New Molecular Modeling Tools Using Three-Dimensional Chemical Substructures

Mark G. Bures,* Elizabeth Danaher, Jerry DeLazzer, and Yvonne C. Martin

Pharmaceutical Products Division, Abbott Laboratories, 1 Abbott Park Road, Abbott Park, Illinois 60064-3500

Received June 25, 1993*

Evaluation of the three-dimensional (3D) similarity in a set of chemical structures is a key aspect of molecular modeling. Described here are three new approaches to 3D similarity analysis, each based on interpoint distance comparisons. The approaches are embodied in three computer programs, DISCO, FAMILY, and COMPAIR. Recent applications of these programs discussed are a pharmacophore mapping study for a set of platelet-activating factor (PAF) antagonists, grouping compounds identified from a 3D database search for potential PAF antagonists and comparing search hits to a reference antagonist or to a model of the receptor binding site. The methods and algorithms used in these approaches are briefly described, followed by a discussion of the three recent applications.

INTRODUCTION

The manipulation and evaluation of three-dimensional (3D) chemical structures is the cornerstone of molecular modeling and design. Four important facets of 3D structure handling are substructure searching, pharmacophore mapping, *de novo* design, and clustering. Over the past several years, a number of reviews and research articles have elaborated the techniques and applications involved in these four areas.[1-5]

The concept of 3D similarity is common to these methodologies. Evaluation of two-dimensional (2D) similarity of chemical substructures is well-established.[5] Recently, approaches used to study 2D similarity have been extended to 3D structures. Much of the work in the area of 3D similarity has recently been reviewed.[6] Over the past several years, researchers have studied a wide range of 3D similarity measures. Representative examples include comparing atom pairs,[7] molecular shape and volume,[8] electrostatic potentials,[9] and molecular fields.[10]

Two approaches to 3D similarity receiving considerable attention are substructure searching (3D database searching)[11,12] and interatom and interpoint distance comparisons.[3] 3D database searching is now a widely used technique, showing utility in evaluating 3D similarity in sets of compounds, identifying novel lead compounds for drug design, and locating atoms, points, and substructures of interest in molecular structures.

Willett and colleagues are actively analyzing and developing 3D similarity measures based on comparisons of interatomic distance matrices.[13,14] One application of this work is the identification of 3D substructures common to a pair or larger set of structures.[15] The focus of this report is the development of new methodologies for evaluating 3D molecular similarity based on interpoint distance comparisons. Applications of this work, as elaborated here, include pharmacophore mapping, clustering sets of compounds, and proposing bioactive superpositions and receptor binding modes of compounds of interest. Background information on the techniques and methods used is presented first. This is followed by a description of how the computational tools generated from this work have been used in molecular modeling exercises.

METHODS

The approaches to evaluating 3D similarity described here have been implemented in a computer program called DISCO (DIStance COmparisons). In addition, two derivative programs of DISCO, FAMILY, and COMPAIR, have been developed and their utility is also described.

DISCO

DISCO[3] is a program for identifying and evaluating 3D similarity based on interpoint distance comparisons. A main objective of a DISCO analysis is to find sets of points common to a collection of structures. The points considered include individual atoms, projections from atoms to hypothetical intermolecular hydrogen-bond donating or accepting sites, and points at the center-of-mass of collections of atoms such as rings. In our work, these points are generated by our 3D searching program ALADDIN.[2,3] The program has found much utility as an aid to pharmacophore mapping, i.e., determining the chemical and spatial requirements for bioactivity.[3] A representative pharmacophore mapping exercise using DISCO is discussed below.

The algorithm used in DISCO has been extensively described;[3] therefore only highlights are discussed here. Running DISCO requires a reference structure, which is compared to every conformation of each compound in the set studied. The first step is calculation of the distances between all pairs of points being considered in each structure. Next, distances that are within a given tolerance and that are considered to be of the same type (as defined by the user) are noted, and this information is stored in a correspondence table. Tolerance is a parameter used in DISCO to specify a value at which two interpoint distances are considered equivalent. Bron-Kerbosh clique detection[15,16] is then used to find sets of common distances, i.e., cliques. Chirality is also considered at this time, by evaluating the sign of the torsion angle between four points not in a plane. DISCO solutions (e.g. pharmacophore maps) are those cliques that meet the criteria specified by the user.

