutions more than an r.m.s. distance of 0.5 Å from all others being selected. As the aim was to establish potential binding orientations irrespective of conformations, only the orientational component of the distance matrix was considered (i.e., the projected distance matrix). The plot of the coordinates after multidimensional scaling is displayed in Figure 9. The best-scoring data points (shown as blue dots in Figure 9) fall into two well-defined clusters, containing 29 and 21 alignments, respectively. These clusters are relatively far from each other in the plot, as they correspond to the head-

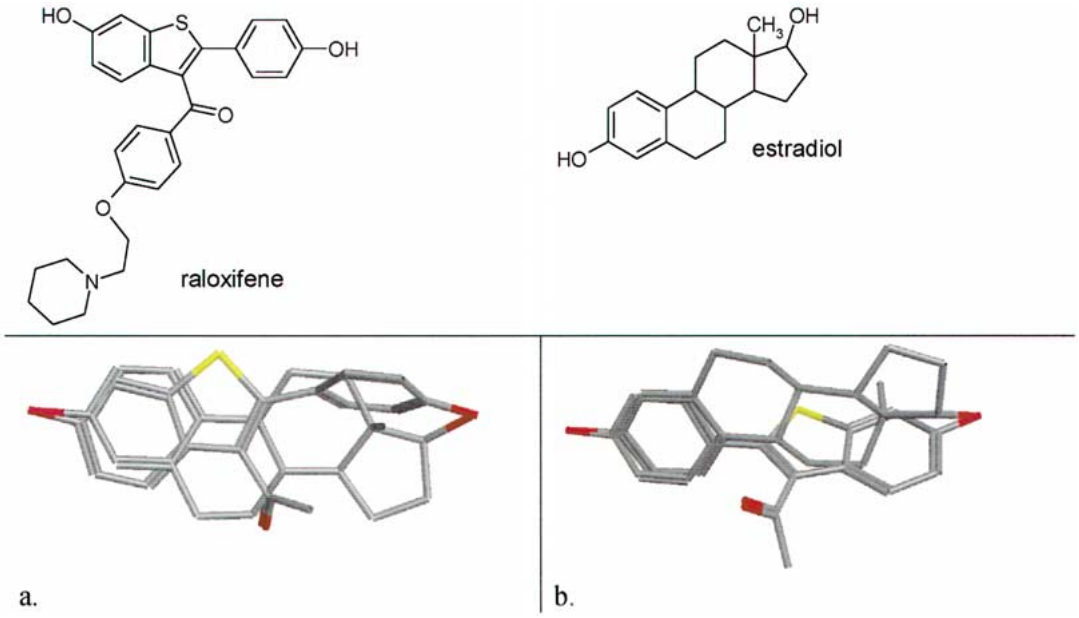

Figure 8. The two observed binding orientations in the flexible alignment of estradiol and raloxifene without considering the tail section (a) the highest scoring head-to-head solution (includedness similarity score 0.98) (b) the highest scoring head-to-tail solution (includedness similarity score 0.97). For clarity, the tail section is also omitted from the figure.

