Consensus Molecular Alignment Based on Generalized Procrustes Analysis[||]

P. M. Kroonenberg,[†] W. J. Dunn III,*[,‡] and J. J. F. Commandeur[§]

Department of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands, Department of Medicinal Chemistry, University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612, and Institute for Road Safety (SWOV), P.O. Box 1090, 2260 BB Leidschendam, The Netherlands

Received May 9, 2003

One of the most serious problems in three-dimensional quantitative structure−activity relationship (3D-QSAR) studies is selection of an alignment rule for molecular super position of the compounds in the data set. In 3D-QSAR analyses of structure−activity data, a reference compound in a defined conformation is chosen, and all structures in the data set are aligned with the reference in a pairwise manner. In subsequent steps, conformation/alignment-dependent descriptors are computed for the compounds and compared to those of the reference. This approach gives much weight to the arbitrarily chosen reference molecule and can introduce a bias in the results. Here an alternative, and more general, approach to molecular alignment is presented that is based on Generalized Procrustes Analysis (GPA). The result is a consensus alignment that uses all molecules in the data set and avoids the bias introduced in the pairwise alignment strategy.

INTRODUCTION

In the transition from two-dimensional quantitative structure activity relationship, (2D-QSAR) studies to 3D-QSAR, it became necessary, when considering receptor alignment as a variable, to construct molecular descriptors that are conformation/alignment dependent. The method that evolved from this, comparative molecular field analysis (CoMFA),[1] initially used visual analysis of superimposed structures to develop an alignment rule. While CoMFA was a breakthrough, alignment selection in 3D-QSAR continues to be a problem. Although it has seen considerable attention, as discussed by Norinder[2] no generally acceptable method has evolved. Gerber and Muller[3] have used what appears to be principal components analysis to superimpose molecules based on several sets of atomic coordinates. A review of the alignment problem has recently been published.[4]

The approach taken here is a geometric one as the alignment is based on the relative positions of common atoms in space. Alignments can be done based on electrostatic and steric overlap.[5,6] If this type of alignment is desired, the method discussed here could be used and would be considered a preprocessing step if more than two molecules are to be superimposed.

In 3D-QSAR analyses, drug-receptor binding is assumed to result from an optimal fit of the ligand in the receptor's binding site. Fit is determined by orientation of the functional groups of the ligand in a manner that provides optimal interactions with their complementary groups in the binding site. First, it is necessary to determine the functional groups of the ligand that are necessary for binding. Structure−activity data obtained by selective replacement of functional groups and site-specific mutations or the receptor protein can be used to identify those groups.

Two critical assumptions are made. The first is the assumption that the geometry of the receptor is rigid or semirigid. The second assumption is that the ligands bind in a low-energy conformation. Thus an optimal ligand alignment is necessary. X-ray crystallographic images of ligand-active site complexes support these assumptions.[7] In most drug-design problems, however, the geometry of the receptor is unknown so the optimal alignment must be determined based on other arguments.

Within the social, behavioral, and food sciences, GPA,[8−11] named for the innkeeper in Greek mythology that shaped the torsos of his guests in various ways so as to give them an optimal fit in their beds, is often used to align configurations from different analyses to assess them with respect to each other. Procrustes analysis has been used in QSAR studies to compare different sets of properties of diverse compounds in a similarity study,[12] and it has been used in X-ray crystallography to align two molecules in a pairwise manner.[13]

For the application of GPA to the general alignment problem, each molecule is represented by a matrix in which the rows are the atoms and the columns are their coordinates in three-dimensional space so that each molecule is represented by a matrix with differing numbers of rows. Commandeur[14] developed a variant of GPA that will handle differing numbers of rows in different samples. This feature makes it possible to use the technique for aligning molecules that have some common atoms and some noncommon ones. Some or all of the common atoms may be used for the alignment by GPA. Aligning on common atoms automatically orients the geometric arrangements of the other atoms of the molecules in order that they can be assessed with respect to the other. By using publicly available software,

* Corresponding author phone: (312)996-5232; fax: (312)996-7107; e-mail: wjdiii@uic.edu.
[†] Leiden University.
[‡] University of Illinois at Chicago.
[§] Institute for Road Safety (SWOV).
[||] This paper is dedicated to Professor Corwin Hansch.