User-specified criteria include the number and types of points required, tolerance range, and potential compounds to eliminate from consideration during the run. Each solution is a set of conformations and points that match the same set of points in the reference structure. Typically, the distance tolerance is iterated until either solutions are found or the user-specified maximum is reached. The algorithm used in DISCO is depicted in Figure 1.

* Abstract published in *Advance ACS Abstracts*, January 15, 1994.

0095-2338/94/1634-0218$04.50/0

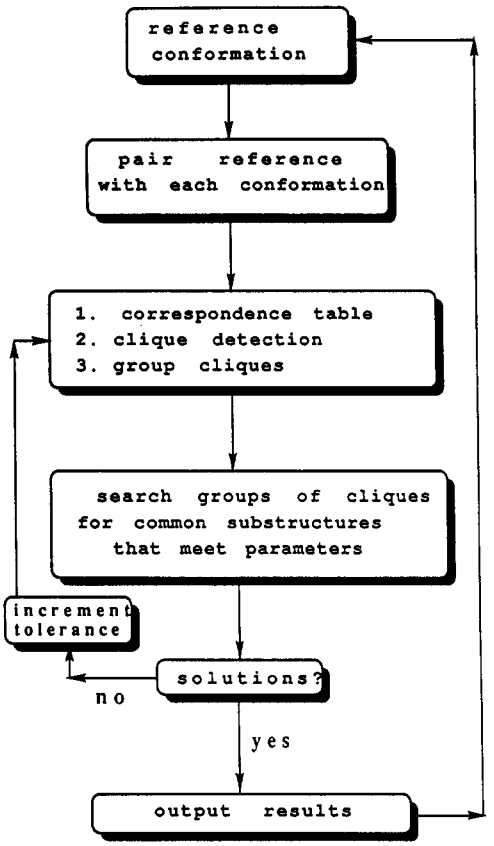

Figure 1. Flow chart of the algorithm used in the DISCO program.

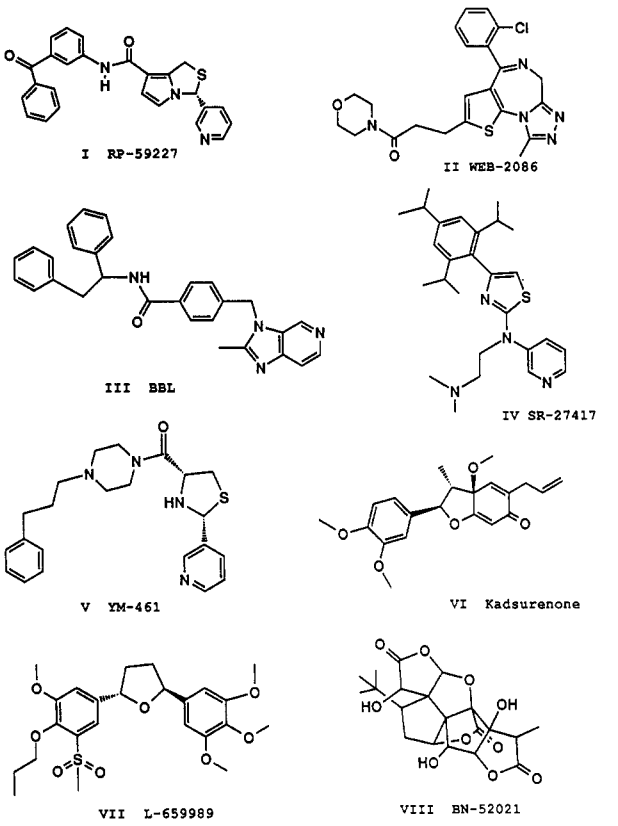

Figure 2. Set of representative platelet-activating factor antagonists studied.

FAMILY

Comparing interpoint distances in sets of 3D structures using clique detection has found additional utility in molecular modeling studies. A derivative of DISCO, a program called FAMILY, that groups compounds based on interpoint distances has been developed. In FAMILY, the emphasis is on clustering those compounds that have a common set of interpoint distances, rather than finding a collection of points common to an *entire* set of compounds (as in DISCO). As discussed below, one important application of this program is to group compounds found from 3D database searching.

