Three-Dimensional Shape-Based Searching of Conformationally Flexible Compounds

Mathew Hahn

Molecular Simulations Incorporated, 9685 Scranton Road, San Diego, California 92121

Received June 30, 1996[⊗]

Steric shape plays a crucial role in receptor-ligand binding. This paper describes a three-phase database searching strategy for rapidly finding compounds similar in shape to a given shape query. The query is constructed from a receptor surface model which is derived from one or more structures. The first phase of the search screens the database, using shape-based indices as screens, for compounds which are potentially similar in shape. The second phase aligns the candidate structures and evaluates shape similarity based upon a grid-volume comparison and optional volume overlap optimization. The third phase flexibly fits candidate structures to the surface model and evaluates steric strain and electrostatic complementary. Conformational flexibility is first addressed by representing compounds in the shape database as a set of diverse conformers and later by the flexible fitting process.

INTRODUCTION

Shape information is central to a number of approaches in molecular design, for example, QSAR analysis[1−3] and three-dimensional database search.[4] The representation of shape is an important component of these methods. Molecular shape has been variously represented by fields,[5] geometrical points,[6] surfaces,[7,8] volumes,[9] indices,[10] and three-dimensional topology.[11,12]

We have previously described a novel method for representing the shape and electronic characteristics of a putative receptor site using a surface representation, called a *receptor surface model* (RSM).[13] The model is constructed from one or more molecules believed to bind at a site and can be used to evaluate and guide the construction of new candidate molecules. Further work has demonstrated the utility of this approach in QSAR analysis.[14] This work explores using a receptor surface model as a database query to search a database for hits that fit a particular query's shape. Such a method could be useful in a number of contexts including database screening, database mining, and combinatorial library diversity analysis.

A RSM can be used to evaluate a candidate molecule by fitting the molecule inside the model. The molecule being evaluated "feels" the surface and is flexibly fit into the surface, yielding a shape that is consistent with the model. The result of the evaluation is an estimate of the strain energy of the fitted molecule and an estimate of the interaction energy (VDW and electrostatic) between the molecule and the RSM. The evaluation process typically takes about 1 s per candidate (SGI R4400). Applying this process to a set of compounds in a database would result in a list of candidate molecules which optimally fit the model. However, at 1 s per compound, applying this process to large databases of tens or even hundreds of thousands of structures would be excessively time consuming.

Since shape is at least a prerequisite to binding, an attempt was made to see if RSM shape information could be used in a database search to rapidly screen for candidates with *similar* shape. Those candidates passing this rough shape similarity filter could then be evaluated with the fitting procedure for a more rigorous steric and electrostatic analysis.

Such a two-phase approach will work for large databases if the first phase (shape similarity screening) is both fast and significantly reduces the number of potential candidates. This screening approach is analogous to 2D substructure searches which use topological bit screens before undertaking the algorithmically time-consuming atom-by-atom comparison. It must be practical for databases of potentially millions of compounds.

A number of approaches to shape-based database searching have been described.[15−17] The following section describes how the RSM methodology has also been extended for shape-based searching.

METHODS

Multiconformation Database Creation. This approach to shape-based searching first requires the creation of a compound database containing multiple 3D conformations per compound. Clearly, any shape searching approach that does not take into account the flexibility of the target molecules in the database will miss many viable candidates. This problem is addressed by representing every molecule in a database as a set of conformers, selected for diversity in conformation space. The method for conformational sampling and insuring conformation coverage has been described elsewhere.[18−20] This method represents the conformational space of a molecule with a relatively small set of diverse conformers (commonly about 20−30).

Compounds and their associated conformations are stored in a Catalyst database.[21] The process of creating the database is time-consuming, because of the conformational analysis required for each molecule. Typically, about 10 000 compounds can be processed per CPU day (SGI Indigo R4400). After the compound database has been created, a *shape filter* database is created. The shape filter database contains information for rapidly screening the database for shape candidates. The shape filter database is constructed by retrieving each conformer from the compound database, computing a set of volume and shape indices, and storing these per-conformer shape indices in the filter database. Shape filter database creation is fast relative to database creation and typically takes less than 30 min per million conformations processed.

[⊗] Abstract published in *Advance ACS Abstracts,* January 1, 1997.

S0095-2338(96)00108-4 CCC: $14.00

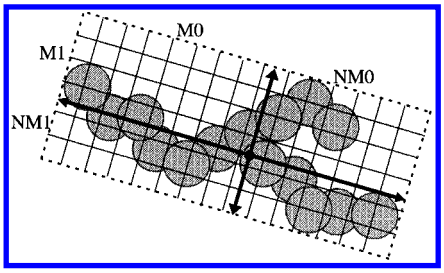

Figure 1. Illustration of the computed shape indices (in two dimensions only). The spheres represent the atoms of conformer or a query molecule(s). A volumetric grid enclosing the atoms is defined, and the interior grid points are determined. The geometric center of the interior points is found (shown a black circle). The principal axis are computed, and the maximum extents along the positive and negative directions are found. These extents are labeled M0 and NM0 for the first principal axis, M1 and NM1 for the second axis, and M2 and NM2 for the third principal axis (in three dimensions).