10.1021/ci0302916 CCC: $25.00
Published on Web 08/29/2003

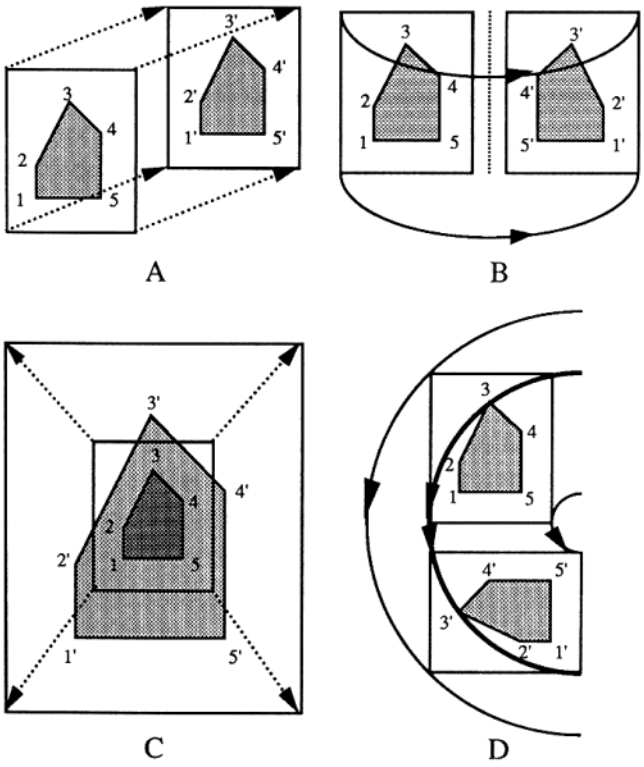


Figure 1. Admissible transformations in Generalized Procrustes Analysis.

representations of the aligned molecules can be easily viewed. Further, it should also be possible to interface the GPA technique to other software in order to carry out 3D-QSAR analyses using conformation/alignment dependent descriptors computed using GPA-derived consensus alignment rules, thus removing the bias introduced with the pairwise strategy.

GENERALIZED PROCRUSTES ANALYSIS

The general problem to be solved in GPA is to match a set of configurations as closely as possible. In this case, the configurations are a common set of atoms in a group of molecules. The atoms are described by their three-dimensional coordinates and are represented in matrix form. The standard criterion for GPA is that the sum of the squared distances between corresponding sets of points is as small as possible, under the restriction that the relative distances between the points of a configuration remain unchanged. In terms of molecules this means that the initial atom bond lengths and angles within each molecule, that is, the relationships between the atoms, are not affected. The general case requires only that the *relative* distances between the points (rows) of a configuration should not change. There are four transformations that can be used to match the configurations: *translation*, that is, the origin of the coordinate system can be shifted arbitrarily (Figure 1A); *reflection*, that is, taking a mirror image of the configuration (Figure 1B); *central dilation* or *uniform scaling*, that is, stretching or shrinking the configuration in a uniform fashion without affecting the *relative* distances (Figure 1C); and *rotation*, that is, any configuration may be arbitrarily subjected to a rigid rotation to optimize the fit (Figure 1D).

In medicinal chemistry, reflections are not allowed because the absolute stereochemistry of molecules must be maintained, and dilations are not allowed because the distances between atoms are fixed/rigid. Receptors are chiral, and thus stereoisomers can have different biological properties. Therefore, a proper selection of the atoms on which the molecules are to be aligned has to be made, such that reflections do not occur.

One of the "novel" aspects of this Procrustes analysis, from a technical point of view, is that only parts of the configurations are used for alignment, that is, the other parts of the molecules are treated as missing rows in all configurations (for a similar attempt at aligning parts of molecules, rather than complete ones see ref 15. The noncommon parts of the molecules generally have different numbers of atoms (=rows), and the alignment atoms have different numbers depending on the presence of other atoms in the molecule. To perform

