Articles

Comparative Molecular Active Site Analysis (CoMASA). 1. An Approach to Rapid Evaluation of 3D QSAR

Takayuki Kotani* and Kunihiko Higashiura

Institute of Bio-Active Science, Nippon Zoki Pharmaceutical Company Ltd., Kinashi, Yashiro-cho, Kato-gun, Hyogo 673-1461, Japan

Received July 30, 2003

We have developed a rapid evaluation method, comparative molecular active site analysis (CoMASA), for obtaining 3D QSAR. CoMASA has three major advantages: (1) the CoMASA results would easily transform to pharmacophore and/or queries required for 3D database searches, (2) the CoMASA method is not required to consider orientation and translation of molecules against a lattice, and (3) standard PCs can be used to perform the analysis. The potential of these improvements and possible further enhancements are discussed.

Introduction

We have developed a rapid evaluation method for obtaining three-dimensional quantitative structure−activity relationships (3D QSAR) using extraction of representational coordinates for the molecules instead of the lattice points around the molecules that are used in comparative molecular field analysis (CoMFA), comparative molecular similarity analysis (CoMSIA), and other 3D QSAR methods. In 3D QSAR analyses such as CoMFA and CoMSIA, molecular orientation against the lattice points sometimes affects 3D QSAR accuracy. In addition, these methods require not only high-performance calculation power but also huge memory space. To solve such issues, we have developed comparative molecular active site analysis (CoMASA). CoMASA has three major advantages: (1) the resulting maps are easily understandable so that the CoMASA results would easily transform to pharmacophore and/or queries required for 3D database searches, (2) the CoMASA method is not required to consider orientation and translation of molecules against a lattice, and (3) because of its rapid scoring functions and reducing interaction points, standard PCs can be used to perform the CoMASA analysis. The potential of these improvements and possible further enhancements are discussed.

Since its advent, comparative molecular field analysis (CoMFA)[1] has become one of the most powerful tools for three-dimensional quantitative structure−activity relationship (3D QSAR) studies. For over a decade, this approach has been widely used to clarify the mechanism of interaction between various receptors and ligands. In CoMFA, molecules are represented and compared by their steric and electrostatic fields sampled at the intersections of a lattice (or grids, or boxes) spanning a three-dimensional region. Thus, each CoMFA descriptor column contains the magnitude of either the steric or electrostatic field exerted by the atoms in the tabulated molecules on a probe atom located at a point in Cartesian space. Because this table usually has many more CoMFA steric and electrostatic descriptors than compounds, standard multiple regression is practically impossible. Instead, partial least squares (PLS) analysis is applied to derive the final CoMFA model. A cross-validated r^2 (q^2) usually serves as a quantitative measure of the statistical quality of the models.

In CoMFA, molecular fields are used as descriptors in regression models. Lennard−Jones potentials and electrostatic potentials against simple probe ions are used as steric fields and electrostatic fields, respectively. Other workers have extended the methods to use hydrophobic fields (e.g., HINT),[2] Gaussian pseudopotential (comparative molecular similarity analysis, CoMSIA),[3−5] a water probe (GRID),[6−8] intersection volume field (INVOL),[9] molecular lipophilic potential (MLP),[10−12] etc.[13−16] CoMSIA uses a Gaussian-type distance dependence, and no singularities occur at the atomic positions. The strongly increasing steric potentials beyond van der Waals (vdW) contact interatomic distances produce certain difficulties in the CoMFA method. The acute steepness of the function increases the sensitivity of the statistic results to the precise docking of the aligned set of molecules into the 3D grid as well as to the accuracy of the alignment rule formation. The INVOL varies smoothly with interatomic distance to conquer the above problem. Accordingly, no arbitrary definitions of cutoff limits and deficiencies due to different slopes of the fields are uncounted. The alignment function used in SEAL[17,18] overcame the outline problems. The alignment condition in SEAL is based on mutual similarity indices pairwise calculation between all atoms of the molecules that are compared. The 3D QSAR methods, which use a lattice, have artifacts that may be introduced using difference grid

* To whom correspondence should be addressed: E-mail t-kotani@nippon-zoki.co.jp; Fax +81-795-42-5332.

10.1021/jm030364c CCC: $27.50
Published on Web 04/16/2004

spacings. Orientation and translation of molecules against a lattice also influence the 3D QSAR results.[19,20] In addition, they are a time- and memory-consuming process and require a high-performance computer.

Hypothetical active site analysis (HASL)[21−23] is another 3D QSAR procedure which is accomplished through the intermediate conversion of a superposed set of molecules to a set of regularly spaced points (lattice) defined by Cartesian coordinates (*x*,*y*,*z*) and atom types. The resulting lattice is then related to individual activity by endowing each lattice intersection point with a binding value so that the sum of the points belonging to a particular molecule results in the original value for that molecule. The binding activity of a new molecule is predicted through the summation of partial values associated with lattice points corresponding to the new molecule. HASL is especially different from CoMFA and the other 3D QSAR methods as it handles relationships in a 3D molecular space, whereas CoMFA is field analysis.

The minimum topological difference (MTD)[24−27] method and the molecular field topology analysis (MFTA)[28] method are the other approaches for much less CPU-time-consuming QSAR analyses. The hypermolecule, which is used in the MTD method, is the result of the approximate (non-hydrogen) atom-per-atom superposition of the molecules in the investigated series. The vertexes of the hypermolecule correspond to the positions of these atoms. Vertexes are binned in three categories, beneficial, detrimental, or irrelevant, and are mapped into an occupancy/nonoccupancy type of interaction.

Alternatively, molecular similarity calculations are now being widely applied in QSAR of drugs and agrochemicals. Although 3D QSAR analyses such as CoMFA and CoMSIA have been widely used for drug design with an increase in efficacy, superposition of the molecules is one of the persistent obstacles to obtain reproducibility and precise 3D QSAR models. We previously reported a rapid similarity calculation method for standard PCs and its application to the superposition of molecules.[29] This method for calculations of similarity indices remarkably reduced the hitherto required time (especially 2 or 3 orders of magnitude faster than previous grid-based evaluation techniques) without any clear loss of preciseness.

In contrast to the superposition methods available on standard PCs, almost all 3D QSAR methods are still time-consuming processes and require a high-performance computer with a huge memory area. We describe a rapid evaluation method for obtaining 3D QSAR, CoMASA, using extraction of the representational coordinates for the molecules instead of the lattice points around the molecules that are used in CoMFA, CoMSIA, and the other 3D QSAR methods.

Methods

All calculations were carried out on a standard Pentium III 600 MHz PC with 256 MB memory. All programs were written in Fortran and C languages and Tcl/Tk. Visualization of the results of CoMASA was performed by Weblab ViewerLite 4.0.[30]

Extraction of Representational Coordinates for the Molecules by Cluster Analysis. Minimum collection of points which are spatially occupied by the molecules would be essentially sufficient to elicit a biological response. As replacement of one functional group to the other one is often carried out to obtain expected biological activity, synthetic chemists sometimes need information like pharmacophore perception. We described the possibility of cluster analysis to extract the minimum collection of points to avoid lattice-based 3D QSAR problems.

The CoMASA method solves arbitrariness with the extraction of minimum collection points by cluster analysis. Cluster analysis is a multivariate analysis technique that seeks to organize information about variables so that relatively homogeneous groups or "clusters" can be formed. We applied the joining (tree clustering) method to the reduction of atomic coordinates according to the procedure shown in Chart 1.

Extraction of the molecular represented points using cluster analysis provided another advantage. Selection of the template molecule is also varied by the researcher's sense, and it would become an artifact for QSAR results. If the molecular represented points are defined by comparison of the molecules onto the template, the points would give different results on how to choose the template. The cluster analysis is suited for avoiding the above issue and reproducibility of 3D QSAR results.

Though the threshold for the cluster analysis is arbitrary, we used 0.75 Å as the default threshold. Addition of ring centroids for the extraction of atomic represented points by cluster analysis is also attempted.