Shape Indices and Query Construction. The shape indices used to describe a conformer, and which are stored in the filter database, are the same indices that are used to characterize a shape query. This section describes how these indices are generated from either a conformer when building a database or from a query when searching the database.

A shape query is represented as an RSM and is constructed from one or more aligned structures. A surface enclosing the structures is computed and representes the VDW surface of the aggregate. The surface encloses a defined volume, which is represented as a grid (0.5−1.0 Å spacing). The surface is represented as a set of points of uniform average density. Both the surface points and the interior volume grid points represent the shape of the query. One can either use the surface points or the volume grid points to define the shape indices. We have found that both surface points and grid points give similar shape indices. These shape indices are derived from either set of points as follows.

First, the geometric center and three principal component vectors of the set of points are computed. No special weighting (either VDW radius or atomic mass) is used in the centroid calculation. Next, the maximum extents along each principal axis are found. This is illustrated for a set of volume grid points in Figure 1 (for two dimensions). The arrows indicate the principal axis. The arrows cross at the geometric center of the points. The bounding box represents the maximum extents along the positive and negative principal axes. M0 and NM0 are the extent lengths along the positive (longest) and negative (shortest) direction of the major axis, respectively. M1 and NM1 are the positive and negative extents along the minor axis. In three dimensions, the third axis contains M2 and NM2 components. In addition to these six indices, the total volume of the query (or conformer) is computed from the total number of surface interior grid points and the grid resolution. These seven indices are stored per-conformer in the shape filter database when constructing the database. The same indices generated for a query are used in the screening process. The indices provide a simple and compact way of representing the gross overall size and shape of a query.

Database Screening. The database screening process for a given query is as follows. The volume and six shape indices are computed for the query. These indices are then compared with the corresponding indices for each conformation in the shape filter database. The filter database is actually sorted on the first index so that only a subset of the indices need be compared. This process quickly eliminates conformations that do not have similar shape, as defined by these indices. A user-settable tolerance on the indices defines what is possibly "similar". This tolerance specifies the plus and minus variation allowed for the extents and volume indices. For example, a tolerance on extents of +10% to −30% will retain molecules that are bigger by up to 10% and molecules which are smaller by up to 30%. This will tend to bias the search to molecules which are smaller than the query. Likewise, the plus and minus variation on the volume can be specified. Default values of ±20% on the size extents and the volume are typically used.

The tolerances can also be dynamically scaled at run time so that no more than a user-defined percentage of the database will pass the screen.

Volumetric Shape Comparison. The database screening phase results in a list of candidate conformations that have shape indices similar to the query. These candidates are retrieved from the compound database. The query and candidate structures are aligned based upon their principal axis (the details of the alignment process are described later). The grid volumes of the query and target are then compared to determine shape similarity using a Tanimoto score estimate similarity.

The Tanimoto score is the *intersection* divided by *union* volumes of the query and target. A score threshold can be specified prior to search, to pass only hits that lie above the score threshold. This threshold acts as a secondary screen to the indices-based screen.

It is possible to calculate the Tanimoto score such that target volume that lies outside the query volume is more heavily weighted than query volume that lies outside the target volume. This penalizes excluded volume and can be important if one wants to favor molecules that tend to fit inside the query shape over molecules that are similar. The weighted Tanimoto is given in eq 1:

$$\text{Tanimoto similarity} = Q \,\&\, T/(w_1(Q - T) + w_2(T - Q) + Q \,\&\, T) \quad (1)$$

In eq 1, Q and T are the Query and Target volumes, respectively. $(Q - T)$ is the query volume that is not contained in the target. $(T - Q)$ is the target volume not contained in the query. w_1 and w_2 are weighting factors and are set to 1.0 by default. If w_2 is larger than w_1, then target volume sitting outside of query volume is penalized, resulting in a lower score for targets with volumes outside the query volume.

One limitation of this approach is that any molecule fitting entirely inside a query will give the same similarity score no matter how it is aligned. A volumetric comparison technique which attempts to deal with this problem has recently been described.[9]

Alignment and Symmetry Equivalence. Calculating the Tanimoto score requires the alignment of query and target volumes. This alignment is accomplished by translating the target center onto the query center and rotating the target such that major (M0) and minor (M1) axis of the query and target are aligned. The grid volume of the target is calculated and compared with the precalculated query volume for Tanimoto similarity.