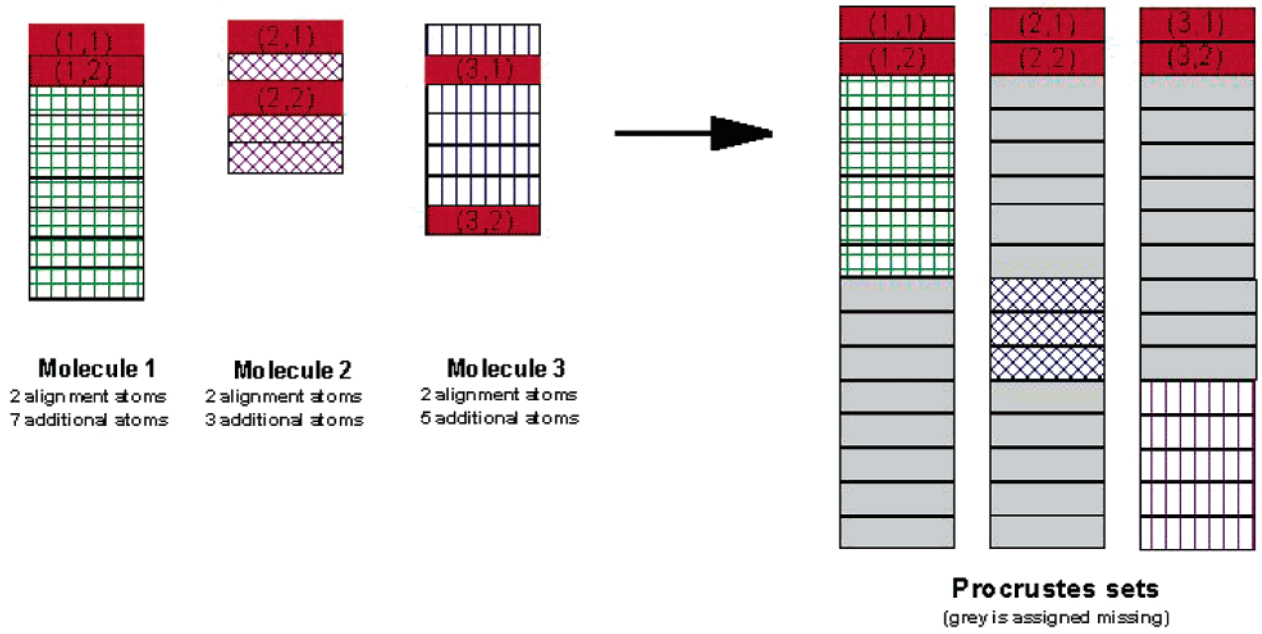


Figure 2. Creation of Procrustes coordinate sets from coordinate sets of separate molecules. In the Procrustes set on the right, the top rows are the alignment atoms, the textured rows are the atoms unique to each molecule. and the gray rows are treated as missing data.

the GPA on the selected atoms and to ensure that the other parts of the molecules are transformed as well, special measures must be taken. In particular, each configuration is expanded with the nonalignment atoms of the other molecules such that all molecules' matrices have the same number of rows. The total number of rows is therefore the sum of the number of common alignment atoms and the sum of the total number of unique atoms across all molecules (Figure 2).

In GPA an optimal centroid is determined for the alignment atoms of all molecules, and all alignment atoms of the molecules are rotated into maximal agreement with the centroid. As its overall orientation is arbitrary in a GPA, the centroid configuration of the common atoms is subjected to a principal components analysis so that the first dimensions successively account for as much variance of the centroid configuration as possible. The transformation used to optimally align the common atoms with the centroid is also applied to the noncommon atoms of the molecules. The quality of the alignment for the common part of the molecules can be computed by what has been called *analysis of variation* (see, e.g., refs 11 and 14.

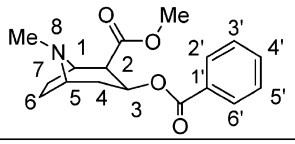
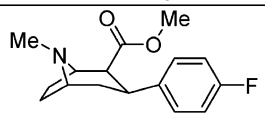
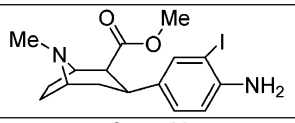
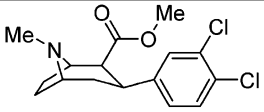
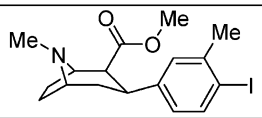
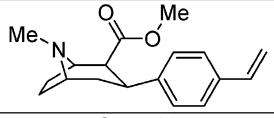
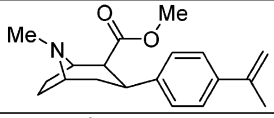
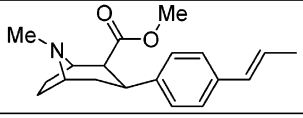
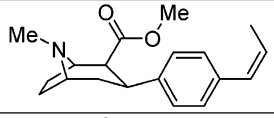
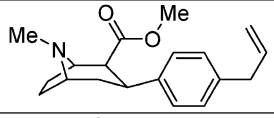
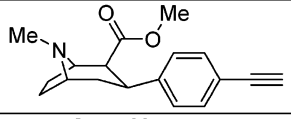
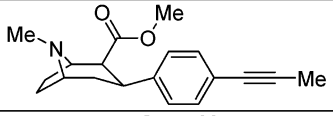
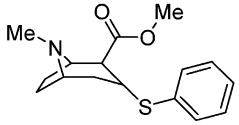
One could try to develop measures for the quality of alignment for complete molecules, but the problem is that the alignment could be perfect for the alignment atoms but inadequate, for instance due to chirality, in atoms that were not included in the alignment and there is no clear way to specify misfit in this case. Robinson et al.[15] remark on this point. "..there is no accepted metric for quantifying the quality of an alignment between two structures that do not exhibit an obvious atom correspondence. If such a metric existed, there is little doubt that it would have been employed for the purposes of molecular alignment a long time ago."

METHODS

GPA attempts to match a number of configurations in such a way that the sum of squared distances between corresponding points in the configurations is as small is possible, under the constraint that the relative distances within a configuration remain unchanged.[10] The theory of the method is presented in the Appendix. This method has been programmed in Fortran, and the GPA was run on a PC. The software reads and writes molfiles (*.mol), so that the common orientation of the molecules can be viewed with standard modeling programs (for further details about the programs used, see *http://three-mode.leidenuniv.nl)*. The figures in the paper were produced by the free WebLab Viewerlite (*http://molsim.vei.co.uk/weblab/*).

To illustrate the utility of GPA in the alignment of molecules, a set of cocaine derivatives, given in Table 1, was chosen. These, and similar compounds, were the subject of a recent tensor decomposition 3D-QSAR study.[16] The structure of and atom numbering for cocaine is given. Since this basic structure has three chiral carbons, there are eight possible stereoisomers. The compounds in the data set are all potent inhibitors of dopamine uptake by the dopamine transporter protein.[16] They are somewhat flexible, and there is evidence that there are conformational preferences in binding of cocaine and its derivatives to the protein.[17] In this study, it was realized that a generalized consensus alignment rule should be used to study the effect of conformation/alignment on binding to the transporter protein. We show here that GPA can provide these optimal consensus alignment rules.