The algorithm used in FAMILY will be reviewed in the literature shortly.[17] The program uses the same clique detection strategy employed in DISCO. In FAMILY, however, cliques are searched between the first structure in a list and each successive structure, comparing only two structures at a time. Once a clique is found for the first structure, this clique is searched for in each remaining structure in the list. Therefore, only the first clique found for the pair of structures being considered is used to search the list further. Structures that contain this clique are placed in the same group (family). The process continues by using the next ungrouped structure as the starting structure. The list is traversed until all structures are grouped. It is important to note that the order of structures in the list directly affects the clusters generated because the first structure is always used to search for the first clique. Application of FAMILY to grouping hits from searches for potential platelet-activating factor (PAF) antagonists is presented in the next section.

COMPAIR

A third derivative of DISCO, a program called COMPAIR, is similar in concept to FAMILY. COMPAIR also compares two structures at a time, a user-supplied reference structure and the structure of interest. In this way, a set of compounds can be compared to one reference structure and different cliques can be found for each pair. In the 3D searching context, this translates to accepting partial matches to a complex query. Two applications of this program discussed below are comparing structures identified from 3D database searching to a reference structure and orienting search hits in a receptor binding site.

APPLICATIONS AND DISCUSSION

Pharmacophore Mapping. DISCO has been used to propose a pharmacophore map for a set of representative PAF antagonists, shown in Figure 2. The compounds chosen for this study display a K_i of binding to PAF receptors from ca. 1–100 nM.[18] The compounds span a range of structural diversity and conformational flexibility.

The typical first step in a DISCO study is to generate a set of representative low-energy conformations for the compounds being evaluated. A starting structure for each compound was generated using CONCORD,[19] except for compound **VIII**. For this rigid structure a crystal structure from the Cambridge Structural Database[20] was used and no additional conformations were generated. The distance geometry program DGEOM[21] was used to generated approximately 20 representative conformations for each of compounds **I–VII**. The conformations were minimized using AMPAC[22] (AM1 Hamiltonian). All conformations within ca. 10 kcal/mol of the lowest energy conformation generated for each compound were examined in DISCO.

The potential pharmacophore points considered for this set of compounds were identified by using our 3D substructure searching program ALADDIN.[2] The set of points was comprised of hydrogen-bond donating and accepting site points, one type of ligand point, and points representing

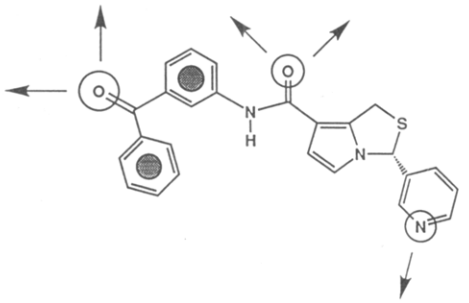

Figure 3. Set of potential pharmacophore points for a representative PAF antagonist. Arrows represent hydrogen-bonding site points, shaded circles represent hydrophobic/dispersion interaction sites, and open circles represent ligand site points.

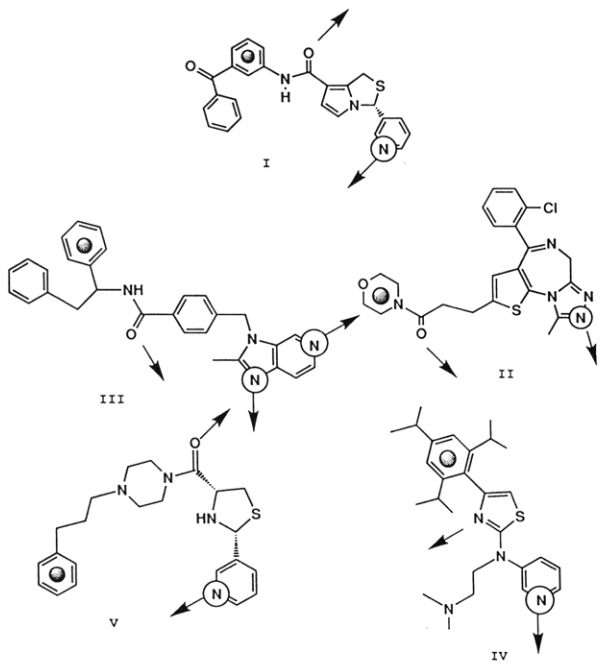

Figure 4. 2D representation of a pharmacophore map generated by using DISCO for the representative PAF antagonists. See text or Figure 3 for an explanation of the symbols used.