Table 1. Shape Indices for the Antimycin Analog and Methotrexate Molecules Used as Shape Queries[a]

molecule	M0	NM0	M1	NM1	M2	NM2	VVol
ant. analog	15.0	9.5	6.4	5.2	3.4	2.8	476
methotrexate	9.2	7.8	6.9	6.1	4.4	4.1	471

[a] Axis indices and volume are in units of Angstroms and Angstroms^3 respectively.

Clearly, if the query or target molecule have any symmetry or near-symmetries, aligning on only the principal axis may not be adequate. The six values indicating the extents along the positive and negative directions of the three principal axis (M0, NM0, M1, NM1, M2, NM2) are used to test for possible symmetries of the molecule. The positive and negative extents are compared, and, if similar, the symmetry equivalent alignments are also compared volumetrically. For example, if the M0 and NM0 extent values are similar, the molecule may have *C*2 symmetry along one of the other two orthogonal axis. If M2 and NM2 are also similar, then the molecule can be rotated 180 degrees about the M1 axis. A "spherical" molecule with all extents equal requires 24 rotational alignments and comparisons. After trying all symmetry-equivalent permutations, the alignment yielding the best volume similarity is retained.

Volume Optimization. In an effort to test the quality of principal axis-based alignment, a descent optimization algorithm was added that attempts to improve the volume overlap of axis-based alignment. This iterative procedure tries a set of rotational and translational realignment adjustments and chooses the adjustment which best optimizes volume overlap. The same set of adjustments is applied again, iteratively, choosing adjustments until a better move cannot be made. The move set consists of six translations and six rotations (three positive and three negative) about the geometric center of the molecule. Testing indicates that a translational step of 0.5 Å and a rotational step of 5 degrees works reasonably well (results for this procedure are described later). This procedure converges to a local minima in a small number of steps.

After volume comparison, the hits that pass the threshold test are then ranked by similarity score. The hit list, sorted by similarity, can be saved and browsed, or can be passed on for the final phase of the search procedure.

Flexible Fitting. After the screening and volume comparison, a set of candidate structures that are shape-similar to a query remain. The candidates have similar total volumes, similar overall lengths, and similar shape as defined by Tanimoto volume similarity. Three-dimensional features of the query (i.e., h-bonding, hydrophobic, and charged groups) have not been taken into account during the search, and so each hit may or may not have electrostatic similarity to the query. Shape-only matches may be useful when looking for molecules or classes of molecules which share spatial but not necessarily topological or electrostatic similarity with another class. If one is interested in only shape-similar compounds, the hits, sorted by Tanimoto similarity, can be browsed or written to a file.

To ensure that the hits are both sterically and electrostatically consistent with the query, the hits are flexibly fitted into the query RSM. This evaluation procedure minimizes each hit into the RSM, flexibly-fitting each geometry to be consistent with the shape and electrostatics of the model.[13]

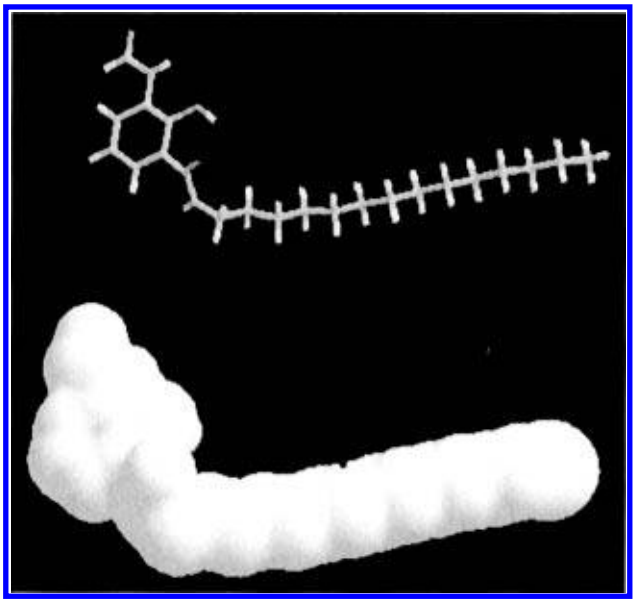

Figure 2. Antimycin analog and representative shape query constructed from the analog.