Table 1. Structures Used in This Study

Compound	Structure
1 (cocaine)	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	

A conformational analysis was performed on the compounds in Table 1 to identify favored conformations.[16] A systematic energy search was performed in which the torsion bonds were rotated in 30° increments, and the energy-minimized structures were considered for a 10 kcal/mol

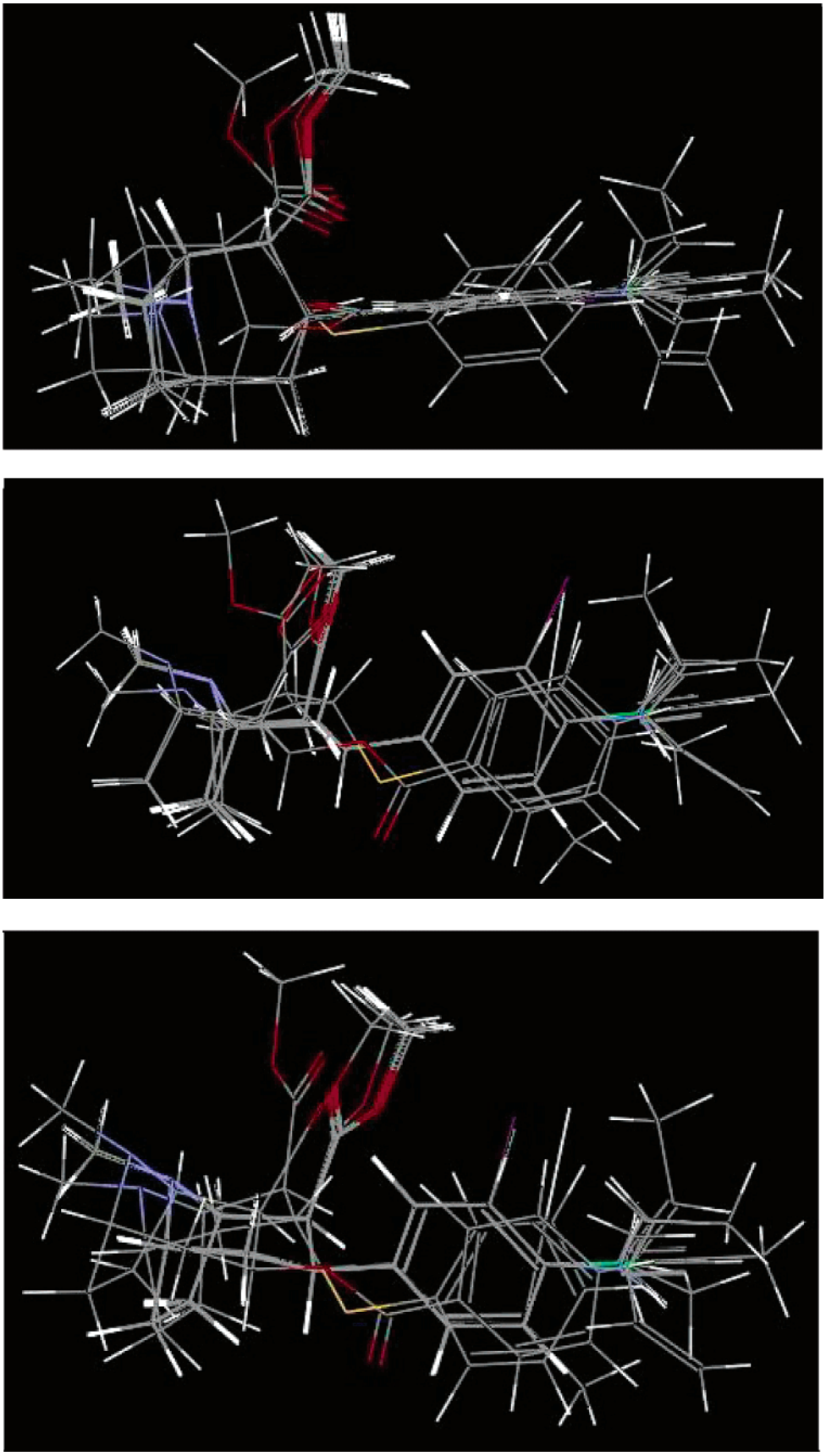


Figure 3. Three views of the results of the Procrustes analysis with Alignment 1: view 1 (top), in the plane of the phenyl ring; view 2 (middle), almost perpendicular to the plane of the phenyl ring; and view 3 (bottom), rotated to optimally show the orientation of the ester group.

threshold above the lowest identified local minima. This conformational analysis ensures that only lower energy wells are considered. Unless specified, all calculations were performed using the Universal Open Force Field.[18,19] The