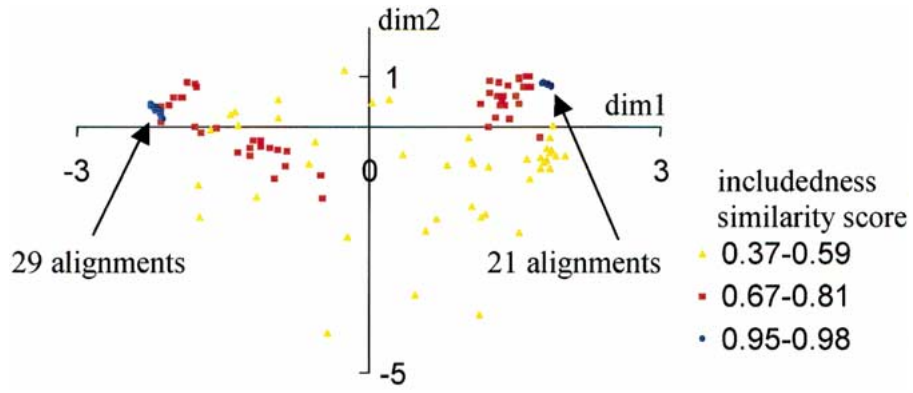

Figure 9. Multidimensional scaling of the orientational component of the most different overlay orientations (projected rmsd >0.5 Å from all previous solutions) for the estradiol-raloxifene pair with the tail section excluded from the fit (STRESS: 0.19). The plot is colored by the includedness similarity coefficient of the overlays. The best scoring solutions, shown in blue, form two well-defined clusters in this plot, which may be associated with two possible mutual orientations of these molecules during binding. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

to-head and head-to-tail solutions. Even some of the worse scoring points (shown as pink squares) are scattered around these two clusters and only the worst solutions (shown in orange) do not appear to cluster at all. The best-scoring overlays cluster so tightly because the solutions within each cluster differ only in the conformation of the individual molecules but not in their relative orientations. The two orientations of the estradiol-raloxifene pair (see the two clusters of blue dots in Figure 9) may be associated with two possible binding orientations of these molecules. The similarity scores and the number of solutions within these clusters are quite similar, hence they are expected to be similarly likely from a purely probabilistic point of view.

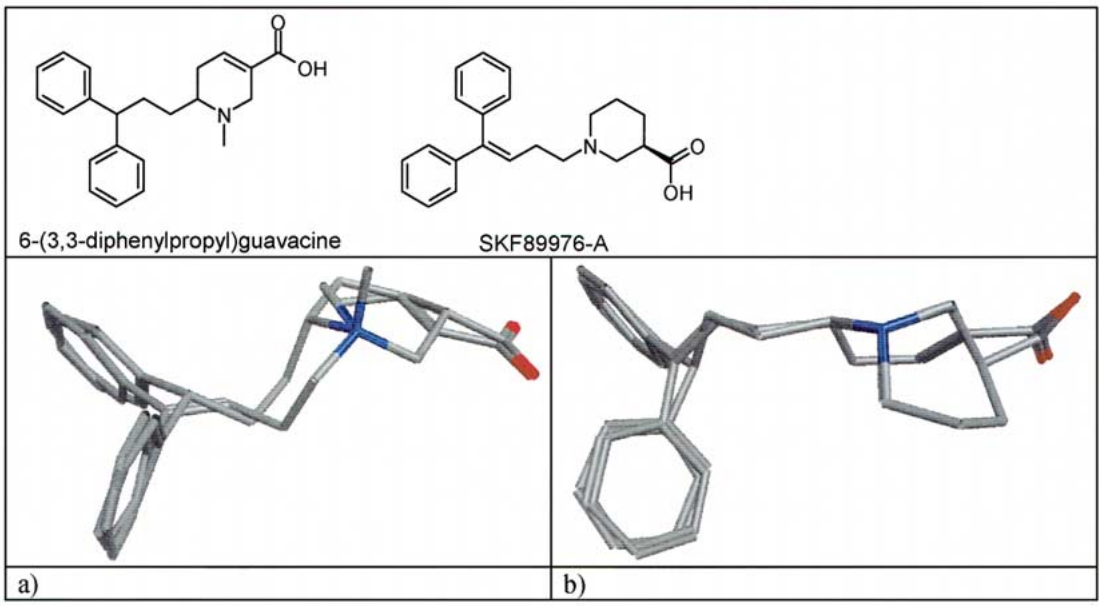

Figure 10. Calculated alignments for the GABA uptake inhibitors 6-(3,3-diphenylpropyl)guavacine and SKF8997-A (a) the best scoring solution *in vacuo* (b) the best scoring solution in an aqueous environment (i.e., using a continuum solvation model)