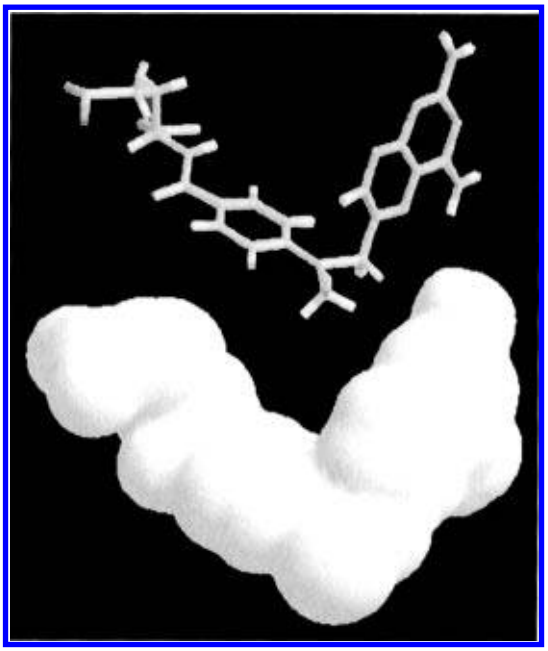

Figure 3. Methotrexate geometry and representative shape query constructed from the molecule.

The evaluation procedure estimates both intramolecular strain energy and intermolecular interaction energy between the hit and the surface model. This strain energy is the difference between the surface-minimized conformational energy and the conformational energy of the nearest local minima. The interaction energy is the sum of the VDW and electrostatic nonbonded energy between the molecule and the surface.

To arrive at a final set of shape matches, the evaluated structures are sorted by strain energy and all structures with a strain energy greater than a specified threshold are discarded. The default threshold is 20 kcal/mol. To measure electrostatic similarity, the remaining candidate list can be resorted on increasing interaction energy.

RESULTS

Two shape screen databases were created from commercially available compound databases. The first was

Table 2. Search Results for the Two Queries over the WDI and ACD Databases[a]

search query/DB	search type	search range %		% screened	# hits	best score	% first axis best	time h:min
antimycin/WDI	fast	+20	−20	98.0	1393	0.70	67.7	0:05
antimycin/WDI	best	+20	−20	98.0	2643	0.75	58.8	0:36
antimycin/ACD	fast	+20	−20	98.3	4460	0.77	78.1	0:10
antimycin/ACD	best	+20	−20	98.3	7172	0.78	68.5	1:19
antimycin/WDI	fast	+10	−10	99.8	563	0.67	67.1	0:02
antimycin/WDI	best	+10	−10	99.8	862	0.75	59.1	0:05
antimycin/ACD	fast	+10	−10	99.7	2252	0.76	80.0	0:03
antimycin/ACD	fast	+10	−10	99.7	2974	0.78	70.7	0:15
mtx/WDI	fast	+20	−20	79.3	6149	0.65	47.1	0:27
mtx/WDI	best	+20	−20	79.3	11042	0.69	47.2	6:04
mtx/ACD	fast	+20	−20	89.8	8983	0.66	48.3	0:37
mtx/ACD	best	+20	−20	89.8	20636	0.69	48.7	8:31
mtx/WDI	fast	+10	−10	94.7	3185	0.64	52.0	0:18
mtx/WDI	best	+10	−10	94.7	6309	0.69	50.0	1:45
mtx/ACD	fast	+10	−10	99.8	3915	0.66	54.0	0:24
mtx/ACD	best	+10	−10	99.8	9665	0.69	52.5	2:08

[a] Column 1 indicates the query and the database searched. Columns 2 and 3 show the parameters of the search: the search type; and the allowed range on extents and volume (the threshold Tanimoto volume similarity was set to 0.50). The remaining columns (4−8) show the search results: percentage of database screened out by the indices; the number of compound hits; the best Tanimoto score found for all hits; the percente of time the first principal axis yields alginment; the elapsed time of the search (200 MHZ R4400 Indigo).

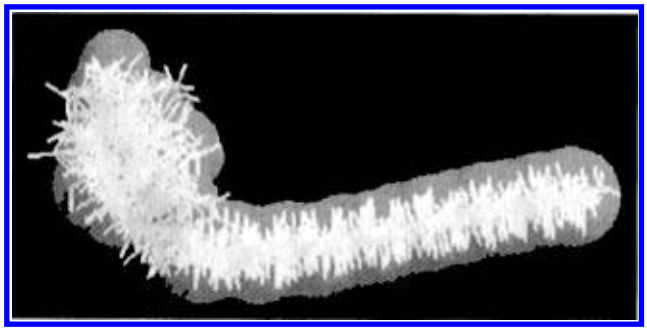

Figure 4. 20 hits from ACD with highest Tanimoto score aligned on shape query. This is the alignment produced by fast search without volume optimization.

Figure 5. The 20 hits from ACD after flexible fitting into query.