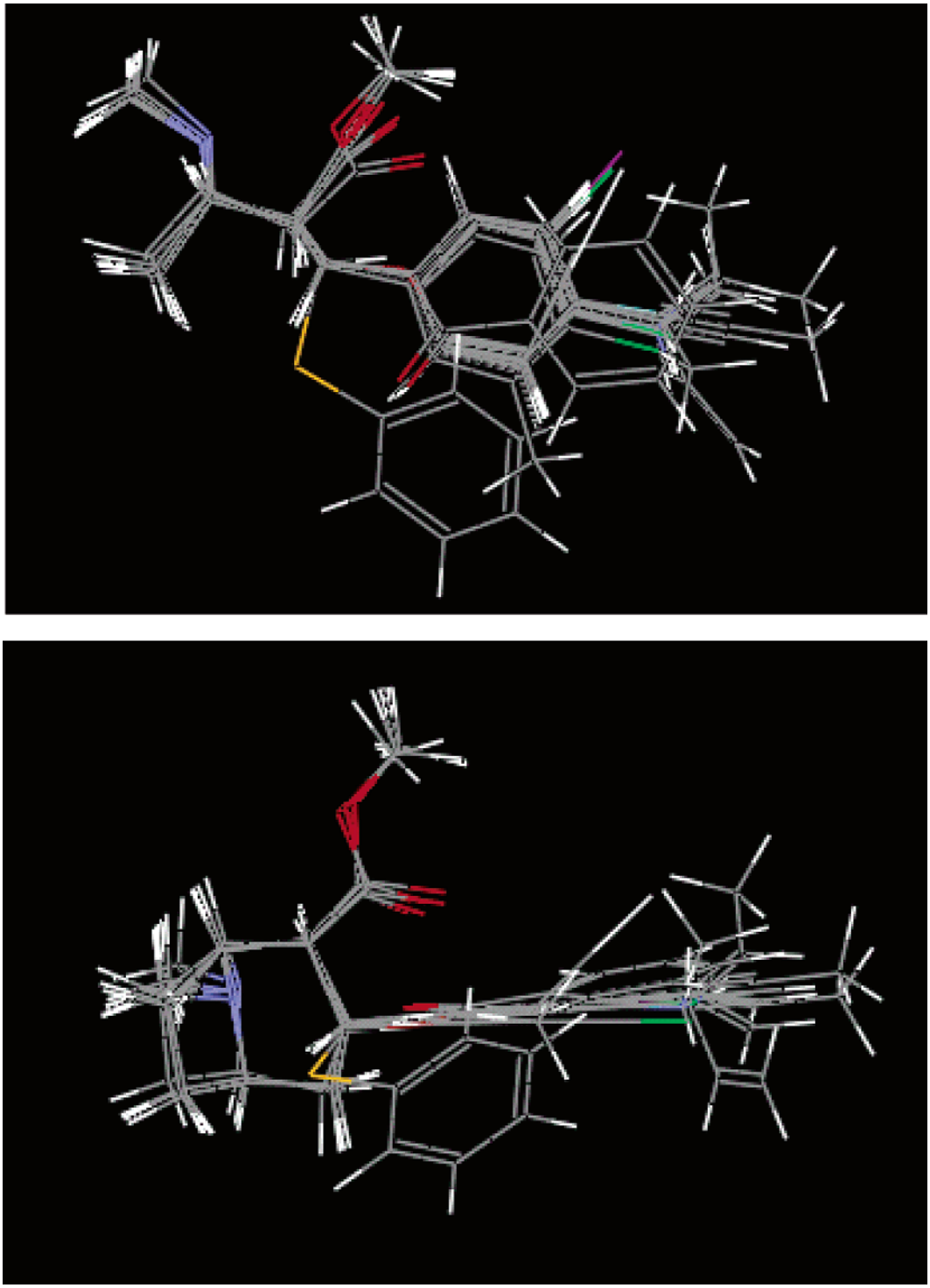


Figure 4. Two views of the result of the Procrustes analysis for Alignment 2: view 1 (top) on the left the tropane ring seen on its side and view 2 (bottom) about 90° rotated with respect to view 1.

coordinates of the low-energy conformation of each are available as Supporting Information.

ALIGNMENT RULES

To realize an alignment of the molecules in the study a number of decisions have to be made. At least three common atoms have to be chosen for the alignment rule. The first decision is whether it is desirable to align on (1) a limited number of atoms, or (2) on all atoms that are common to all molecules, or (3) on all atoms that are common to all molecules as well as on those atoms that are common to two or more of the molecules. In the present specification (option 1) the compounds are aligned on atoms that intuitively would be clustered around the centroid of all of the structures. Structure−activity studies suggest that the 8-nitrogen, 2β-ester, and 3β-aryl groups are involved in receptor binding. In the first alignment the 8-nitrogen, the carbonyl oxygen in the 2β-position, and the 4′-carbon of the aryl group were selected for superposition. This presumably would target the alignment on the approximate center of the binding site.

A second alignment based on three atoms of the tropane part of the molecules was chosen. In particular, we aligned on atoms in the tropane ring—these would be the nitrogen, and any two or more of the tropane ring carbons. The numbering of the phenyl ring goes either clockwise or counterclockwise from the carbon attached to the tropane ring. This results from the ordering rules for substituents placed on the phenyl ring and the fact that in some compounds, due to conformational preferences, the 2- and 6-positions can become switched. Since the chiral centers are in the tropane ring system, their numbering then is fixed and unambiguous.