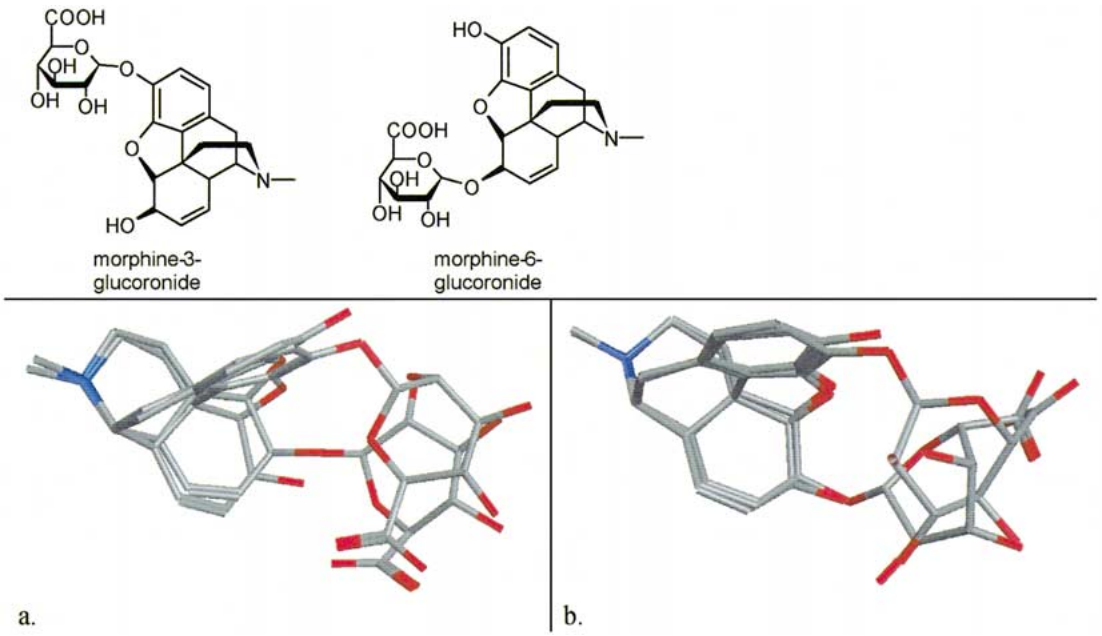

Figure 12. The results of the flexible overlay of morphine-3-glucoronide and morphine-6-glucoronide in (a) vacuum (b) in water.

It is important to note that although in this case the second possible orientation could be detected in the first few alignment solutions, this may not always be the case. Even then, however, multidimensional scaling can help to map the possible relative orientations of the aligned molecules and thus is beneficial in rapidly identifying potential alternative binding orientations.

Differences in behavior in binding and permeability: The flexible alignment of two GABA-receptor antagonists

In the experimental study of two GABA-uptake inhibitors [27], 6-(3,3-diphenylpropyl)guavacine and SKF 89976-A (see Figure 10) it was found that although the two molecules have similar binding affinities in vitro, only the latter one is active *in vivo*. The authors attributed this to differences in permeability through the blood-brain barrier. It was later proposed [28] that the effect may result from a stabilization of either the neutral form or the zwitterionic form of SKF 89976-A by an intramolecular hydrogen bond.

Flexible overlays model the ability of two molecules to assume similar conformations. As the two molecules have similar binding affinities, they are likely to be able to adopt similar conformations in an aqueous environment. On the other hand, the two molecules appear to differ in their ability to cross from an aqueous into a non-aqueous environment while passing through the blood-brain barrier, even though they have highly similar structures. The lipophilicities of guavacine and the SKF molecule are expected to be similar, based upon their $\log D$ values at $\mathrm{pH} = 7$ of 2.67 and 2.87, calculated in this work. This makes it likely that the dissimilarity between the two molecules indeed lies in the differences in their conformational behavior. Our aim was to investigate how flexible overlays with the described scoring and scaling tools can be applied in the study of such problems.