CoMASA Methodology. Similar to the usual CoMFA approach, a data table was constructed by calculating the interaction at the molecular represented points. Interactions between the compounds at the molecular represented points were calculated according to the following evaluation functions. Details of each method are summarized in the appendix.

Study 1. Steric and Electrostatic Descriptors and Evaluation Functions on the CoMASA Method. Four studies were carried out on the CoMASA method using 21 steroids that have been described in the literature[1] to clarify the influence of steric and electrostatic descriptors. Atom-by-atom alignment reported in the CoMFA and CoMSIA methods were used for the calculation of the binding affinity to human corticosteroid-binding globulins (CBG).

Study 1-1. Rapid Similarity Index Calculation Method as an Evaluation Function (Method A). We previously reported rapid similarity indices,[29] considering only the nearest atomic distances of each template atom. This technique is somewhat coarse and rough but extremely rapid. The strategy for calculation of similarity indices has been applied to an evaluation function of the CoMASA method.

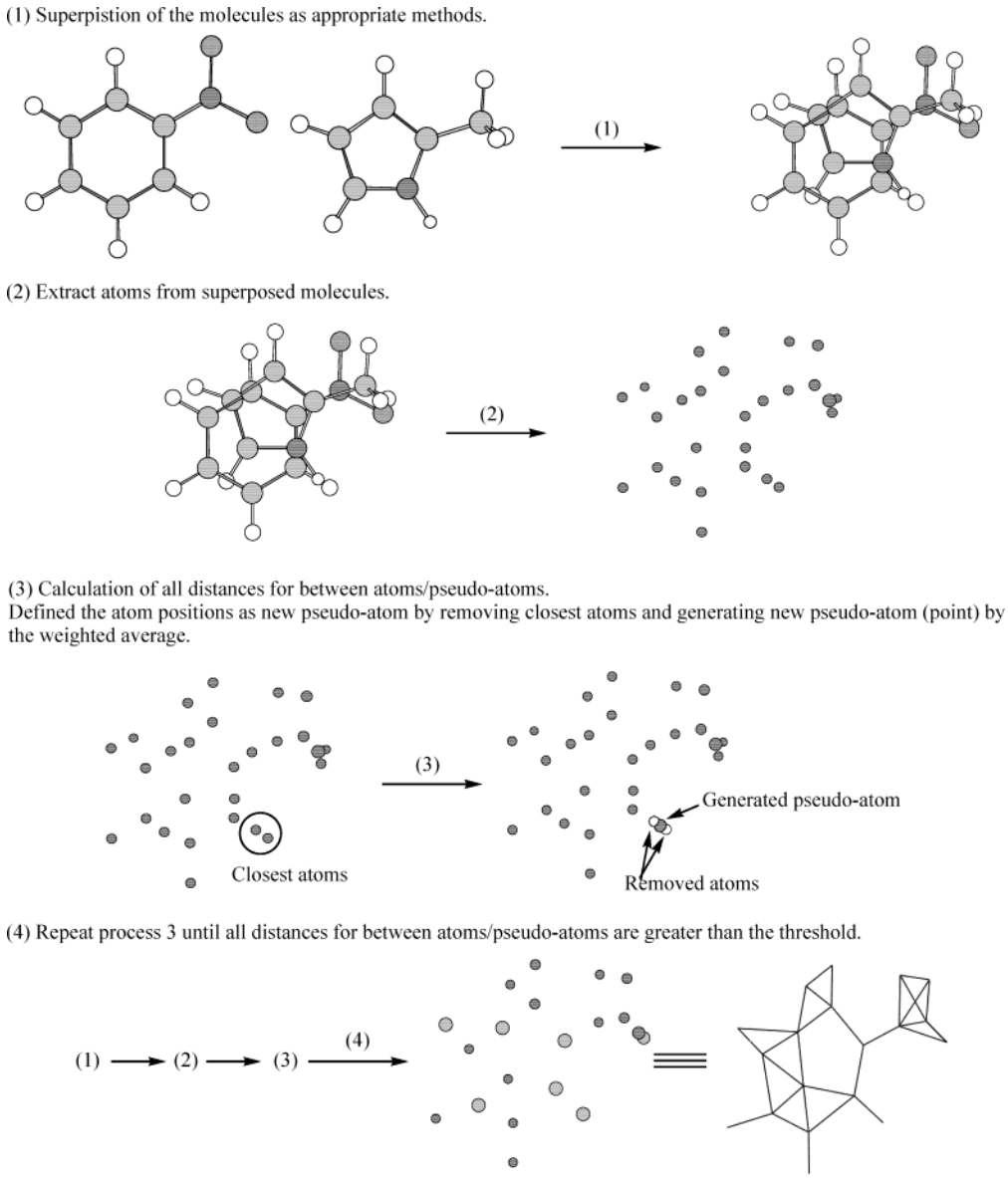
Study 1-2. SEAL as an Evaluation Function (Method B). Similar to the CoMSIA approach, SEAL similarity indices were used as an evaluation function.[3−5] Similarity indices $A_{F,k}$ between compounds of interest and a probe atom were calculated according to (i.e., at the molecular represented point q for the molecule j of the data set):

$$A_{F,k}{}^{q}(j) = -\sum_{i=1}^{n} (w_{\mathrm{probe},k} w_{i,k} e^{-\alpha r_{iq}{}^{2}})$$

where i = summation index over all atoms of the molecule j under investigation, $w_{i,k}$ = actual value of the physicochemical property k of the atom i, $w_{\mathrm{probe},k}$ = probe atom with charge +1 and radius 1 Å, α = attenuation factor, and r_{iq} = mutual distance between the probe atom at the molecular represented point q and the atom i of the test molecule. Details are summarized in the appendix.

Study 1-3. Good's Gaussian Similarity Indices as an Evaluation Function (Method C). Good's similarity indices[31,32] were applied to the CoMASA approach as well as SEAL. Good et al. reported the shape and electrostatic similarity indices using the two and three Gaussian function approximations to the atomic orbital electron density and the electrostatic potential, respectively. Electrostatic and shape similarity indices P_{el} and P_{st} between compounds of interest and a probe

Chart 1. The Processes of Extraction of Representational Coordinates for the Molecules by the Cluster Analysis Using Nitrobenzene and 2-Methylpyrrole as Sample Molecules. After the Compete Division (process 4), the Representational Coordinates Are Developed



atom were calculated according to (i.e., at the molecular represented point k for the molecule j of the data set):

$$P^k(j) = P_{el}{}^k(j) + P_{st}{}^k(j)$$

$$P_{el}{}^k(j) = -\sum_{i=1}^{n} q_i(G_{el1} + G_{el2})$$

$$P_{st}{}^k(j) = -\sum_{i=1}^{n} (G_{st1} + G_{st2} + G_{st3})$$

where $G_Z = \gamma_Z e^{-\alpha_Z r_{ik}{}^2}$ and i = summation index over all the atoms of the molecule j under investigation, q_i = actual charge of the atom i; γ_Z = proportionality constants and α_Z = exponential constants for Gaussian function approximation defined by the atom type, and r_{ik} = mutual distance between the probe atom at the molecular represented point k and the atom i of the test molecule.