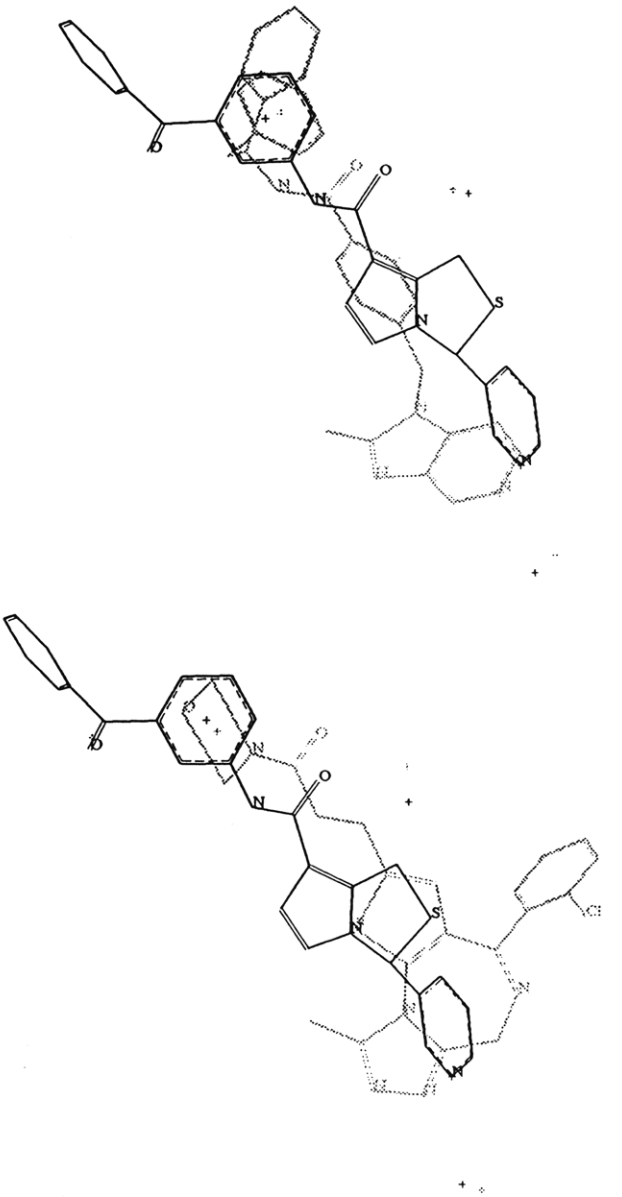

Figure 5. Selected superpositions of representative PAF antagonists, using a pharmacophore map from the DISCO analysis. Asterisks represent the proposed pharmacophore points generated by using DISCO. (a, top) Compound **I** (bold lines) is superimposed with **III**. (b, bottom) Compound **I** (bold lines) is superimposed with **II**.

hydrophobic/dispersion interaction regions. These points are illustrated in Figure 3 for a representative compound from the set.

Running DISCO is typically an iterative process, with processing times generally from 1 to 10 min on a VAX-9000 or SGI-4D/340. In this study, compound **VIII** was initially chosen as the reference structure, because of its rigid nature. However, because the distance between pharmacophore points in **VIII** is significantly shorter than those in **I–VII**, no solutions in which there was a good overall superposition were found. Therefore, compound **VIII** was removed from consideration in further DISCO runs.

Next, compound **I** was chosen as the reference compound, because we had proposed a bioactive conformation for this structure based on a 3D-QSAR and a receptor model (unpublished results). Examining compounds **I–VII**, DISCO found several pharmacophore models, each with a tolerance in the range between 3 and 4 Å. Because a large tolerance value was required, each compound had multiple orientations and conformations that fit this solution. Analysis of the results showed that compounds **VI** and **VII** were forcing high-tolerance values to be used in order to find a solution.

At this point, we decided to focus on compounds **I–V**, eliminating **VI–VIII** from further DISCO runs. Considering the substructural similarity present in compounds **I–V** (Figure 2), it is reasonable to propose that these compounds would be binding in the same manner to one receptor binding site. Conversely, considering the structural diversity of compounds **VI–VIII**, relative to **I–V**, it is perhaps not surprising that our pharmacophore mapping exercises were not able to produce a common superposition for the entire set of compounds. However, the fact that we could feel confident that no conformer of **VI–VIII** matches **I–V** increases the objectivity of the pharmacophore analysis.