In order to study this transport process, 1000 alignments were produced both with continuous solvation and in vacuum. The best scoring alignments in the two media are shown in Figures 10a and 10b. As can be seen in these Figures, the functional groups and much of the bulk in both media are overlaid in a similar fashion. The only major difference between the two best scoring solutions is the overlay of the piperidine and the tetrahydro-pyridine rings; *in vacuo*, the two rings closely overlap, whereas they are practically perpendicular when a solvation model is employed. Among the ten best solutions, the overlays in vacuum and in water become qualitatively similar and the corresponding scores are also comparable. There is, however, a major difference in the relative internal energies of these solutions. Whereas the best solutions in vacuum have relatively high internal energies (25–33 kcal/mol, compared to the sum of the energies of the relaxed molecules), the analogous aqueous solutions are much less energetic (7–14 kcal/mol). The high strain energy *in vacuo* arises as a result of an internal hydrogen bond being formed in all low energy solutions of the SKF molecule. As this internal hydrogen bond has no equivalent in guavacine, good alignments between the two molecules can be achieved only at the cost of straining the molecules. In contrast, internal hydrogen bonds are less favorable in water (as a result of preferential hydrogen-bond formation with the solvent) and thus less strain in required to align the two systems properly. In conclusion, flexible alignments neatly show that the passage from an aqueous to a non-aqueous environment induces significant changes that render the two molecules dissimilar.

Multidimensional scaling was applied to gain information about the remaining alignments. A diverse subset of alignments was selected (those which had a minimum r.m.s. distance of 1 Å from all the others) and the orientational component of the distance matrix was scaled. Figure 11a presents a projected map for the alignments *in vacuo*, whereas Figure 11b shows the plot for the overlays in water with the coloring scheme representing the quality of the alignments. As conformational effects were projected out, proximity of the points in the map relates to the differences in the relative orientation of the two molecules. Figures 11a and 11b illustrate that the distribution of similar quality solutions is similar in both plots. It can be concluded that there is only one high-quality common orientation of the two systems, which can still be attained through a number of conformations.

The major difference between alignments in vacuum and in water can be discerned if the relative strain energy of these overlays is also examined. Figures 11c and 11d display the 250 best scoring overlays, with coloring of the points representing the relative strain energy. We can see that the first 250 overlays cover a relatively small area in the plot (essentially within the $[-1,1]$ interval along both axes). This arises because all top scoring overlays place the carboxyl groups on top of each other, limiting the opportunities for alternative overlays. The coloring scheme reveals the high internal energy of all vacuum alignments; there is essentially no convincing low energy overlay between the two molecules. In contrast, the solutions in water have much lower internal energies. Hence the multidimensional plots also support the conclusion described above that guavacin and the SKF molecule are similar in an aqueous environment but they lose their similarity on passing to a non-aqueous medium, due to differences in their conformational behavior.