Study 1−4. Newly Introduced Simple Indicator Variables as an Evaluation Function (Method D). Newly introduced simple indicator variables were adopted to clarify the scope and limitations of the CoMASA methodology. Simple indicator variables are a variant of the topological index and are expected to reduce calculation time without clear loss of preciseness. The MTD method, which is based on a very similar concept, has the three category bins, beneficial, detrimental, or irrelevant, according to occupancy/nonoccupancy type of interaction. Our newly introduced indicator variables are three simple categories depending on interatomic distances. Although the evaluation function of method A has been applied for steric interaction, other effects such as electrostatic and hydrophobic interaction were not included. Threshold 0.75 Å, which has been used for cluster analyses threshold, is applied as the default value. Indicator variables I between compounds of interest and a probe atom systematically replaced at the molecular represented points have been calculated according to (i.e., at the molecular represented point k for the molecule j of the data set):

$$I^k(j) = I_{el}{}^k(j) + I_{st}{}^k(j)$$

where the indicator valuable I^k at the molecular represented point k for the molecule j of the data set is simply defined by the distance r_k from the point k to the nearest atom i of the

molecule j. Electrostatic and steric indicator variables I_{el} and I_{st} are derived form the distance r_k:

$$I_{el,k} = \begin{cases} q(k) & \text{if } r_k \le \text{th} \\ \dfrac{q(k)}{2} & \text{if th} < r_k \le 2 \times \text{th} \end{cases}$$

$$I_{st,k} = \begin{cases} 1.0 & \text{if } r_k \le \text{th} \\ 0.5 & \text{if th} < r_k \le 2 \times \text{th} \end{cases}$$

th = threshold for cluster analysis, q = charge at the atom of the molecule j closest to the molecular represented point k. This method also has the great advantage of reducing calculation time because if the atomic distance between the atom j and the molecular represented point k is less than th, then further atomic distances based on the molecular represented point need not be calculated. As in method A, all of the distances r_k need not be calculated, and it helps to save calculation times.

Study 2. Hydrophobic Descriptors on the CoMASA Method. HINT,[2] GRID,[6−8] MLP,[10−12] etc.,[9,13−16] were used for hydrophobic contribution calculations of the ordinary 3D QSAR methods. Here we describe the two parametrization methods, AlogP[33−36] and hydrophilic parameters used in the FlexS flexible ligand superposition method.[37] These two types of hydrophobic descriptors were carried out in combination with SEAL and the simple indicator variables as evaluation functions on the CoMASA method (see methods B and D).

Study 2-1. ALogP Hydrophobic Parameters with SEAL Type Function (Method E). Similar to the CoMSIA approach, AlogP[33−35] hydrophobic parameter was applied to the CoMASA method using the Gaussian function. The AlogP method is, in constant, a direct, easy-to-computerize atomic constant approach to predict log P and is shown to exhibit a relatively robust performance, though the method has important limitations.[35,36] An interaction involves two or more atoms; hence, setting a unique atom type for all involved atoms in a particular interaction will lead to linearly dependent atom type columns, thereby intuitively equating similar atom environments into one class. Similarity indices $A_{F,k}$ between compounds of interest and a probe atom were calculated as described in Study 1−2, where w_{ik} = actual value of the physicochemical property k of atom i, i.e., AlogP value of atom i, and $w_{probe,k}$ = probe atom with hydrophobicity = +1. The attenuation factor α is used the same value as that of Study 1-2.

Study 2-2. ALogP Hydrophobic Parameters with Simple Indicator Variables Type Function (Method F). Simple indicator parameter was applied using the simple indicator variables type function described in Study 1−4. Indicator variables I_{lipo} were calculated in the same manner as in method D:

$$I_{lipo,k} = \begin{cases} a_i(k) & \text{if } r_k \le \text{th} \\ \dfrac{a_i(k)}{2} & \text{if th} < r_k \le 2 \times \text{th} \end{cases}$$

where a_i = AlogP hydrophobic value at the atom of the molecule i closest to the molecular represented point k.

Study 2-3. FlexS Hydrophilic Parameters with SEAL Type Function (Method G). FlexS[37] is a method for fast flexible ligand superposition developed by Lemmen et al. It uses the absolute value of an atom a, |chr(a)|, and classifies atoms according to the following scheme into hydrophobic, hydrophilic, and ambiguous.

$$\text{if } |\text{chr}(a)| > \text{th}_1(\text{type}(a)) \Rightarrow \text{hydrophilic}$$

$$\text{else if } |\text{chr}(a)| < \text{th}_2(\text{type}(a)) \Rightarrow \text{hydrophobic}$$

$$\text{else} \Rightarrow \text{ambiguous}$$

Table 1 shows the classification of atom types and the corresponding threshold values used in the FlexS method. The heights are 1.0 for hydrophobic atoms, −1.0 for hydrophilic atoms, and 0.0 for ambiguous atoms. SEAL type Gaussian function is used for the evaluation of the interaction between the probe atom at the molecular represented point q and the atom i of the test molecule.

Table 1. Hydrophobicity Classification Scheme of FlexS[a]

type(a)	th_1	th_2
H	0.1	0.06
C, N, O, F, B	0.2	0.1
P, Cl, Br, I, S	∞	0.1

[a] Type(a) indicates the element of the corresponding atom a, and th_1 and th_2 give the upper and lower thresholds, respectively.

Study 2-4. FlexS Hydrophilic Parameters with Simple Indicator Variables Type Function (Method H). FlexS hydrophilic parameters were used in the same manner as in Studies 1-4 and 2-2.

Study 3. CoMASA Method with the Steroid External Data Set. The CoMASA method is applied for the external test set in order to validate the derived models in a more rigorous way than by simply using leave-one-cut cross-validation procedure. It is well-known that such an approach has a tendency to overestimate the predictive performance of the derived models. Steric and electrostatic parameters and evaluation functions were selected from the results of the Study 1. Hydrophilic properties and evaluation functions were also selected from Study 2. Prediction of the CBG receptor affinity based on the model obtained from 21 compounds and the test data set of 10 compounds used by Cramer et al.[1] to predict activities of examples not included to the training set have been performed with a combination of the good methods obtained from Studies 1 and 2.

Study 4. 1,5-Diarylpyrazole Analogue-Type COX-2 Inhibitors. The CoMASA method is applied to 40 1,5-diarylpyrazole analogue-type COX-2 inhibitors.[38] In this study, the probable binding conformational alignment provided by Liu et al. was used without any modification. The binding conformations and their alignment in the binding site of COX-2 were extracted from the AutoDock diarylpyrazole-COX-2 complexes. Steric and electrostatic parameters and evaluation functions were selected from the results of Study 1. Hydrophilic properties and evaluation functions were also selected from Study 2. 3D QSAR analyses of COX-2 inhibitors were carried out using a combination of the good methods obtained from Studies 1 and 2.

PLS Calculation. To extract a QSAR equation from the high-dimensional data table, the partial least-squares (PLS) method was used. The SAMPLS[39] method reported by Bush and Nachbar was applied for the above purpose. The externally produced data tables were subjected to "block-scaling",[40] which is called CoMFA_std scaling in SYBYL, before PLS calculation to avoid complications arising from the variance contributed by the numerous variables. PLS components were extracted as long as the q^2 increases. The statistical evaluation for the CoMASA was performed in the same way as described for CoMFA. The obtained statistical parameters are reported in Table 2.

QSAR Coefficient Maps. The visualization of the results of the comparative analyses in terms of field contributions was performed by computer graphics enclosing the volumes above and below particular fields values by isocontours (isopleths). We have used the field type "the CoMASA coefficient × standard deviation (STDEV × COEFF)" to obtain maps that elucidate the relationships between the differences in the molecular active site and variations in the dependent variable. The maps obtained in the various analyses are shown in Figures 3−11 and Figures 1S−7S (Figures 1S−7S are shown in Supporting Information).

Results and Discussion

Extraction of Representational Coordinates for the Molecules by Cluster Analysis. A minimum

Table 2. Summary of the Results of Different 3D QSAR Analyses of the Steroid Data Set

			CoMASA without ring centroids										CoMASA with ring centroids
	CoMFA (run 1)	CoMSIA (run 2)	method A[a] (run 3)	method B[a] (run 4)	method C[a] (run 5)	method D[a] (run 6)	method E[a] (run 8)	method F[a] (run 9)	method G[a] (run 10)	method H[a] (run 11)	method I[b] (run 12)	method J[b] (run 13)	method B[a] (run 7)
q^2	0.662	0.662	0.760	0.528	0.822	0.798	0.381	0.707	0.754	0.743	0.742	0.779	0.521
r^2	0.719	0.763	0.899	0.915	0.984	0.982	0.568	0.879	0.921	0.938	0.931	0.929	0.915
no. of comp	2	4	2	4	4	4	1	2	4	3	2	2	4
fraction													
electrostatic	–	0.535	–	0.757	0.458	0.500	–	–	–	–	0.272	0.263	0.783
steric	1.000	0.086	1.000	0.243	0.542	0.500	–	–	–	–	0.182	0.334	0.217
hydrophobic	–	0.378	–	–	–	–	1.000	1.000	1.000	1.000	0.545	0.403	–
no. of interaction points		7200	92	92	92	92	92	92	92	92	84	84	97

[a] Alignment according to ref 1. [b] SEAL-like superposition using MOPAC 93.01 AM1 charge calculation.