RESULTS AND DISCUSSION

Alignment 1. The results from Alignment 1 are given in Figure 3. In the top portion of Figure 3 the view is from above the plane of the tropane ring and in the plane of the phenyl ring. It shows that most of the compounds occupy similar positions in the space of the tropane ring and the phenyl ring. There is considerable structural variation around the 4′-position of the phenyl group. The thiophenyl substituent in compound 13 does not coincide with the other phenyl substituents, and this is as expected since it is in the 3α-position of the tropane ring.

The second part of Figure 3 shows again the close correspondence of the tropane rings in the alignment. There is also the correspondence of the phenyl groups with the exception of the phenyl group in cocaine. The ester linkage extends the phenyl group further from the tropane group than in the other compounds. The 3α-orientation of the thiophenyl group also is obvious as is variation in structure at the 3′- and 4′-positions. The third view in Figure 3 illustrates the variable positions of the 2β-ester group and illustrates, again, the structural variation in the 4′-position.

Alignment 2. The Procrustes alignment based on Alignment 2 is shown in Figure 4. This alignment is based on the tropane ring that is common to all compounds in the set. It is clearly seen that the phenyl rings of cocaine and the 3α-thiophenyl derivative are sitting apart from the rest. This is not surprising and emphasizes the idea that alignment highlights atoms remote from those used in the alignment rule, that is, those removed from the centroid.

A problem encountered when applying GPA to align compounds was preserving the original bond lengths and angles of the aligned structures. This did not turn out to be serious and could be easily checked by comparing distance matrices before and after alignment. After a GPA has been carried out it is possible, via an analysis of variation, to get an idea to what extent the different dimensions contribute to the alignment part of configurations of each of the molecules. If, as is the case in the present example, some of the nonalignment atoms are part of an identical structure in all or most molecules such as the ester groups connected to the tropane one could calculate their relative distances to assess the similarity of the molecules notwithstanding the fact that these structures were not part of the alignment rule.

CONCLUSIONS

Using GPA a simultaneous geometric alignment of molecules was performed. This clearly shows that a consensus alignment with GPA can allow all molecules in a data set to be simultaneously compared in a quantitative way. The benefit of using the present procedure over the conventional pairwise alignment is that it eliminates the bias introduced by comparison to a reference and gives each molecule in the set equal weight in the alignment. As we have presented alignment by GPA, it can be considered a preprocessing step in molecular modeling. By varying conformation and selecting various atoms for alignment, conformation/alignment dependent descriptors can be computed that are not based on the pairwise procedure. These can be submitted to 3D-QSAR software by interfacing the GPA procedure to modeling software. Further, the computed common volume element of the aligned data set should be a close approximation of the shape of the ligand binding site. A display of the aligned compounds can be easily viewed with public domain software as discussed earlier.

APPENDIX

Technical Description of Generalized Procrustes Analysis (GPA). Generalized Procrustes Analysis attempts to match K configurations in such a way that the sum of squared distances between corresponding points in the configurations is as small is possible, under the constraint that the relative distances within a configuration remain unchanged. Gower[10] showed that for I points in a J dimensional space, the sum of the squared distances between these I points is always equal to I times the sum of squared distances between the I points and their centroid. This observation is the basis for the presentation here, and the development will be based exclusively on the latter situation.

We start with the assumption that there are K configurations, $\mathbf{X}_k$ with each J columns, but with I_k rows. To be able to perform the analyses, all configurations will be expanded such that they have an equal number of rows, say I. The arrangement of the rows should be that the entities in the rows that are the same across configurations should be in the same row, that is, have the same row number. The centroid to be determined will be represented by the $I \times J$ matrix $\mathbf{Z}$.

Let us define the $I \times I$ diagonal matrix $\mathbf{M}_k$, such that the diagonal entry $m_{ii} = 1$ if the row element is present in the configuration and $m_{ii} = 0$ if the row element is missing. Using the matrices, the general loss function can be defined as

$$\sum_{k=1}^{K} tr\,(\mathbf{X}_k - \mathbf{Z})'\mathbf{M}_k(\mathbf{X}_k - \mathbf{Z}) \text{ with}$$

$$\mathbf{Z} = \left(\sum_{k=1}^{K} \mathbf{M}_k\right)^{-1}\left(\sum_{k=1}^{K} \mathbf{M}_k\mathbf{X}_k\right)$$

where tr is the trace function.

