# Toward General Methods of Targeted Library Design: Topomer Shape Similarity Searching with Diverse Structures as Queries

Katherine M. Andrews and Richard D. Cramer*

*Tripos, Inc., 1699 South Hanley Road, St. Louis, Missouri 63144*

*Received January 3, 2000*

A promising strategy for selecting synthetic targets is similarity-based searching of very large "virtual libraries", which comprise all structures accessible by linking two or three commercially available building blocks with combinatorial syntheses. To assess the general applicability of this strategy, leading structures taken from each of 34 recent medicinal chemistry publications were used as queries to search a virtual library containing $2.6 \times 10^{13}$ products from seven reactions, using a topomer shape similarity metric. Eighty-five percent of these searches succeeded, by yielding, with a search radius no greater than 120 topomer shape units, either at least 400 hits or hits from at least six sublibraries. From these 34 sets of search results, 122 representative structures were selected, illustrating potential "lead hops", or otherwise novel structures. Overall shape similarity to the query structure was confirmed for up to 95% of these representative structures, according to FLEXS, an algorithmically distinct program. Experimentally, there were 28 structures among those reported in the 34 query publications that were identified within the virtual library. Among these, the frequency of high activity was 87% for the 16 structures whose similarity to their query was 90 topomer units or less, compared to a frequency of 50% for the other 12 structures.

## Introduction

Success in drug discovery continues to depend greatly on effectiveness in selecting the next compound(s) to test, even as high throughput methods increase the rates and volumes of testing. Traditionally the selection domain consists of the structures that an individual chemist can envision synthesizing, perhaps augmented by computational selections among inventories of compounds at hand or commercially available. Most would agree that the selection mechanism dominating individual chemists' "intuition" is structural similarity.[1] Perhaps it is not as widely appreciated that most formal compound selection methods also depend on similarity assumptions, though usually unstated. For example, when a computational model derived from previous test results or a receptor structure suggests a change that could improve the properties of a series of molecules, this change will usually be local, expected to be effective only within an otherwise similar structure.

Thus it is understandable that a particular challenge in compound selection is to "lead hop", that is, to select, from the vastness of structurally dissimilar molecules, structures most likely to have the same critical biological property as some compound of interest. This challenge has motivated creation of such powerful computational methodologies as "docking"[2] (of available structures into receptor site models) and "pharmacophoric searching"[3] (among available structures for those which can present a particular geometric relation among elements thought critical for biological activity). While these methodologies have had their successes, their emphasis on conformational flexibility does introduce a problem dimension so enormous as to computationally limit the practical size of their selection domains to perhaps $10^7$ candidate structures.[4]

Whether or not "lead-hopping" is an explicit goal, we can imagine the ideal or optimal compound selection process. Its domain would comprise "all readily synthesizable structures" and that domain would be searched by a technique capable of rapidly identifying a small number of structures most likely to have some desirable biological activity. Topomer shape similarity searching of ChemSpace virtual libraries is a reasonable approximation to this ideal.[5] Its search domain is the particular subset of "all readily synthesizable structures" that is most accessible to combinatorial chemistry. As a searching method, topomer shape has been found in retrospective studies to be the molecular descriptor whose similarity was most consistently and strongly related to similarity in biological behavior.[6] Topomer shape similarity then performed very well in an unambiguously prospective experiment, for example by indicating the 15% of a combinatorial array of prospective angiotensin II antagonists that proved to contain all the structures of highest potency[7] and by making useful contributions to rapid discovery of new serotonin antagonist classes.[8]

These recent practical validations of the ChemSpace virtual library technologies encouraged us to explore topomer searching as a very general method for targeted library design, by trying to answer questions such as the following. Will searches of a virtual library yield hit[9] structures for most lead queries? Will these structures tend to be novel, unlikely to have been considered otherwise, but yet plausibly active? Will they indeed be synthetically accessible? To carry out this exploration in as realistic a way as possible, we built an unusually large virtual library and performed topomer shape

* To whom correspondence should be addressed. Phone: 314-647-1099. Fax: 314-647-9241. E-mail: cramer@tripos.com.