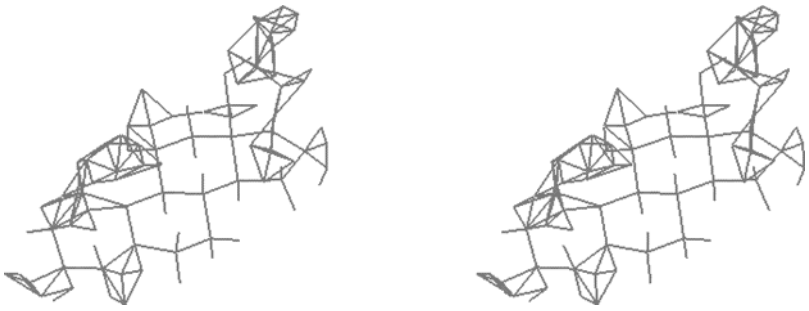


Figure 1. Stereodiagram of the molecular represented point of 21 steroids without the addition of ring centroids derived by cluster analysis.

collection of points which are spatially occupied by the molecules would be essentially sufficient to elicit a biological response. As replacement of one functional group to the other one is often carried out to obtain expected biological activity, synthetic chemists sometimes need information like pharmacophore conception. Cho et al. reported cross-validated r^2-guided region selection (q^2-GRS)[19,20] to resolve the above issue. They first subdivide the rectangular lattice initially obtained with conventional CoMFA into small boxes and perform independent analyses using probe atoms placed within each box. They then select only those small boxes for which a q^2 is higher than a specified optimal cutoff value. Finally, CoMFA with the union of small boxes selected at the previous step is repeated. The q^2-GRS method proved to be orientation-independent and derived a model having a high q^2, exceeding the one obtained with conventional CoMFA. They concluded that this method should be routinely used in future CoMFA studies to guarantee the reproducibility of reported q^2 values.

However, CoMFA and relative 3D QSAR results use a number of lattice points for the calculation of interactive effects with molecules, while the contour maps are represented outside the molecule in most methods due to described cutoff settings and the steepness of the potentials close to the molecular surface.

The MTD method uses a very similar concept to reduce complexity. However, the hypermolecule used in the MTD method is constructed by sequential superposition of the molecules of the investigated series upon the template molecule. The hypermolecule would be different depending on the template molecule, since it remains subject to the arbitrariness of the researchers such as atom-to-atom superposition procedure. An intuitive graphical interpretation of the models is important for understanding the QSAR results and the structural modifications of the ligands. The MTD method presents some difficulty for visualization of the results such as the CoMFA method because it uses topology for the QSAR analysis.

We described the possibility of cluster analysis to extract the minimum collection of points to avoid lattice-based 3D QSAR problems and the topological method issue.

Study 1. Two stereodiagrams of Figures 1 and 2 indicated the molecular represented points derived by the cluster analyses with and without the addition of ring centroids. As anticipated, the greater the diversity of the molecular structures in arbitrary areas, the closer are the extracted points.

CoMASA functions using the evaluation methods A−D with and without the addition of ring centroids for the extraction of molecular represented points were carried out. The q^2 (as defined by Cramer et al.[1] and Waller and Marshall[41,42]) amounts for the analyses based on methods A−D without the addition of ring centroids and method B with the addition of ring centroids to 0.760, 0.528, 0.822, 0.798, and 0.521, respectively, are shown in Table 2. These methods also gave a high r^2 value compared with the CoMFA and the CoMSIA methods even if the q^2 values were lower. CoMASA with simple indicator variables (method D, run 6) gave higher r^2 and q^2 than the CoMFA and the CoMSIA methods. In the present study, the addition of ring centroids did not result in any significant improvement (runs 4 vs 7, Figures 3 and 4 vs Figures 4S and 5S). Practically the same results were obtained with or

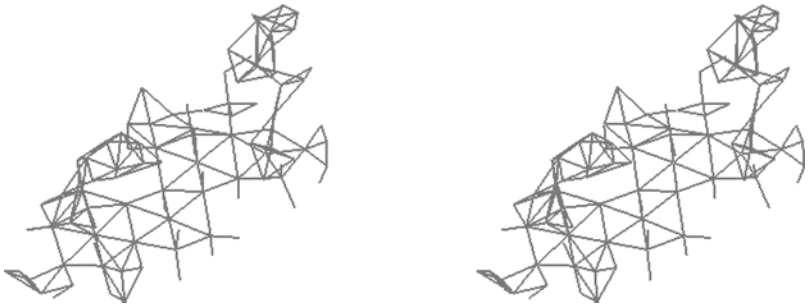


Figure 2. Stereodiagram of the molecular represented point of 21 steroids with the addition of ring centroids derived by cluster analysis.

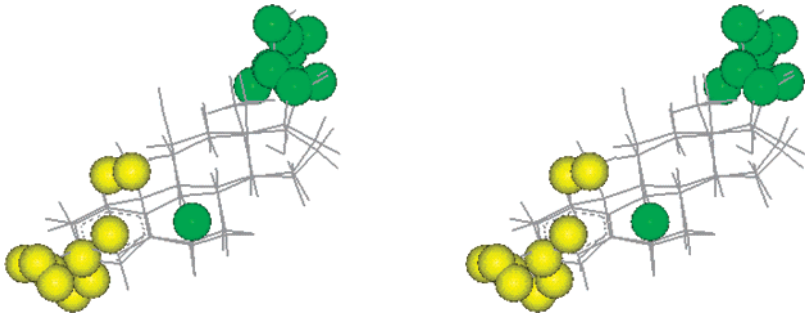


Figure 3. Stereoview of the map of method B (run 4) for the steric properties obtained for the set of steroids, the green balls (STDEV*COEFF = 0.1) indicate regions where any occupation with sterically demanding groups will enhance the activity. The yellow balls (STDEV*COEFF = −0.1) indicate where steric bulk should be reduced. Superimposed onto the map are cortisol and estradiol exhibiting high and low CBG receptor affinity, respectively.

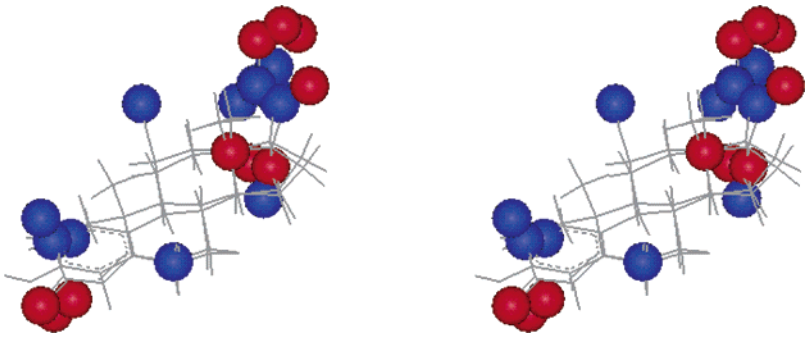


Figure 4. Stereoview of the map of method B (run 4) for the electrostatic properties obtained for the set of steroids. The red balls (STDEV*COEFF = 0.2) indicate regions where an increase in negative charge will enhance affinity; the blue balls (STDEV*COEFF = −0.2) indicate where a more positively charged group will improve the binding properties. Superimposed onto the map are cortisol and estradiol exhibiting high and low CBG receptor affinity, respectively.

without ring centroids when using methods A and D (data not shown). As expected, the maps of method B (runs 4 and 7, Figures 3, 4, 4S, and 5S) were quite similar to that of the CoMSIA result, even though q^2 was smaller (vs run 2). Although method C (run 5, Figures 2S and 3S), which used a Gaussian-type evaluation function similar to method D, gave higher q^2 than the other methods, the map of both steric and electrostatic properties indicated different tendencies from the maps generated by the other methods. Particularly, the areas that show high contributions of electrostatic properties appeared on molecular skeletons. Therefore, the evaluation function of method C seems to be inadequate for the CoMASA 3D QSAR method. It is noteworthy that methods A and D showed higher r^2 and q^2 than the CoMFA and CoMSIA methods (runs 3 and 6 vs 1 and 2, Figures 1S, 5, and 6). These methods remarkably reduced the required time to prepare descriptors compared with method B as discussed in the CoMASA methodology section. Additionally, method D was superior to method A because method D was able to calculate electrostatic property in addition to steric property.