As before, a proposed bioactice conformation of compound **I** was used as the reference. In this run, we evaluated a four-point model, comprised of a hydrophobic region, a hydrogen-bond donating site, a hydrogen-bond accepting site, and an aromatic nitrogen ligand site (this point was chosen to serve as a common "anchor" to help orient the compounds). One solution was found, at 1.7-Å tolerance, for the one conformation of compound **I** studied. A 2D representation of the pharmacophore points in this solution is shown in Figure 4. The

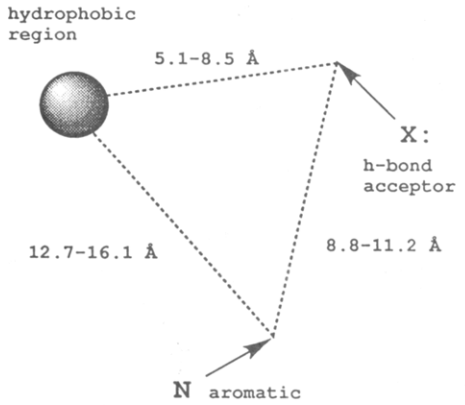

Figure 6. Schematic of the 3D search query used to identify potential PAF antagonists.

shaded circle represents the hydrophobic site, the arrows represent the hydrogen-bonding sites, and the open circle denotes the aromatic nitrogen anchor point. Selected superpositions from this solution are shown in Figure 5. Compound **III** had two superposition rules, resulting from two different aromatic nitrogen anchoring points (Figure 4), while the remainder of the compounds had only one superposition. The proposed pharmacophore map shows a good overall superposition for each compound. This illustrates the utility of this approach in selecting a common superposition rule for a set of structures with multiple conformations and potential pharmacophore points.

The information generated by using DISCO in this pharmacophore mapping analysis are representative of many of our experiences with this approach. Often, it is most fruitful to run the program in an iterative process, using the current results to guide succeeding evaluations. Accordingly, the algorithm and computational techniques used in DISCO facilitate running the program interactively, with typical run times between 1 and 10 min.

Clustering Structures. In this part of the study, the program FAMILY was used to group compounds, based on interpoint distances, identified from a 3D database search for PAF antagonists. The pharmacophore map discussed above was used to develop a target for the 3D search, depicted in Figure 6. The points required in the search for those in the map, exclusing the nitrogen anchor point: a hydrophobic point, a hydrogen-bond accepting site, and a hydrogen-bond donating site. The distance ranges between these points (Figure 6) were derived by subtracting and adding the tolerance value in the DISCO pharmacophore map (1.7 Å) from correpsonding distances in compound **I**.

A 3D structural database of commercially available compounds, called the *Available Chemicals Directory*,[23] was searched, and 90 compounds which matched the target were identified. ALADDIN[2] was used to perform the search and generate the input for the clustering analysis using FAMILY. Input consisted of the points to consider for interpoint distance comparison; in this case the points from the proposed pharmacophore map discussed above were used.

Using a tolerance value of 0.6 Å, this evaluation resulted in the 90 hits being clustered into 12 groups of varying sizes: 1 member, 4 groups; 2–10 members, 4 groups; 11–20, 3 groups; 21–30, 1 group. 2D structures of the members of two representative groups are shown in Figure 7. Group one illustrates that a set of substructural analogs will generally be placed in the same group because their interpoint distances will naturally be similar. The last compound in group one and those in group two illustrate another important point. As