Figure 6. The 20 hits from ACD after relaxation minimization. Most of the hits remain entirely within the query.

constructed from the ACD[22] database, stored in Catalyst multiple-conformation format. This database consists of 166 705 compounds and 5 528 677 total conformations. The average volume of the compounds is 337.9 Å.[3] The average principal axis lengths (sum of positive and negative extents) for the molecules are 14.5, 9.5, and 7.1 Å. The second database was the Derwent World Drug Index (WDI).[23] This database contains 45 579 compounds and 1 949 459 total conformations. The average volume is 516.7 Å^3, with average lengths of 17.3, 11.4, and 8.5 Å. As might be expected, the average size of the drug molecules in Derwent is larger than the average size of chemicals in ACD.

Search results over both databases are illustrated with two molecules: an antifilarial antimycin analog from the Selwood et al. dataset[24] and methotrexate. For the antimycin analog, a low energy extended conformation was used to generate the shape query. For methotrexate, the DHFR bound conformation[25] was selected. These two molecules have similar volumes but significantly different shape indices. The antimycin analog is relatively asymmetric, while methotrexate is symmetric, as defined by these indices. These differences will help illustrate the strengths and limitations of this approach. Table 1 shows the shape indices for the two molecules.

Figure 2 shows the antimycin analog geometry and the shape query constructed from the molecule. Figure 3 shows the methotrexate geometry and the associated query.

For the searches, the allowed ranges on the extents were varied between ±20% and ±ten%. The Tanimoto score threshold was set to 0.5. Searches with volume optimization turned both off and on were performed. Searching with no volume optimization is termed a *fast* search, while a *best* search includes the volume optimization step. Table 2 shows the results of the searches.

Antimycin Analog Query. Using the antimycin analog query, a fast search was performed over WDI with tolerances of ±20% on extents and volume. This tolerance produced 94 457 candidate conformations, which represents a 98% screening rate of the full database of 1 949 459 conformations. Each of these candidate conformations was volumetrically compared with the query, and all candidates with similarity less than the threshold of 0.5 were discarded, resulting in 1393 compound hits. (If a compound was