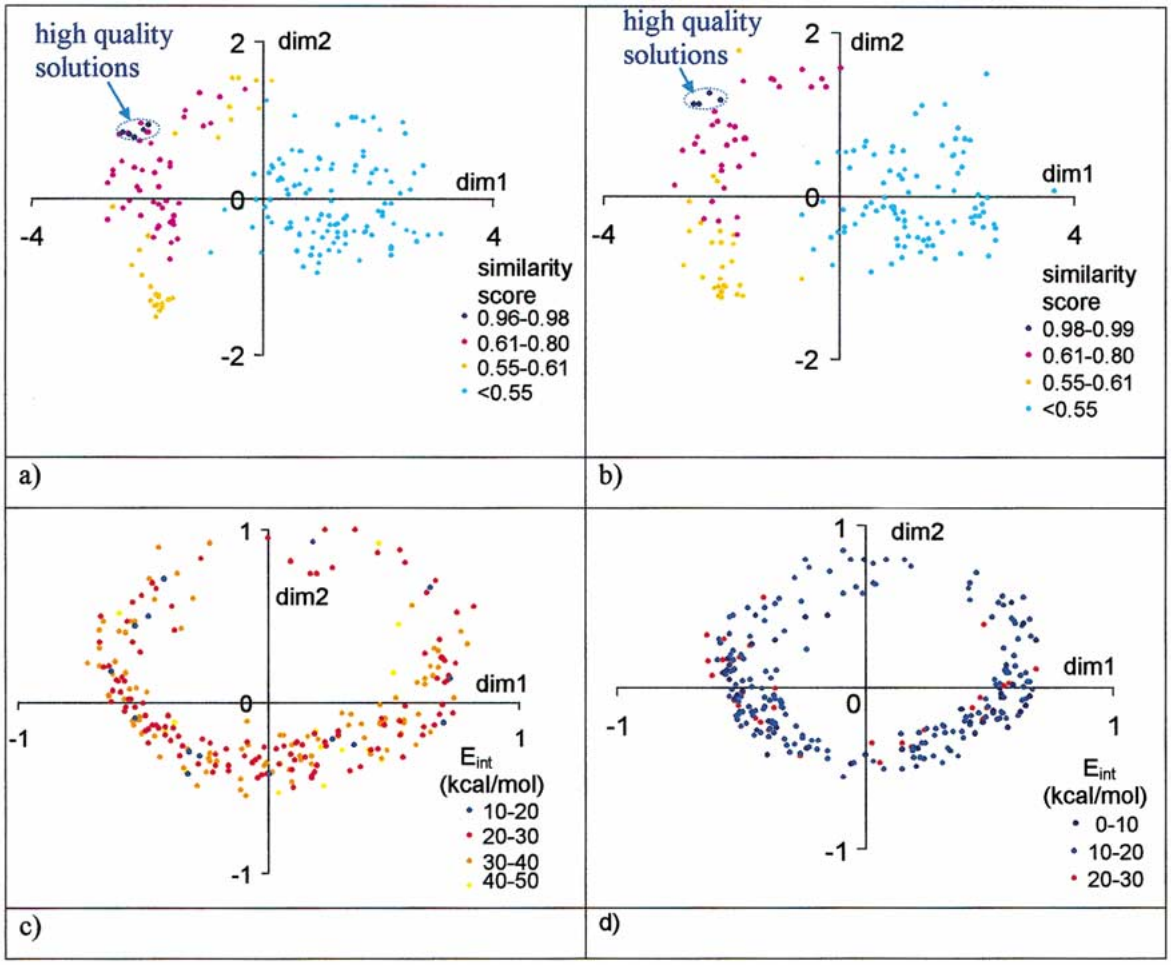

Figure 11. Multidimensional scaling of the orientational components from flexible alignments for the GABA uptake inhibitors 6-33,diphenyl-propyl-guavacine and SKF89976-A (a) the most different alignment solutions (total rmsd >1 Å from all previous solutions) *in vacuo*, colored by the similarity score of the overlays (STRESS: 0.12) (b) same in water (STRESS: 0.12) (c) the 250 best solutions *in vacuo*, colored by the internal energy (STRESS: 0.19) (d) same in water (STRESS: 0.16). The plots show that although the general distribution of the alignments is similar in vacuum and in water, the two sets differ considerably in the internal energy of the solutions. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

The detection of chameleonic behavior: The flexible alignment of morphine-3-glucoronide and morphine-6-glucoronide

The molecules morphine-3-glucoronide (M3G) and morphine-6-glucoronide (M6G) are the two major metabolites of morphine. Despite their polarity and similar lipophilicities, M6G has been shown to penetrate well through the blood-brain brrier, while M3G penetrates to a lesser extent [29]. As in the previous example, our aim was to test whether flexible alignments are helpful in detecting and explaining such dissimilarities.

The different blood-brain barrier penetration of M3G and M6G is surprising in view of the fact that both have similar experimental octanol-water partition coefficients (at pH 7.4 the log D values are 0.08 and 0.2, respectively [30]). These experimental values also show that the two molecules are much less hydrophilic than would be expected from their polarity. (The experimental values can be compared, for example, to the calculations in this work of log $D_{7.4}$ values -3.75 for M3G and -3.35 for M6G). The discrepancy between the expected and the observed hydrophilicities can be explained by the conformational behavior of the two molecules [29, 31]. In an aqueous solution, extended conformers predominate that expose