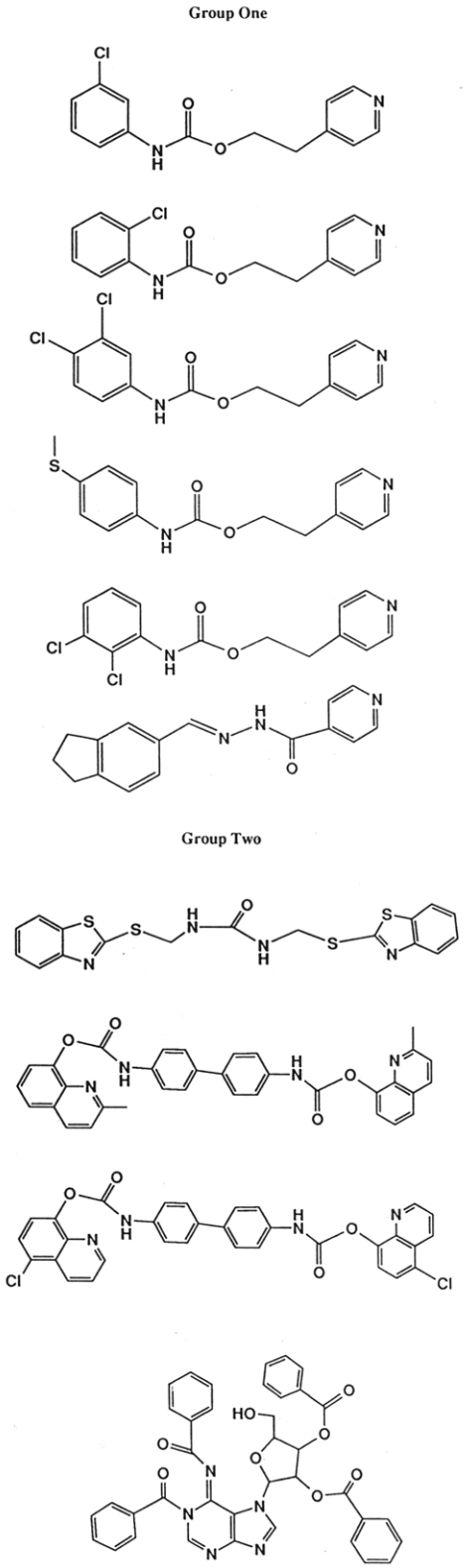

Figure 7. Two groups generated by clustering PAF search hits using the program FAMILY.

a result of the compounds being clustered on the basis of interpoint distances, groups can consist of a structurally diverse set of compounds.

We are justing beginning to explore the utility of this approach to clustering structures. The results discussed show how FAMILY can be used to group compounds identified from 3D database searching. Selecting one member from each group (e.g. for screening purposes) may be a good representation of the entire set.

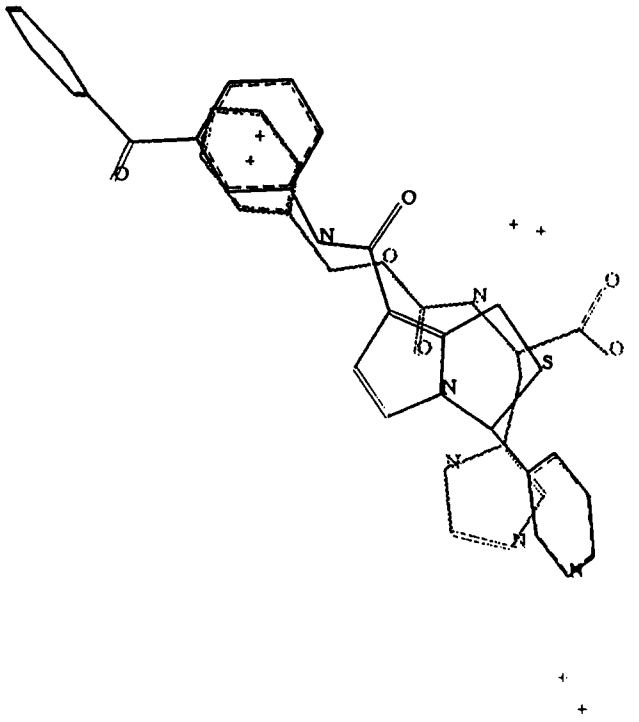

Figure 8. Superposition of a representative search hit with reference compound I (bold lines), using a clique generated by COMPAIR. Proposed pharmacophore points are represented by asterisks.

Orienting Structures. The final aspect of this study uses the DISCO derivative program COMPAIR to evaluate cliques between a reference structure and another structure of interest. Two comparisons were made: one between the proposed pharmacophore points identified in compound I and corresponding points identified in the 3D database search hits discussed above, and the other between pharmacophore points present in a 3D model of the PAF receptor binding site (unpublished) and corresponding points in the search hits. The cliques identified were used to superimpose the search hits either on reference compound I or on the binding site model. Another objective of using COMPAIR is the potential identification of cliques different from those found using DISCO, since only two structures are compared at a time.