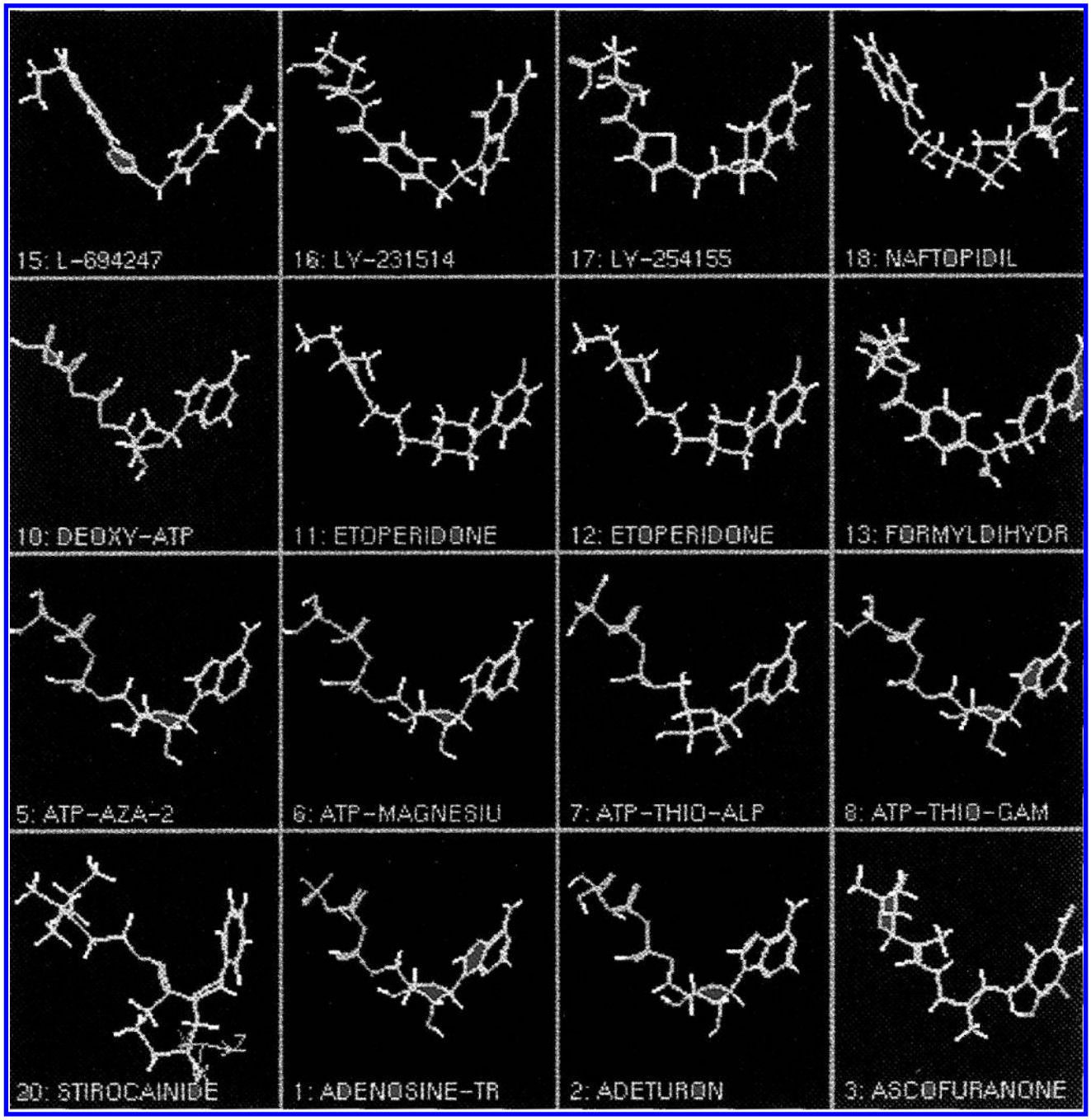

Figure 7. Twenty hits from WDI database with the highest Tanimoto score searching for methotrexate.

represented by more than one conformation, only the best scoring conformation was retained). The best Tanimoto score found was 0.70. (Recall that for best searching, multiple orientations of a conformation are volume compared, depending on the symmetry of the shape indices.) The second to last column in Table 2 shows the percentage of time the first principal axis alignment yields the best volumetric overlap. For the antimycin analog query, the first alignment is the best alignment between 58 and 80% of the time. The total elapsed time for the fast search was 5 min on a 200 MHZ Indigo R4400. The elapsed time for the best search was 36 min.

Figure 4 shows the top 20 hits returned from the fast search, oriented by axis alignment. Most of the hits fit well within the shape of the query, but there are portions of several of the molecules which lie outside the surface. Each of the molecules was then flexibly fit to the surface model. Figure 5 shows the 20 hits after flexible fitting to the surface. Note that now, all molecules lie within the surface. Some of the molecules have internal strain due to the fitting process. To alleviate the strain, the molecules are reminimized without the presence of the query surface. Figure 6 shows the hits after relaxation minimization. The relaxed structures still fit well within the surface and are at a close local minima to the flexibly fit structures.

For a given Tanimoto score threshold, fast search and best search (which includes volume optimization) will produce a different set of hits. To test if the differences are significant, a fast and best search of WDI and ACD was performed, keeping a hit list of the top 200 compounds. These lists were then compared to see the amount of overlap between them. For WDI, 128 of the 200 compounds were found in both lists. For ACD, 109 were common to both. The best similarity found was always higher with volume

optimization, indicating that volume optimization provides additional improvement. However, the search times with volume optimization are about 3 to 12 times slower than fast search (which does not perform volume optimization).

Tightening the allowed range on the extents improves the screening ratio and decreases search times. Table 2 shows the effect of tightening the search range from 20% to 10%. For the antimycin analog these tighter tolerances lead to a screening of over 99.7% for both ACD and WDI. To test the sensitivity of the tolerances a search with tolerances of 10, 20, and 30% was performed over WDI and ACD, keeping a hit list of the top 200 compounds. These lists were compared to determine the amount of overlap between them. For WDI, comparing tolerances of 20 and 10%, 118 of the 200 compounds were found in both lists. For ACD, 105 were common to both. Comparing the tolerances of 20−30% found only an additional five compounds for WDI and seven compounds for ACD but increased the search time by about a factor of 3. This indicates that a tolerance of 20% is a reasonable compromise between search speed and search coverage.

Methotrexate Query. Methotrexate is a more symmetrical molecule than the antimycin analog, as defined by the shape indices (see Table 1). Further, the indices are very similar to the average indices of the conformations in WDI. Therefore, the screening ratio is poorer. For WDI, about 80% of the database is screened out. For ACD, the screening ratio with the methotrexate indices is also poorer than with the antimycin indices. About 90% is screened out for ACD, when tolerances of 20% are used. When the tolerances are tightened to 10%, about 95% of the WDI and over 99% of the ACD are screened out. Search times are about an order of magnitude slower than for the antimycin analog. This is due to two factors. First, since the screening ratio is poorer, more candidates are evaluated volumetrically (about four times as many as for the antimycin analog). Second, since the molecule is more symmetrical, more symmetry-equivalent orientations are tried. This is shown in the second to last column in Table 2 which lists the percentage of time the first principal axis alignment yields the best volumetric score. These values drop by 10−20% over the corresponding antimycin values. Because of the more spherical shape of methotrexate, alignment on principal axis only is less likely to be the optimal alignment.