The maps of steric properties obtained by method D (run 6) showed that the sterically favorable site was sandwiched between two steric disfavorable sites around the "3-position" of the A-ring (Figure 5). The map of electrostatic properties also showed that the positive charge favorable site was sandwiched between two negative charge favorable sites in the same region (Figure 6). These maps showed that the CoMASA method would give sharper results than the other methods.

The relative contributions of steric and electrostatic properties were different between the CoMSIA method, method B, and method D. In the CoMSIA method, the electrostatic contribution was significant, whereas method D showed equal contribution (Table 2). The relative contribution to method B exhibited similar tendencies to that of the CoMSIA method since both of them used

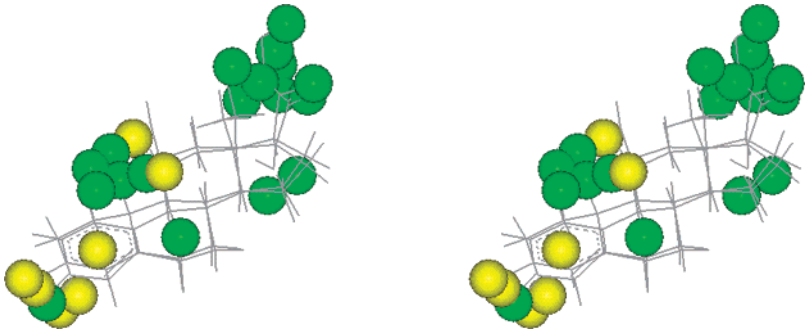


Figure 5. Stereoview of the map of method D (run 6) for the steric properties obtained for the set of steroids. The color legend is as in Figure 3 (STDEV*COEFF = ±0.15).

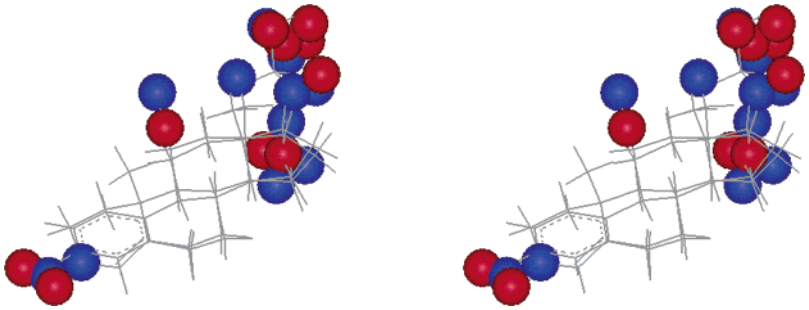


Figure 6. Stereoview of the map of method D (run 6) for the electrostatic properties obtained for the set of steroids. Color legend is as in Figure 4.

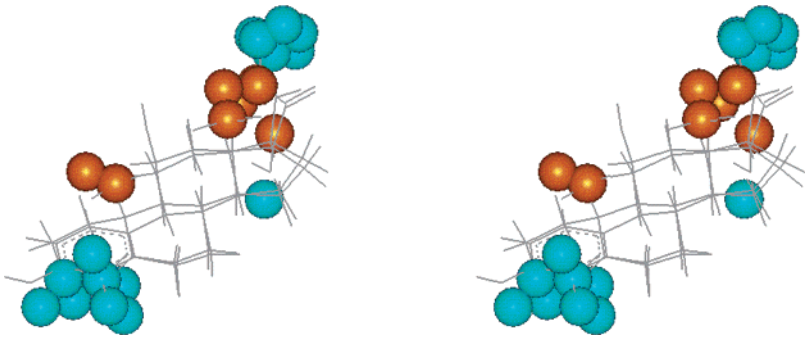


Figure 7. Stereoview of the map of method G (run 10) for the electrostatic properties obtained for the set of steroids. The orange balls (STDEV*COEFF = 0.02) indicate regions where a decrease in hydrophobicity will enhance the affinity; the sky-blue balls (STDEV*COEFF = −0.02) indicate regions where a more hydrophilic group will improve the affinity. Superimposed onto the map are cortisol and estradiol exhibiting high and low CBG receptor affinity, respectively.

the SEAL fitting function for the generation of descriptors.

Study 2. Great differences appeared in the selection of hydrophobic parameters. Table 2 listed the results of Study 2. ALogP hydrophobic parameters with the SEAL type function (method E) gave only the region where a more hydrophilic group will improve the affinity at 0.02 of STDEV × COEFF, in addition to relatively low r^2 and q^2 (run 9, Figure 6S). For method F, as r^2 and q^2 showed the highest values, the regions, where a decrease in the hydrophobicity will enhance the affinity and where a more hydrophilic group will improve the affinity, were intermingled (run 10, Figure 7S). For that reason, hydrophobic parameters developed by Viswanadhan et al. seemed to be unsuitable for the CoMASA method. Alternatively, hydrophobic parameters used in FlexS gave satisfactory results on the CoMASA method as shown in Figures 7 and 8 (methods G and H, runs 10 and 11). Both r^2 and q^2 values were high enough compared with the ordinary 3D QSAR methods. Especially, the CoMASA method indicated that the introduction of a hydrophilic group at the terminus of the side chain increases CBG binding affinity. These results agree with the importance of the hydroxy group on deoxycortisol and deoxycprticosteron for affinity to human CBG.

Study 3. The CoMASA method was applied for 31 steroids included the external test set in order to validate the derived models in a more rigorous way than by simply using leave-one-cut cross-validation procedure. The training set was made up 31 molecules, a steroid set that binds to CBG receptors. Steroid geometries used in this work have been kindly provided by the Gasteiger group.[43] Partial charges of steroids were calculated by MOPAC 93.01[44] using AM1 Hamiltonian. Alignments of the data sets were used the SEAL-like method with consideration of steric interactions (in-house program). Molecular represented points were extracted from the 21 steroids (training set), in which 84 points were obtained. Prediction of the CBG receptor affinity based on the model obtained from the 21 compounds and the test data set of 10 compounds used

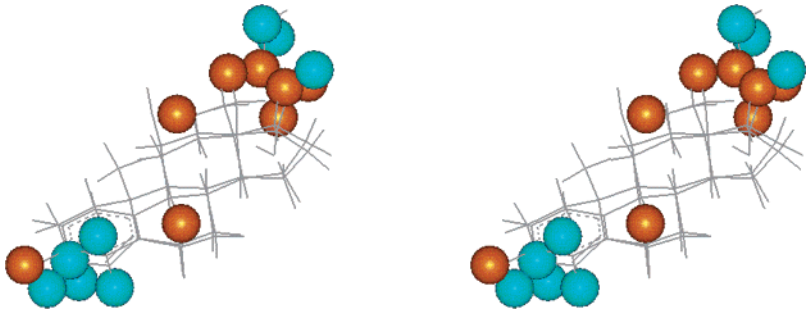


Figure 8. Stereoview of the map of method H (run 11) for the electrostatic properties obtained for the set of steroids. The color legend is as in Figure 14.