In the first use of COMPAIR the reference structure was a proposed bioactive conformation of compound I and the points examined were the pharmacophore points (excluding the nitrogen anchor point) identified in the pharmacophore mapping exercise discussed above. The reference structure was compared to each of the 90 search hits, using all potential pharmacophore points present in the hits. As in running DISCO, a tolerance range can be used to guide clique detection. Here, the maximum tolerance value allowed was 2.0 Å.

One or more cliques were found for each hit compound. One clique for each compound was the set of pharmacophore points that correspond to the pharmacophore points identified in the pharmacophore mapping study described above. In this way, COMPAIR was useful for identifying a superposition rule for each search hit, such that the compounds could be overlaid on the reference structure, compound I. An example of these superpositions is shown in Figure 8 for one representative search hit. The additional cliques identified for some of the hits represent alternative superpositions (not shown).

The second aspect of using COMPAIR involves comparing potential pharmacophore points in each of the search compounds to proposed pharmacophore points present in a model of the PAF receptor binding site. The PAF receptor is a

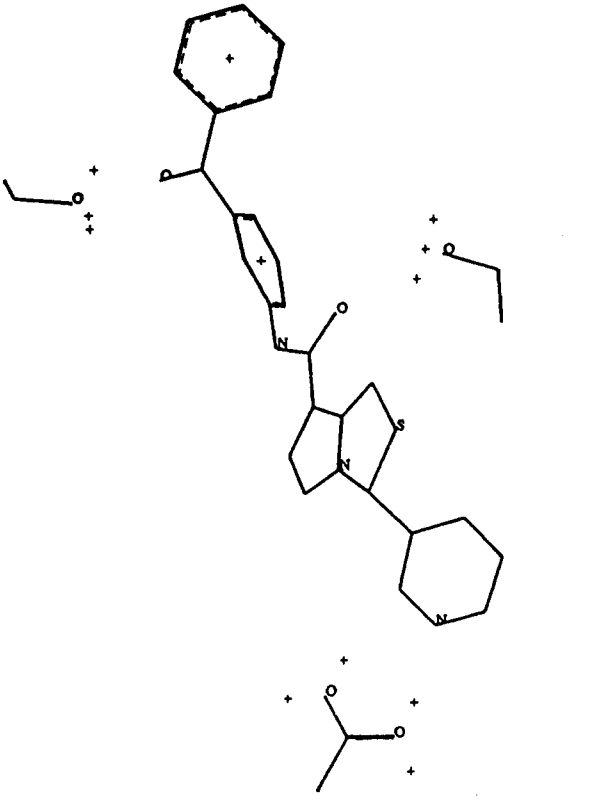

Figure 9. Model of compound I bound in a putative PAF receptor binding site. Potential receptor interaction sites are depicted as asterisks.

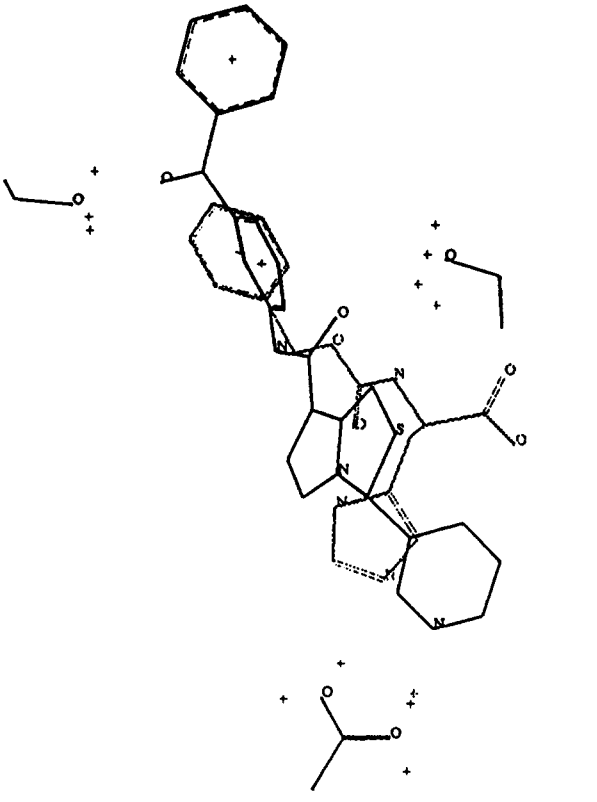

Figure 10. Superposition of a representative search hit (light lines) with the PAF receptor binding site model, using a clique generated by COMPAIR. Asterisks denote the points superimposed.