Figure 7 shows the top 20 hits returned by the fast search of methotrexate against WDI. Figure 8 shows the hits inside the query, after the flexibly fitting and relaxation minimization. While methotrexate is contained in the WDI database, methotrexate was not found in the top 20 hits, rated by Tanimoto score, when searching this database (Deazamethotrexate is found.). This indicates that the bound crystal conformation of methotrexate is not generated by the conformational analysis program used to populate the database. However, methotrexate and a number of methotrexate analogs are still found within the top 100 structures, indicating that a conformation close to the bound crystal structure is produced for these structures. This highlights the fact that this methodology can be sensitive to the quality and coverage of the conformers provided by the conformation generation algorithm.

The percentage of time that the methotrexate first alignment is the best alignment is also lower than for the antimycin analog. This is expected given the more symmetric shape of the methotrexate; therefore, structure alignment on principal axis only is less likely to be optimal. The best search times are significantly slower for methoxtrexate than for the antimycin analog over both databases. Because of the symmetry of the molecule, the number of steps for volume optimization convergence is, on average, higher for methotrexate.

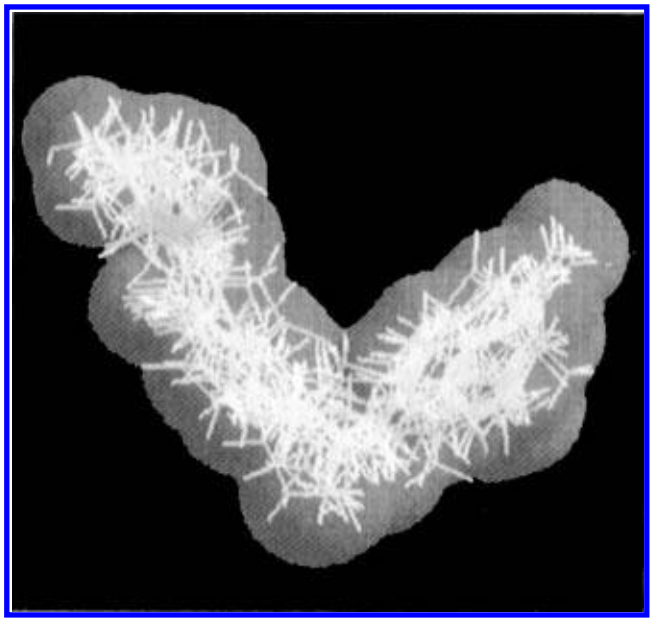

Figure 8. Twenty hits from WDI, shown after flexible fitting into the query and then relaxation minimization.

CONCLUSIONS

This paper has described a method for shape-based searching that is rapid and general. The three-phase search strategy allows the searching of large databases (>100 000 compounds) in reasonable time. The shape indices are compact. The use of principal axis magnitudes provides good screening capability. The use of the principal axis for volume alignment gives reasonable results, though optimizing the volume after principal axis alignment provides improvement. The problem of conformation flexibility is addressed by both the use of multiconformation databases and the flexible fitting of candidates against the RSM query. However, as was seen with methotrexate, this method is sensitive to the quality and coverage of the conformational models stored for each compound in the database. One potential way to deal with this problem is to apply the flexible fitting process, allow the fit structures to relax, and then recompute and reorder the hits based upon the new volume similarity.

This approach, as with any grid-based method, is also sensitive to the choice of an appropriate grid spacing. In this case, volume similarity is sensitive to the granularity of the volume. A small grid spacing increases the number of points by a factor of n^3 and slows the search speed, while a large grid spacing can lose spatial resolution. This work and others[9] indicates that a grid resolution of 1.0 Å provides a reasonable compromise between speed and quality.

The fast screening phase of the search procedure uses only shape information. One attractive area for future extension of the approach is to add to the screening procedure electrostatic information. The charge distribution of the

molecule could be characterized, for example, by molecular moments,[26] and these could be added as additional indices.

The shape indices can be used to characterize the 3D shape of molecules. By taking averages and ranges of the shape indices of all conformations for a given compound, whole molecule descriptors can be derived which represent shape and size variability. Such descriptors should be useful in diversity and similarity analysis.

ACKNOWLEDGMENT

The author would like to thank Dr. David Rogers and Dr. Marvin Waldman for their timely and insightful contributions in the preparation of this manuscript. The methods described in this paper are implemented and available as part of MSI's Catalyst and Cerius[2] Drug Discovery Workbench and Diversity products.

REFERENCES AND NOTES

(1) Cramer, R.; Patterson, D.; Bunce, J.; Comparative Molecular Field Analysis (CoMFA) 1. Effect of shape on Binding of Steroids to Carrier Proteins. *J. Med. Chem.* **1988**, *110*, 5959−5967.