To introduce the relative distance-preserving transformations, that is, translations and the orthonormal transformations (which include rotation and reflection), and uniform rescaling, we define the $I \times 1$ unit vector $\mathbf{1}_I$, the $J \times 1$ translation vectors $\mathbf{u}_k$, the $J \times J$ orthonormal transformation matrix $\mathbf{R}_k$, and K uniform scaling factors s_k. With these quantities the fundamental loss function for GPA becomes

$$f(\mathbf{U},\mathbf{s},\mathbf{R}) = \sum_{k=1}^{K} tr(s_k(\mathbf{X}_k - \mathbf{1}_I\mathbf{u}_k')\mathbf{R}_k - \mathbf{Z})'\mathbf{M}_k(s_k(\mathbf{X}_k - \mathbf{1}_I\mathbf{u}_k')\mathbf{R}_k - \mathbf{Z})$$

where

$$\mathbf{Z} = \left(\sum_{k=1}^{K} \mathbf{M}_k\right)^{-1}\left(\sum_{k=1}^{K} s_k\mathbf{M}_k(\mathbf{X}_k - \mathbf{1}_I\mathbf{u}_k')\mathbf{R}_k\right)$$

The definition of the centroid **Z** is such that given the other parameters it yields the global minimum so that we may plough it back into the equation and solve for the other parameters independently. Furthermore, explicit formulas expressed in the other parameters exist for the translation vectors $\mathbf{u}_k$ that also can be used to simplify the loss function. After the other parameters have been determined the centroid follows from its definition, and the translation vectors from their explicit solution. Estimating the rotation and scaling parameters can only be done iteratively, and the algorithm used in the paper due to ref 13 is sketched below but will not be given here in all its details. The algorithm proceeds iteratively with solving least squares problems for each of the sets of parameters in turn until convergence. Reference 13 also gives formulas for the analysis of variation to assess how well each row element is fitted by the model and by each of the dimensions and similarly for each of the columns of the configurations and for each of the configurations themselves. This makes a detailed assessment of the fit possible. After the solutions have been determined, often the centroid is rotated to its principal axes, and the individual configurations are then rotated with the same transformation matrix.

Outline Algorithm. *1. Initialization Step*. Compute the $I\times I$ centering matrix $\mathbf{C}_k$ for each configuration (molecule):

$$\mathbf{C}_k = \mathbf{M}_k\left(\mathbf{I} - \frac{\mathbf{1}_I\mathbf{1}_I'\mathbf{M}_k}{\mathbf{1}_I'\mathbf{M}_k\mathbf{1}_I}\right)$$

Determine the Moore-Penrose inverse of the sum of the centering operators $\mathbf{C}_k$:

$$\mathbf{C}^- = \left(\sum_{k=1}^{K} \mathbf{C}_k\right)^-$$

Normalize the K configurations such that $\sum_{k-1}^{K} tr\mathbf{X}_k'\mathbf{C}_k\mathbf{X}_k = K$ (optional)

Construct the diagonal matrix $\mathbf{W}^{-1/2}$, with $w_{kk} = 1/\sqrt{\mathbf{X}_k'\mathbf{C}_k\mathbf{X}_k}$

Initialize all K scaling constants $s_k = 1$, and set all transformation matrices $\mathbf{R}_k$ equal to $\mathbf{I}_J$.

2. Orthonormal Transformation Step. Determine for each k a new orthonormal matrix $\mathbf{R}_k$ using the singular value decomposition $\mathbf{X}_k'\mathbf{C}_k\mathbf{C}^- (\sum_{k'\neq k} s_{k'}\mathbf{C}_{k'}\mathbf{X}_{k'}\mathbf{R}_{k'}) = \mathbf{P}_k\Phi_k\mathbf{Q}_k'$

$$\mathbf{R}_k = \mathbf{P}_k\mathbf{Q}_k'$$

Evaluate loss function

3. Rescaling Step. Compute the $K\times K$ diagonal matrix **W** with as diagonal elements $w_{kk} = tr\mathbf{A}_k'\mathbf{A}_k$ with $\mathbf{A}_k = \mathbf{C}_k\mathbf{X}_k\mathbf{R}_k$ and the $K\times K$ matrix **Y** with elements $y_{kk'} = tr\mathbf{A}_k'\mathbf{C}^-\mathbf{A}_{k'}$

Determine the first eigenvector $\mathbf{p}_1$ of $\mathbf{W}^{-1/2}\mathbf{Y}\mathbf{W}^{-1/2}$

The vector of new scaling constants **s** becomes $s = \sqrt{K}\,\mathbf{W}^{-1/2}\mathbf{p}_1$.

Evaluate loss function

4. Check Convergence. If the difference between consecutive values of the loss function is small enough Stop else go to step 2.

5. Compute Centroid

$$\mathbf{Z} = \mathbf{C}^-\left(\sum_{k=1}^{K} s_k\mathbf{C}_k\mathbf{X}_k\mathbf{R}_k\right)$$

Transform **Z** into its principal axes position: $\mathbf{Z}^* = \mathbf{Z}\mathbf{K}$, with **K** from $\mathbf{Z}'\mathbf{C}\mathbf{Z} = \mathbf{K}\Lambda\mathbf{K}'$

6. Compute the Optimally Transformed Configurations

$$\mathbf{X}_k^* = s_k\mathbf{M}_k(\mathbf{X}_k - \mathbf{1}_I\mathbf{u}_k')\mathbf{R}_k\mathbf{K}$$

This algorithm is assured to converge, but no guarantee can be given that it will always converge to a global optimum; however, so far in the literature no multiple solutions have been reported. In case of doubt, one could start the algorithm with other, orthonormal, matrices $\mathbf{R}_k$ in an attempt to generate other optimal solutions.