member of the G-protein linked class of receptors. A model of the putative transmembrane portion of the receptor has been developed by Hutchins.[24] A model of compound I oriented in the receptor binding site is illustrated in Figure 9. The proposed pharmacophore points from compound I are

shown to interact with key residues in the binding site. For example, the pharmacophore point derived from the pyridyl nitrogen (protonated form) of **I** is interacting with an aspartate residue, and the carbonyl oxygens of **I** are making hydrogen-bonding interactions with the hydroxyl groups of threonine residues.

As in the previous example COMPAIR was used to generate superposition rules for each of the search hits, in order to orient them in the receptor binding site. The maximum tolerance value allowed was 2.0 Å. One superposition rule for each compound was comprised of the points corresponding to the pharmacophore points identified for compound **I**. An examples of these superpositions is shown in Figure 10 for one representative hit compound. It is important to note that these orientations only consider superposition of pharmacophore points and do not evaluate any potential steric clashes between the hit compound and the receptor.

CONCLUSIONS

Three new computational tools, DISCO, FAMILY, and COMPAIR, for evaluating 3D similarity based on interpoint distance comparisons have been developed. These approaches have demonstrated utility in pharmacophore mapping, grouping sets of compounds, and orienting compounds relative to a reference structure or to a receptor binding site.

REFERENCES AND NOTES

(1) Martin, Y. C.; Bures, M. G.; Willett, P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1990; p 213.
(2) Van Drie, J. H.; Weininger, D.; Martin, Y. C. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 225.
(3) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83.
(4) Martin, Y. C. *Tetrahedron Comput. Methodol.* **1990**, *3*, 15.
(5) Willett, P. *Similarity and Clustering in Chemical Information Systems*; Research Studies Press: New York, 1990.
(6) Bures, M. G.; Martin, Y. C.; Willett, P. *Topics Stereochem.*, submitted for publication.
(7) Bakarat, M. T.; Dean, P. M. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 107.
(8) Meyer, A. M.; Richards, W. G. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 426.
(9) Richard, A. M. *J. Comput. Chem.* **1991**, *12*, 959.
(10) Cramer, R. D., III: Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
(11) Bures, M. G.; Hutchins, C. W.; Maus, M.; Kohlbrenner, W.; Kadam, S.; Erikson, J. W. *Tetrahedron Comput. Methodol.* **1992**, *3*, 673.
(12) Bures, M. G.; Black-Schaefer, C.; Garder, G. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 323.
(13) Pepperrell, C. A.; Willett, P. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 455.
(14) Pepperrell, C. A.; Taylor, R.; Willett, P. *Tetrahedron Comput. Methodol.* **1990**, *3*, 575.
(15) Brint, A. T.; Willett, P. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 152.
(16) Kuhl, F. S.; Crippen, G. M.; Friesen, D. K. *J. Comput. Chem.* **1984**, *5*, 24.
(17) Martin, Y. C.; DeLazzer, J. Manuscript in preparation.
(18) Hodgkin, E. E.; Miller, A.; Whittaker, M. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 597. Herbert, J. M.; Valette, G.; Bernat, A.; Savi, P.; Maffrand, J. P.; Le Fur, G. *Drugs Future* **1992**, *17*, 1011.
(19) Pearlman, R. S.; Rusinko, A., III; Skell, J. M.; Balducci, R.; McGarity, C. M. *CONCORD*, distributed by Tripos Associates, Inc., 1969 S. Hanley Rd., Suite 303, St. Louis, MO 63944.
(20) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrea, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187.
(21) Blaney, J. *DGEOM* is available from the Quantum Chemical Program Exchange, Indiana University, Bloomington, IN.
(22) *AMPAC*, version 2.1 (QCPE No. 506), available from Quantum Chemical Program Exchange, Indiana University, Bloomington, IN.
(23) *Available Chemicals Directory* available from Molecular Design Limited, San Leandro, CA.
(24) Hutchins, C. Manuscript in preparation.