Table 3. Data Sets of 21 Steroids Used in the Training Set to Obtain the Different CoMASA Models and 10 Examples Used to Predict Binding Affinities[a]

	pK_I	method I (run 12)	method J (run 13)
aldosterone	6.28	−0.50	−0.61
11-deoxycorticosterone	7.65	0.09	0.20
11-deoxycortisol	7.88	0.35	0.27
dihydrotestosterone	5.92	−0.60	−0.21
estradiol	5.00	−0.01	−0.08
estriol	5.00	0.04	−0.06
estrone	5.00	0.43	0.03
etiocholanolone	5.23	−0.03	−0.17
pregnenolone	5.23	−0.33	0.23
17α-hydroxypregnenolone	5.00	−0.52	−0.17
progesterone	7.38	−0.11	0.12
androstanediol	5.00	0.01	−0.08
17α-hydroxyprogesterone	7.74	0.23	0.36
testosterone	6.72	−0.11	−0.05
5-androstenediol	5.00	0.12	0.31
4-androstenedione	5.76	−0.71	−0.83
androsterone	5.61	0.17	0.21
corticosterone	7.88	0.42	0.32
cortisol	7.88	0.47	0.20
cortisone	6.89	0.14	−0.29
dehydroepiandrosterone	5.00	0.45	0.31
MEAN	6.15	0.00	0.00
ST_DEV	1.18	0.36	0.31
HIGH	7.88	0.47	0.36
LOW	5.00	−0.71	−0.83
prednisolone	7.51	0.43	0.12
cortisolacetat	7.55	0.02	0.01
4-pregnene-3,11,20-trione	6.78	0.09	−0.09
epicorticosterone	7.20	−0.21	−0.18
19-nortestosterone	6.14	−0.45	−0.26
16α,17α-dihydroxyprogesterone	6.25	−1.17	−1.04
16α-methylprogesterone	7.12	−0.33	0.03
19-norprogesterone	6.82	−0.40	−0.28
2α-methylcortisol	7.69	0.72	0.25
2α-methyl-9α-fluoro-cortisol	5.80	−1.18	−1.63
MEAN	6.89	−0.25	−0.31
ST_DEV	0.65	0.61	0.58
HIGH	7.69	0.72	0.25
LOW	5.80	−1.18	−1.63

[a] The measured pK_I values are given together with residuals remained for the different models; for the predictive set the differences to the measured value are listed.

by Cramer et al.[1] to predict activities of examples not included to the training set have been performed.

Two types of methods (methods I and J) were carried out, i.e., a combination of methods B and G (method I), and a combination of methods D and H (method J) (runs 13 and 14, respectively). The results were summarized in Tables 2 and 3 (no figure). Both methods I and J gave satisfactory q^2 and r^2. The experimentally determined binding constants[45] (expressed as pK_I) are listed in Table 3 together with the residuals obtained in the different models. The pronounced outlier is 2α-methyl-9α-fluorocortisol. This is quite similar tendency to the other 3D QSAR methods such as CoMFA and CoMSIA. Since the data set did not contain a compound with a non-hydrogen atom in the 9-position, reliable predictions cannot be expected by the models. The present type of correlation analysis cannot extrapolate into the position where no data have initially been included in the training set.

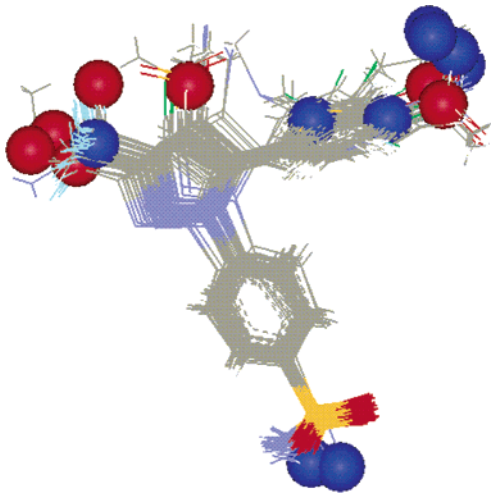


Figure 9. Stereoview of the map of method J (run 15) for the electrostatic properties obtained for the set of COX-2 inhibitors. The red balls (STDEV*COEFF = 0.02) indicate regions where an increase in negative charge will enhance affinity; the blue balls (STDEV*COEFF = −0.02) indicate where a more positively charged group will improve the binding properties. Superimposed onto the map are 40 COX-2 inhibitors.

Study 4. The CoMASA method was applied for the analysis of COX-2 inhibitory activity. Forty COX-2 inhibitors with probable binding conformations[38] were used. Two types of methods (methods I and J) are carried out the same as Study 3. The results are shown in Figures 9−11 and Table 4. For method I, q^2 was small and the number of optimized components were not obtained (data not shown). Conversely, method J gave satisfactory q^2 and r^2.

According to Soliva et al., we used three structural moieties as follows: (i) the central five-membered ring (5MR), (ii) the sulfone/sulfonamide-substituted benzene (SR), and (iii) the other substituted or unsubstituted benzene residue (BR).[46] Surprisingly, big differences were obtained for the steric favorable region around the BR moiety. It seems to be in conflict with the CoMFA and CoMSIA methods. However, the CoMFA and CoMSIA methods use the molecular field. The CoMFA

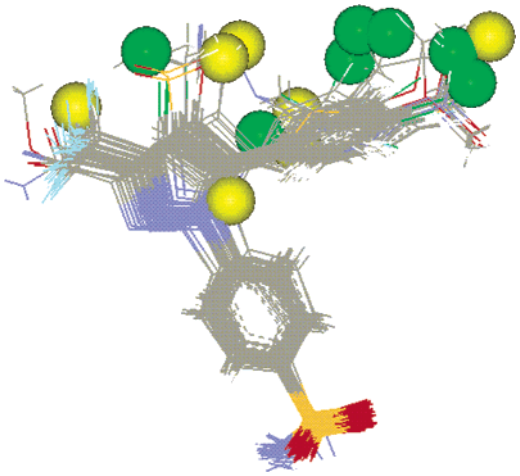


Figure 10. Stereoview of the map of method J (run 15) for the steric properties obtained for the set of COX-2 inhibitors, the green balls (STDEV*COEFF = 0.02) indicate regions where any occupation with sterically demanding groups will enhance the activity. The yellow balls (STDEV*COEFF = −0.02) indicate where steric bulk should be reduced. Superimposed onto the map are 40 COX-2 inhibitors.

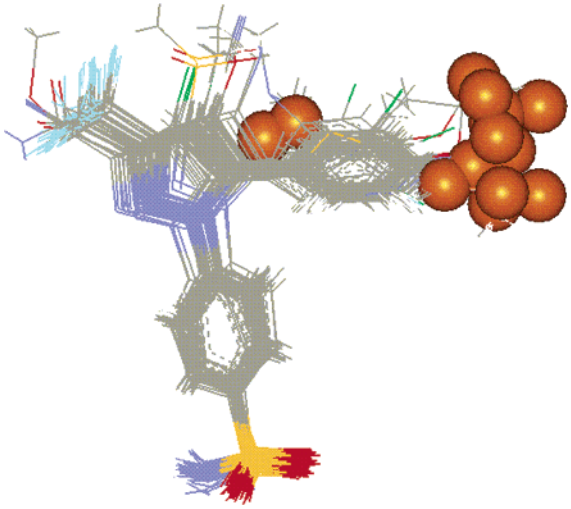


Figure 11. Stereoview of the map of method J (run 15) for the electrostatic properties obtained for the set of COX-2 inhibitors. The orange balls (STDEV*COEFF = 0.03) indicate regions where a decrease in hydrophobicity will enhance the affinity; the sky-blue balls (STDEV*COEFF = −0.03) indicate regions where a more hydrophilic group will improve the affinity. Superimposed onto the map are 40 COX-2 inhibitors.