(2) Burke, B.; Hopfinger, A.; Molecular Shape Analysis: A Formalism to Quantitatively Establish Spatial Molecular Similarity. In *Concepts and Applications of Molecular Similarity*; Johnson, M., Maggiora, G., Eds.; John Wiley and Sons: New York, 1990; pp 173−209.

(3) Doweyko, A. The Hypothetical Active Site Lattice. An approach to modeling active sites from data on inhibitor molecules. *J. Med. Chem.* **1988**, *31*, 1396−1406.

(4) Martin, Y. 3D Database Searching in Drug Design. *J. Med. Chem.* **1992**, *35*, 2145−2154.

(5) Grant, J.; Pickup, B. A Gaussian Description of Molecular Shape. *J. Phys. Chem.* **1995**, *99*, 3503−3510.

(6) Connolly, M. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.

(7) Masek, B.; Marchant, A.; Matthew, J. Molecular Skins: A New Concept for Quantitative Shape Matching of a Protein with its Small Molecule Mimics. *Proteins* **1993**, *17*, 193−202. Masek, B.; Marchant, A.; Matthew, J. Molecular Shape Comparison of Antiotensin II Antagonists. *J. Med. Chem. Proteins* **1993**, *36*, 1230−1238.

(8) Bohacek, R.; McMartin, C. Definition and Display of Steric, Hydrophobic, and Hydrogen-Bonding Properties of Ligand Binding Sites in Proteins Using Lee and Richards Accessible Surface: Validation of a High Resolution Graphical Tool for Drug Design. *J. Med. Chem.* **1992**, *35*, 1671−1684.

(9) Perkins, T.; Mills, J.; Dean, P. Molecular surface-volume and property matching to superpose flexible dissimilar molecules. *J. Comput.-Aided Mol. Design* **1995**, *9*, 479−490.

(10) Todeschini, R.; Lasagni, M.; Marengo, E. New Molecular Descriptors for 2D and 3D Structures. Theory. *J. Chemometrics* **1994**, *8*, 263−272.

(11) Mezey, P. Three-Dimensional Topological Aspects of Molecular Similarity. In *Concepts and Applications of Molecular Similarity*; Johnson, M., Maggiora, G., Eds.; John Wiley and Sons: New York, 1990; pp 321−368.

(12) Mezey, P. *Shape In Chemistry*; VCH: New York, 1993.

(13) Hahn, M. Receptor Surface Models. 1. Definition and Construction. *J. Med. Chem.* **1995**, *38*, 2080−2090.

(14) Hahn, M.; Rogers, D. Receptor Surface Models. 2. Application to Quantitative Structure−Activity Relationships Studies. *J. Med. Chem.* **1995**, *38*, 2091−2102.

(15) Good, A.; Ewing, T.; Gschwend, D.; Kuntz, I.; New Molecular Shape Descriptors: Application in Database Screening. *J. Comput.-Aided Mol. Design.* **1995**, *9*, 1−12.

(16) van Geerstein, V.; Perry, N.; Grootenhuis, P.; Haasnoot, C. 3D Database Searching on the Basis of Ligand Shape Using the SPERM Prototype Method. *Tetrahedron Comput. Methodology* **1990**, *3*, 595−613.

(17) Wild, D.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures. Alignment of Molecular Electrostatic Potential Fields with a Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 159−167.

(18) Smellie, A.; Teig, S.; Towbin, P. *J. Comput. Chem.* **1995**, *16*, 171.

(19) Smellie, A.; Kahn, S.; Teig, S. An Analysis of Conformational Coverage 1. Validation and Estimation of Coverage. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285−294.

(20) Smellie, A.; Kahn, S.; Teig, S. An Analysis of Conformational Coverage 2. Applications of Conformational Models. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285−294.

(21) Catalyst is distributed by Molecular Simulations Inc., San Diego, CA, U.S.A.

(22) ACD database is available from MDL Informations Systems Inc., San Leandro, CA, U.S.A.

(23) WDI database is available from Derwent Information Limited, London, England.

(24) Selwood, D.; Livingstone, D.; Comley, J.; O'Dowd, A.; Hudson, A.; Jandu, K.; Rose, V.; Stables, J. Structure-Activity Relationships of Antifilarial Antimycin Analogues: A Multivariate Pattern Recognition Study. *J. Med. Chem.* **1990**, *33*, 136−142.

(25) The PDB file used to obtain the methotrexate molecule geometry was 4DFR.brk.

(26) Silverman, B.; Platt, D. Comparative Molecular Moment Analysis (CoMMA): 3D-QSAR without Molecular Superposition. *J. Med. Chem.* **1996**, *39*, 2129−2140.

CI960108R