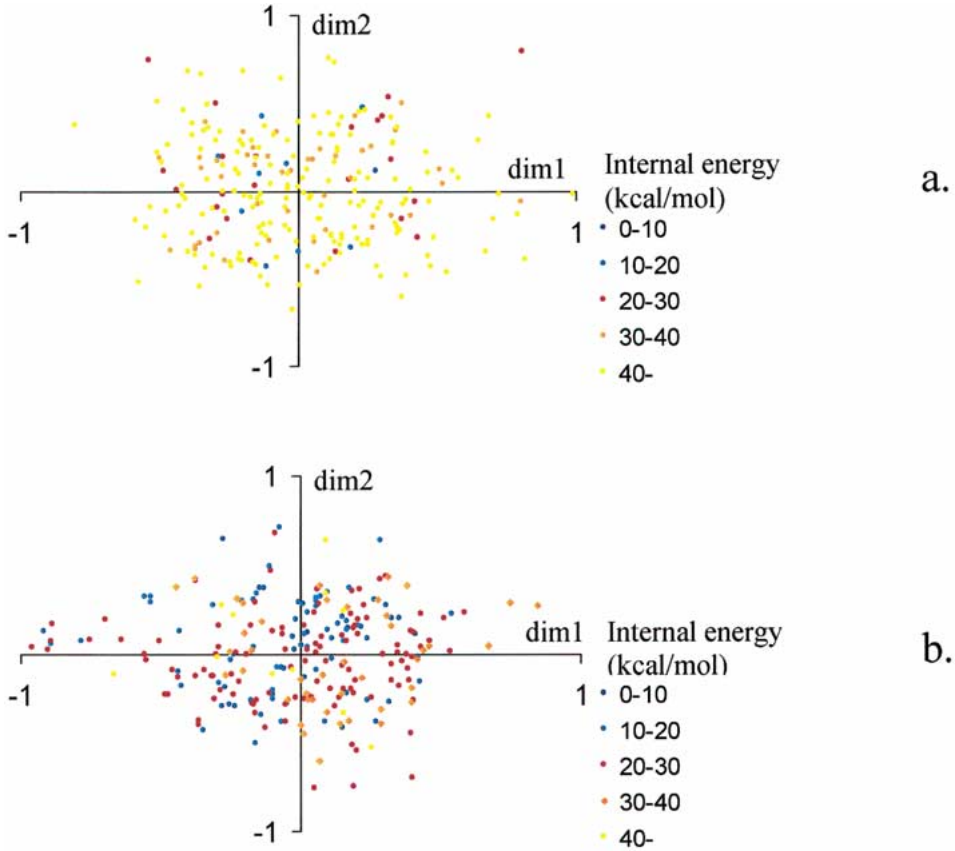

Figure 13. Multidimensional scaling of the orientational component of the distance matrix for the best 250 overlays of morphine-3-glucoronide and morphine-6-glucoronide (a) in vacuum (STRESS: 0.22) (b) in water (STRESS: 0.21). All the solutions appear together in one group because the morphine parts of the two molecules were overlaid in all of these high-scoring solutions. The coloring scheme of the solutions corresponds to their internal energies. The high internal energy of the vacuum overlays and the relatively low energy of the solutions in water indicates their different behavior on crossing from one medium to the other. The axes represent the first two dimensions of the scaled coordinates and are marked as *dim1* and *dim2*.

the polar groups, while in a non-aqueous environment the majority of conformers are folded so that they can efficiently mask the polar groups. This uncommon conformational behavior is usually referred to as 'chameleonic' [29].

A total of 1000 alignments were generated in water *in vacuo*. The best scoring solutions are displayed in Figures 12a and 12b. The major difference between the two alignments is the direction of the potential hydrogen bonding groups. Whereas the vacuum overlays contain conformations that can form internal hydrogen bonds, thus reducing the polarity of the molecules, those in the solvated alignments form hydrogen bonds with the solvent. This observation agrees with earlier conclusions [29, 31]. It can also be ascertained from the measurement of the distance between the 6-OH and the carboxylic carbon in the M3G moiety that this distance is generally greater in water. In the best scoring solution, for example, this C-O distance is 3.1 Å and 5.1 Å in vacuum and water, respectively. In summary, simulations in different media in this work produced identical conclusions to those obtained from virtual lipophilicity calculations [29, 31], confirming the chameleonic behavior of the two systems.