Table 4. Results of Different 3D QSAR Analyses of COX-2 Inhibitors

	method I (run 14)	method J (run 15)
q^2	0.144	0.411
r^2	0.675	0.796
no. of compd	3	2
fraction		
electrostatic	0.478	0.379
steric	0.156	0.244
hydrophobic	0.366	0.377

method especially neglects the lattice points within the molecular clouds; therefore, contour maps appear outside of the molecules. In comparison with them, the CoMASA method directly shows locations of the preferable interactions for the target molecules, since the molecular represented points are generated from the molecular atomic position. The sterically disfavorable region around the 5MR is similar to the CoMFA and CoMSIA results. Our results according to the sterically favorable regions around the BR moiety are agreeable because the BR seems to be less tightly bound to the enzyme.[46] It is noteworthy that occasionally the CoMASA method is spatially sensitive because the disfavorable region appears next to the favorable one. To synthetic chemists this would indicate modifications of the functional groups. The CoMASA results of the electrostatic properties suggest that both trifluoromethyl on the 5MR and substituents on the BR units enhance inhibitory activity. Namely, the region around the 5MR is in accordance with the trifluoromethyl group that has negatively charged atoms around positively charged atoms. The hydrophobic favorable regions appeared around the SR moiety. This demonstrates that the phenyl ring group interacts with the side chains of the hydrophobic residues in the side pockets around Tyr 385.

Computational Speed. In the literature,[3] CoMSIA used 7200 interaction points (descriptors) for the 3D QSAR analysis of the binding affinity to human CBG, whereas CoMASA required only 92 points for the same compounds set. For analyses involving m-descriptor, n-compounds, and h-components, SAMPLS needs ($n \cdot n \cdot h$) multiplications for the regression after a preliminary calculation of covariance using ($n \cdot n \cdot m$) multiplications. SAMPLS also requires ($m \cdot n$) multiplications to convert the regression coefficients back into property space, if this is desired for the CoMFA mapping or other purposes. Since the CoMFA and the CoMASA methods use a different number of descriptors according to PLS procedures, the number of the descriptors, m, plays important roles in calculation time. Thus, CoMASA has an intrinsic advantage of 78 (= 7200/92)-fold in speed for the SAMPLS procedure in the regression after a preliminary calculation of covariance and the regression coefficients back into property space compared with the CoMASA method in the present study. It also has a theoretical speed advantage in descriptor generation, while method B had the anticipated advantage of the same degree with the SAMPLS case in calculation speeds. Using methods A and D reduces the calculation time because, if the atomic distance between the atom and the molecular represented point is less than the threshold, further atomic distances based on the molecular represented point are not required. It is not necessary to calculate all of the distances, so calculation times are reduced. The CoMASA method has an intrinsic speed advantage compared with other methods.

Conclusions

In the present paper, a rapid 3D QSAR approach to compute property fields based on the extraction of molecular represented points by cluster analysis has been described. The evaluation functions to calculate fields of different physicochemical properties are successful with both SEAL and simple indicator variables without occurring singularities at the atomic positions. A data set of steroids was analyzed by the usual CoMFA method (Lennard−Jones and Coulomb potentials), CoMSIA method (SEAL function, Gaussian-type potentials), and the CoMASA method with four different evaluation functions (rapid similarity index calculation,

SEAL function, Good's similarity index function, and simple indicator variables) with and without the addition of ring centroids. In addition, two types of hydrophobic parameters with different types of evaluation functions were examined.

CoMASA with simple indicator variables (methods D, H, and J) gave higher r^2 and q^2. This convenient method for the calculation of physicochemical properties would have great advantage over the ordinary 3D QSAR methods. In the present study, the addition of the ring centroids did not result in any significant improvement.

The CoMASA approach is by far superior to the usual 3D QSAR methods such as the CoMFA and the CoMSIA methods according to the following four advantages: (1) the resulting maps are easily understandable so that the CoMASA results would easily transform to pharmacophore and/or queries required for 3D database searches, (2) the CoMASA method is not required to consider orientation and translation of molecules against a lattice, and (3) standard PCs are available to run the CoMASA method because of its rapid scoring functions and reducing interaction points. Consequently, the CoMASA method became a useful tool for convenient 3D QSAR, especially for synthetic chemists who do not require the precise 3D QSAR results. Investigations are currently underway for the inclusion of hydrogen-acceptor/donor properties.

Appendix

Study 1-1. Rapid Similarity Index Calculation Method as an Evaluation Function (Method A). Rapid similarity indices S between compounds of interest and a probe atom systematically replaced at the molecular represented points were calculated according to (i.e., at the molecular represented point k for the molecule j of the data set):

$$S^k(j) = I_k$$

where the indicator valuable I_k at the molecular represented point k for the molecule j of the data set is simply defined by the distance r_k from the point k to the nearest atom i of the molecule j.

$$I_k = \begin{cases} 1.0 & \text{if } r_k \le \text{rlim1} \\ 0.5 & \text{if rlim1} < r_k \le \text{rlim2} \end{cases}$$

where

$$\text{rlim1} = \alpha_1 \frac{\text{vdW}_i + 1.0}{2}$$

$$\text{rlim2} = \alpha_2 \frac{\text{vdW}_i + 1.0}{2}$$

and vdW_i = vdW radii of the nearest atom i of the molecule j which is used for calculation of r_k. Similarly to the CoMSIA approach, the vdW radius of the probe atom has been set to 1 Å. This method has the great advantage of reducing calculation time because if the atomic distance between the atom j and the molecular represented point k is less than rlim1, then further atomic distances based on the molecular represented point need not be calculated. All of the distances r_k are not needed, and it helps save calculation time. In the present study α_1 and α_2 were set to 1.3 and 0.6, respectively, the same as the rapid similarity calculation method.

Study 1-2. SEAL as an Evaluation Function (Method B). Similarity indices $A_{F,k}$ between compounds of interest and a probe atom were calculated according to (i.e., at the molecular represented point q for the molecule j of the data set):

$$A_{F,k}{}^q(j) = -\sum_{i=1}^{n} (w_{\text{probe},k} w_{i,k} e^{-ar_{iq}{}^2})$$

where i = summation index over all atoms of the molecule j under investigation, $w_{i,k}$ = actual value of the physicochemical property k of the atom i, $w_{\text{probe},k}$ = probe atom with charge +1 and radius 1 Å, α = attenuation factor, and r_{iq} = mutual distance between the probe atom at the molecular represented point q and the atom i of the test molecule. Large values of α will result in a strong attenuation of the distance-dependent consideration of molecular similarity. Accordingly, there is little averaging of local feature matches of the molecules being compared. The global molecular similarity becomes less important. With small values of α, the global molecular features become more important. In the present study, α has been set to 0.3 as in the CoMSIA approach. With this selection, at the given molecular represented point, the property value of an atom of the molecule under investigation (i.e., the partial atomic charge) is observed at 1 Å distance by 74.1%, in 2 Å by 30.1%, and in 3 Å by 6.7% of its total value.

Acknowledgment. We would like to thank Professor Miki Akamatsu (Kyoto University, Japan) for her valuable comments. We also thank Professors Hong Liu, Jianhua Shen, and Hualiang Jiang (Shanghai Institute of Materia Medica, People's Republic of China) for their generosity in providing us the alignment files of COX-2 inhibitors.

Supporting Information Available: Figures 1S−7S. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cramer, R. D., III; 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) Kellogg, G. E.; Abraham, D. J. Hydrophobicity: Is LogP(o/w) More Than the Sum of Its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651−661.
(3) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.
(4) Klebe, G. Comparative Molecular Similarity Indices Analysis: CoMSIA. *Perspect. Drug Discovery Des.* **1998**, *12/13/14*, 87−104.
(5) Klebe, G.; Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) to Study Hydrogen-Bonding Properties and to Score Combinatorial Libraries. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1−10.
(6) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(7) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. *J. Med. Chem.* **1989**, *32*, 1083−1094.