Further Technical Development. As can be seen from Figure 2, even with a few molecules which have many nonalignment atoms, the number of rows, I, of the matrices $\mathbf{X}_k$ can become very large (I = number of common atoms + total number of nonalignment atoms of all molecules together), for example, for Alignment 2 $I = 6916$, while $\Sigma_k I_k = 568$. It is therefore worthwhile to seek alternative algorithms geared specifically to the problem at hand. One option that is investigated is to perform a GPA exclusively on the common atoms, that is, steps 1, 2, 4, and 5. After convergence in step 6 the translation vectors $\mathbf{u}_k$ are computed and used to center the complete molecules. The thus centered molecules are then rotated with the previous derived rotation matrices $\mathbf{R}_k$. Optionally, the principal axis rotation matrix **K** can be applied as well. If the alignment atoms are common to all molecules, the procedure for missing rows can be circumvented, saving on the otherwise costly inversion of the $I\times I$ matrix $\mathbf{C}^-$. If there are some alignment atoms common to some but not all molecules, the dimensions of $\mathbf{C}^-$ will be relatively small. Finally in such an algorithm, the total amount of memory required for storage of the data can be reduced drastically, because only the $I_k\times J$ molecules themselves have to be stored but not their augmented versions of size $I\times J$, which in Alignment 2 would give a savings of $20\,748 - 1704 = 19\,044$ data elements. A further improvement could be to develop a new algorithm which only uses rotations and prohibits reflections, but this has yet to be investigated.

Supporting Information Available: Coordinates of the low-energy conformation of compounds in Table 1. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA): 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(2) Norinder, U. *Recent Progress in CoMFA Methodology and Related Developments*; Kubinyi, H., Folkers, G., Martin, Y. C., Ed.; Kluwer: Dordrecht, The Netherlands, 1998; Vol. 3, pp 41−56.
(3) Gerber, P. R. M.; Muller, K. Superimposing several sets of atomic coordinates. *Acta Crystallogr.* **1987**, *A43*, 426−428.
(4) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput-Aided Mol. Des.* **2000**, *14*, 215−232.
(5) Kearsley, S. An alternaitve method for the alignment of molecular structures: maximizing electrostatic and steric overlap. *Tetrahedron Comput. Method.* **1990**, *3*, 615−663.
(6) Melani, F.; Gratteri, P.; Adamo, M.; Bonaccini, C. Field interaction and geometrical overlap: A new simplex and experimental design based computational procedure for superposing small ligand molecules. *J. Med. Chem.* **2003**, *46*, 1359−1371.
(7) Kim, K. H. *Building a Bridge between G-Protein-Couled Receptor Modelling, Protein Crystallography and 3D QSAR Studies for Ligand Design*; Kluwer: Dordrecht, The Netherlands, 1998; Vol. 3.
(8) Dijksterhuis, G. B. G.; Gower, J. C. The interpretation of generalized Procrustes analysis and allied methods. *Food Qua. Prefer.* **1992**−**3**, *3*, 67−87.
(9) Dijksterhuis, G. B. *Multivariate Data Analysis in Sensory and consumer Science*; Leiden University: Leiden, The Netherlands, 1995.
(10) Gower, J. C. Generalized Procrustes analysis. *Psychometrica* **1975**, *40*, 33−51.
(11) Gower, J. C. *Orthogonal and Projection Procrustes Analysis*; Oxford University Press: Oxford, 1995; pp 113−134.
(12) Rose, F. S.; Rahr, E.; Hudson, B. D. The use of Procrustes analysis to compare different property sets for the characterization of a diverse set of compounds. *QSAR* **1994**, *13*, 152−158.
(13) Robert, D.; Carbo-Dorca, R. AnBlisi de Procrustes i alineament molecular. *Scientia Gerundensis* **1999**, *24*, 175−181.
(14) Commandeur, J. J. F. *Matching Configurations*; DSWO: Leiden, The Netherlands, 1991.
(15) Robinson, D. D.; Lyne, P. D.; Richards, W. G. Partial molecular alignment via local structure analysis. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 503−512.
(16) Appell, A.; Dunn III, W. J.; Reith, M. E. A.; Miller, L.; Flippen-Anderson, J. L. An analysis of the binding of cocaine analogues to the monoamine transporters using tensor decomposition 3D-QSAR. *Bioorg. Med. Chem.* **2002**, *10*, 1197−1206.
(17) Appell, M.; Krunic, A.; Choi, T. J.; Mariappan, S.; Dunn III, W. J.; Reith, M. E. A. NMR study of conformational preferences of inhibitors of monoamine uptake. *QSAR* **2002**, *21*, 38−40.
(18) Casewit, C. J.; Colwell, K. S.; Rappé, A. K. Application of a universal force field to organic molecules. *J. Am. Chem. Soc.* **1992**, *114*, 10035−10046.
(19) Rappé, A. K.; Goddard W. A. Charge equilibrium for molecular dynamic simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.

CI0302916