It has been shown so far how flexible alignments help to explain the blood-brain penetration of the polar compounds M6G and M3G. In order to understand the dissimilarity between them, the distance matrix for the 250 best scoring alignments was projected as scaled. The resulting plots for the two simulations are displayed in Figures 13a and 13b. There is no significant clustering evident in these plots. On the other hand, the first 250 solutions all appear to be within the [−1, 1] range along both coordinates, indicating that all

these alignments are similar to each other. Inspection of the individual alignments reveals that this similarity results from the overlay of the morphine units in all best scoring solutions.

Although the essential distribution of the alignments from multidimensional scaling appears to be similar in vacuum and water, major differences show up when we compre the internal energies of the solutions. This is illustrated by the coloring of Figures 13a and 13b, where the internal energy was calculated as the sum of the individual conformational energies in the alignment. In vacuum, the aligned conformers are highly energetic, the majority being over 45 kcal/mol above the lowest energy solution. In contrast, the great majority of the solutions in water are below 25 kcal/mol in energy. Although the two molecules are similar in terms of the overlay of their functional groups in water, conformational differences between the two can be expected to arise on transfer to a non-aqueous medium. In other words, the flexible alignments demonstrated that the differences on transfer between M3G and M6G are likely to be also conformational in nature.

Conclusions

A new method is described in this work for the rapid visual evaluation of the results of flexible molecular alignments. Since the method is capable of comparing two molecules by considering different orientations and conformations, it can be applied in situations where the molecules must assume a similar position, such as in the study of binding. As a common position can usually be adopted in a number of conformations (where the two molecules both change conformation in the same manner), there is value in considering their orientations separately from their conformations. Mapping common orientations essentially amounts to identifying potential common binding modes, as is demonstrated by the studied examples. It is important to emphasize that potential binding modes are not necessarily the same as experimentally observed ones. Whether two molecules actually bind similarly in a given case depends on the receptor structure and thus similarity is receptor-dependent in this context [32]. The purpose of the described method is merely to identify potential binding modes, which may be (but not necessarily are) accessible in case of a specific receptor.

Another possible application demonstrated here is identifying conformational differences between 'similar' molecules. There are many known cases of two molecules binding to the same target in the same mode and yet having quite different properties on crossing through a membrane or the blood-brain barrier. It could be demonstrated with flexible alignments in the cases shown above that the molecules were highly similar in an aqueous environment but not in vacuum. (Vacuum was used as a model for a non-aquerous medium). Although there are other methods that could be used in the specific examples described, analysis based on multidimensional scaling is a robust, rapid and automatic process. Also, the described process can be applied in conjunction with any alignment tool, although in our experience the FlexAlign method [4] proved to provide results in best agreement with experiments.

It is interesting to explore the described approach as a means for a broader understanding of similarity. In many cases, chemists understand similarity as the resemblance of molecular constitutions, as represented by the 2D or 3D structure. Molecules are dynamic entities, constantly changing their positions and conformations. Therefore, our method with its simultaneous consideration of possible orientational and conformational manifolds should provide a more realistic basis for the description of molecular similarity than single conformer methods. In addition, we can further expand our new understanding of similarity by introducing scroring schemes based on different properties. As our alignment-based similarity scheme seemed to account for the similarities and dissimilarities in binding or partitioning in all the studied cases, it is hoped that the method can be further developed into a quantitative predictor of activities and transport properties.

References

1. Feher, M. and Schmidt, J., J. Chem. Inf. Comput. Sci., 41 (2001) 346.
2. Lemmen, C. and Lengauer, T., J. Comput. Aid. Mol. Des., 14 (2000) 215.
3. Feher, M. and Schmidt, J., J. Chem. Inf. Comput. Sci., 40 (2000) 495.
4. Labute, P., Williams, C., Feher, M., Sourial, E. and Schmidt, J., J. Med. Chem., 44 (2001) 1483.
5. Krzanowski, W.J., Principles of Multivariate Analysis, Clarendon Press, Oxford, 1988.
6. Klebe, G., Mietzner, R. and Weber, F., J. Comput. Aid. Mol. Des., 8 (1994) 751.