(8) Wade, R.; Clerk, K. J.; Goodford, P. J. Further Development of Hydrogen Bond Function for Use in Determining Generically Favorable Binding Sites on Molecules of Known Structure. Ligand Probe Groups with the Ability to Form Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 140−147.
(9) Sulea, T.; Oprea, T. I.; Muresan, S.; Chan, S. L. A Different Method for Steric Field Evaluation in CoMFA Improves Model Robustness. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1162−1170.
(10) Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular Lipophilicity Potential, a Tool in 3D QSAR: Method and Applications. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83−96.
(11) Furet, P.; Cohen, N. C. 3D Molecular Lipophilicity Potential Profiles: A New Tool in Molecular Modeling. *J. Mol. Graph.* **1988**, *6*, 182−189.
(12) Fauchère, J. L.; Quarendon, P.; Kaetterer, L. Estimating and Representing Hydrophobicity Potential. *J. Mol. Graph.* **1988**, *6*, 202−206.
(13) Silverman, B. D.; Platt, D. E.; Pitman, M.; Rigoutsos, I. Comparative Molecular Moment Analysis (CoMMA). *Perspect. Drug Discovery Des.* **1998**, *12/13/14*, 183−196.
(14) Robert, D.; Amat, L.; Carbo-Dórca, R. Three-Dimensional Quantitative Structure−Activity Relationships from Tuned Molecular Quantum Similarity Measures: Prediction of the Corticosteroid-Binding Globulin Binding Affinity for a Steroid Family. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 333−344.
(15) Silverman, B. D.; Platt, D. E.; Pitman, M.; Rigoutsos, I. Comparative Molecular Moment Analysis (CoMMA). *Perspect. Drug Discovery Des.* **1998**, *12/13/14*, 183−196.
(16) Lobato, M.; Amat, L.; Besalu, E.; Carbó-Dorca, R. Structure−Activity Relationships of a Steroid Family Using Quantum Similarity Measures and Topological Quantum Similarity Indices. *Quant. Struct.-Act. Relat.* **1997**, *16*, 465−472.
(17) Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.
(18) Klebe, G.; Mietzner, T.; Weber, F. Different Approaches Toward an Automatic Structural Alignment of Drug Molecules: Applications to Sterol Mimics, Thrombin and Thermolysin Inhibitors. *J. Comput.-Aided Mol. Des.* **1994**. *8*. 751−778.
(19) Cho, S. J.; Tropsha, A. Cross-Validated R^2-Guided Region Selection for Comparative Molecular Field Analysis: a Simple Method to Achieve Consistent Results. *J. Med. Chem.* **1995**, *38*, 1060−1066.
(20) Tropsha, A.; Cho, S. J. Cross-Validated R^2 Guided Region Selection for CoMFA Studies. *Perspect. Drug Discovery Des.* **1998**, *12/13/14*, 57−69.
(21) Doweyko, A. M. Three-Dimensional Pharmacophores from Binding Data. *J. Med. Chem.* **1994**, *37*, 1769−1778.
(22) Doweyko, A. M. The Hypothetical Active Site Lattice. An Approach to Modelling Active Sites From Data on Inhibitor Molecules. *J. Med. Chem.* **1988**, *31*, 1396−1406.
(23) Guccione, S.; Doweyko, A. M.; Chen, H.; Barretta, G. U.; Balzano, F. 3D-QSAR Using 'Multiconformer' Alignment: the Use of HASL in the Analysis of 5-HT1A Thienopyrimidinone Ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 647−657.
(24) Oprea, T. I.; Ciubotariu, D.; Sulea, T. I.; Simon, Z. Comparison of the Minimal Steric Difference (MTD) and Comparative Molecular Field Analysis (CoMFA) Methods for Analysis of Bonding of Steroids to Carrier Proteins. *Quant. Struct.-Act. Relat.* **1993**, *12*, 21−26.
(25) Ciubotariu, D.; Deretey, E.; Oprea, T. I.; Sulea, T.; Simon, Z.; Kurunczi, L.; Chiriac, A. Multiconformational Minimal Steric Difference. Structure-Acetylcholinesterase Hydrolysis Rates Relations for Acetic Acid Esters. *Quant. Struct.-Act. Relat.* **1993**, *12*, 367−372.
(26) Kurunczi, L.; Olah, M.; Oprea, T. I.; Bologa, C.; Simon, Z. MTD-PLS: A PLS−Based Variant of the MTD Methodol. 2. Mapping Ligand−Receptor Interactions. Enzymatic Acetic Acid Esters Hydrolysis. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 841−846.
(27) Oprea, T. I.; Kurunczi, L.; Olah, M.; Simon, Z. MTD-PLS: A PLS−Based Variant of the MTD Methodol. A 3D-QSAR Analysis of Receptor Affinities for a Series of Halogenated Dibenzoxin and Biphenyl Derivatives. *SAR QSAR Environ. Res.* **2001**, *12*, 75−92.
(28) Kroemer, R. T.; Hecht, P. A New Procedure for Improving the Predictiveness of CoMFA Models and its Application to a Set of Dihydrofolate Reductase Inhibitors. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 396−406.
(29) Kotani, T.; Higashiura, K. Rapid Evaluation of Molecular Shape Similarity Index Using Pairwise Calculation of the Nearest Atomic Distances. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 58−63.
(30) WebLab ViewerLite 4.0, Accelrys Inc., San Diego, CA.
(31) Good, A. C.; Hodgkin, E. E.; Richards, W. G. Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188−191.
(32) Good, A. C.; Richards, W. G. Rapid Evaluation of Shape Similarity Using Gaussian Function. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112−116.
(33) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three-Dimensional Structure Directed Quantitative Structure−Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Applications for an Automated Superimposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.
(34) Viswanadhan, V. N.; Ghose, A. K.; Singh, U. C.; Wendoloski, J. J. Prediction of Solvation Free Energies of Small Organic Molecules: Additive-Constitutive Models Based on Molecular Fingerprints and Atomic Constants. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 405−412.
(35) Viswanadhan, V. N.; Ghose, A. K.; Wendoloski, J. J. Estimating Aqueous Solvation and Lipophilicity of Small Organic Molecules: A Comparative Overview of Atom/group Contribution Methods. *Perspect. Drug Discovery Des.* **2000**, *19*, 85−98.
(36) Viswanadhan, V. N.; Reddy, M. R.; Bacquet, R. J.; Erion, M. D. Assessment of Methods Used for Predicting Lipophilicity: Application to Nucleosides and Nucleoside Bases. *J. Comput. Chem.* **1993**, *14*, 1019−1026.
(37) Lemmen, C.; Lengauer, T.; Klebe, G. FlexS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem.* **1998**, *41*, 4502−4520.
(38) Liu, H.; Huang, X.; Shen, J.; Luo, X.; Li, M.; Xiong, B.; Chen, G.; Yang, Y.; Jiang, H.; Chen, K. Inhibitory Mode of 1,5-Diarylpyrazole Derivatives Against Cyclooxygenase-2 and Cyclooxygenase-1: Molecular Docking and 3D QSAR Analyses. *J. Med. Chem.* **2002**, *45*, 4816−4827.
(39) Bush, B. L.; Nachbar, R. B., Jr. Sample-Distance Partial Least Squares: PLS Optimized for Many Variables, with Application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587−619.
(40) Nilsson, J. Multiway Calibration in 3D QSAR. Applications to Dopamine Receptor Ligands. URL at http://www.ub.rug.nl/eldoc/dis/science/j.nilsson.
(41) Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-Dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-Determined Alignment Rules. *J. Med. Chem.* **1993**, *36*, 4152−4160.
(42) Waller, C. L.; Marshall, G. R. Three-Dimensional Quantitative Structure−Activity Relationship of Angiotensin-Converting Enzyme and Thermolysin Inhibitors. II. A Comparison of CoMFA Models Incorporating Molecular Orbital Fields and Desolvation Free Energies Based on Active-Analogue and Complementary-Receptor-Field Alignment Rules. *J. Med. Chem.* **1993**, *36*, 2390−2403.
(43) The steroid structures can be downloaded in the Gasteiger's group Website: http://www2.chemie.uni-erlangen.de/services/steroids/.
(44) MOPAC 93 by Stewart, J. J. P., Fujitsu. MOPAC 93.01 revised for Cygwin by Kotani, T.
(45) Good, A. C.; So, S. S.; Richards, W. G. Structure−activity relationships from molecular similarity matrices. *J. Med. Chem.* **1993**, *36*, 433−438.
(46) Soliva, R.; Almansa, C.; Kalko, S. G.; Luque, F. J.; Orozco, M. Theoretical Studies on the Inhibition Mechanism of Cyclooxygenase-2. Is There a Unique Recognition Site? *J. Med. Chem.* **2003**, *46*, 1372−1382.

JM030364C