7. Molecular Operating Environment, Version 2000.02, Chemical Computing Group Inc., Montreal, Quebec, Canada.
8. Halgren, T.A., J. Comp. Chem., 17 (1996) 490.
9. Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T., J. Am. Chem. Soc., 112 (1990) 6127.
10. Qiu, D., Shenkin, P.S., Hollinger, F.P. and Still, W.C., J. Phys. Chem. A., 101 (1997) 3005.
11. Ferguson, D.M. and Raber, D.J., J. Am. Chem. Soc., 111 (1989) 4371.
12. ACD log *D* suite, version 4.56, Advanced Chemistry Development Inc., Toronto, ON, Canada.
13. NAG Statistical Add-Ins for Excel, Release 1.1, The Numerical Algorithm Group Ltd., Oxford, UK, 1999.
14. Girvan, R. and Grant, F., Scie. Comp. World 52 (2000) 31.
15. McCullough, B.D. and Wilson, B., Comp. Stat. Dat. Anal. 31 (1999) 27.
16. Testa, B., Carrupt, P.-A., Gaillard, P. and Billois, F., Pharm. Res, 13 (1996) 335.
17. Böhm, H.-J., Klebe, G. and Kubinyi, H., Wirkstoffdesign, Spektrum Verlag, 1996, pp. 345–346.
18. Böhm, H.-J., Klebe, G. and Kubinyi, H., Wirkstoffdesign, Spektrum Verlag, 1996, pp. 327–330.
19. Badger, J., Minor, I., Kremer, M.J., Oliveira, M.A., Smith, T.J., Griffith, J.P., Guerin, D.M.A., Krishnaswamy, S., Luo, M., Rossmann, M.G., McKinlay, M.A., Diana, G.T.D., Dutko, F.J. Fancher, M., Rueckert, R.R. and Heinz, B.A., Proc. Natl. Acad. Sci. USA, 85 (1988) 3304.
20. Böhm, H.-J. and Klebe, G., Angew. Chem. Int. Ed. Engl., 35 (1996) 2588.
21. Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A. and Carlquist, M., Nature, 389 (1997) 753.
22. Wurtz, J.-M., Egner, U., Heinrich, N., Moras, D. and Mueller-Fahrnow, A., J. Med. Chem. 41 (1998) 1803.
23. Arevalo, J.H., Hassig, C.A., Stura, E.A., Sims, M.J., Taussig, M.J. and Wilson, I.A., J. Mol. Biol., 241 (1994) 663.
24. Walliman, P., Marti, T., Fürer, A. and Diederich, F., Chem. Rev., 97 (1997) 1567.
25. Zhorov, B.S. and Lin, S.X., Proteins, 38 (2000) 414.
26. Schmidt, J., Mercure, J., Feher, M., Dunn-Dufault, R., Peter, M. and Redden P., De Novo Design i Synthesis and Evaluation of Novel Non-Steroidal High Affinity Ligands for the Estrogen Receptor, to be published.
27. N'Goka, V., Schlewer, G., Linget, J.-M., Chambon, J.-P. and Wermuth, C.-G., J. Med. Chem. 34 (1991) 2547.
28. Carrupt, P.A., Gaillard, P., Billois, F., Weber, P., Testa, B., Meyer, C. and Perez, S., The Molecular Lipophilicity Potential: A New Tool for log *P* Calculations and Docking, and in Comparative Molecular Field Analysis, in Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, 1996, pp. 195–217.
29. Carrupt, P.A., Testa, B., Bechalany, A., Tayar, N.E., Descas, P. and Perrissoud, D., J. Med. Chem. 34 (1991) 1272.
30. Avdeef, A., Assessment of Distribution-pH Profiles, in Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, 1996, pp. 109–139.
31. Gaillard, P., Carrupt, P.A. and Testa, B., Bioorg. Med. Chem. Lett. 4 (1994) 737.
32. Kubinyi, H., Persp. Drug Disc. Des. 9/10/11 (1998) 225.