10.1021/jm000003m CCC: $19.00 
Published on Web 04/06/2000

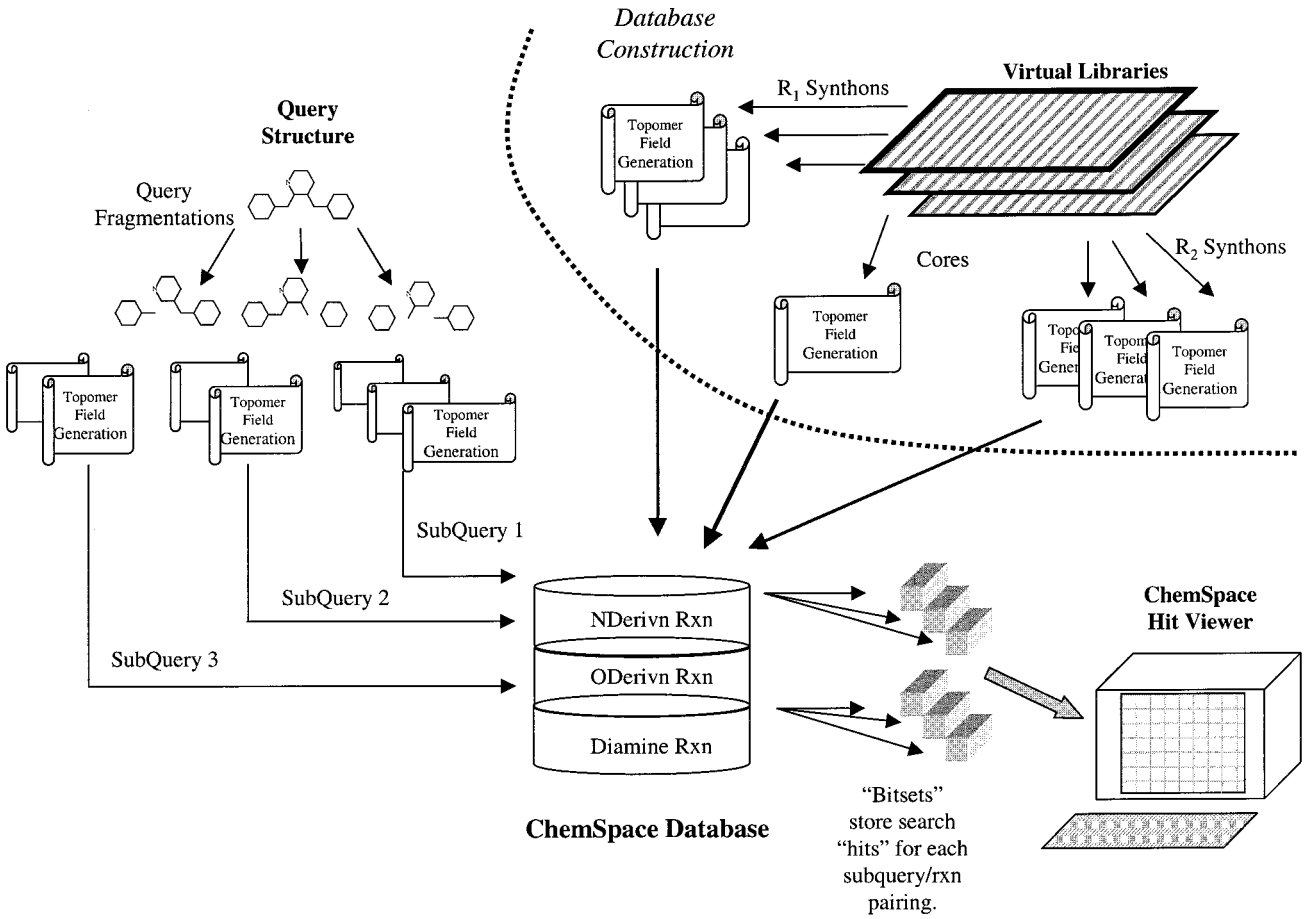



**Figure 1.** Flow diagram for topomer shape similarity searching. The query structure is partitioned into various sets of fragments, and the topomer conformation and steric field are calculated for each fragment. The fields are compared with the stored fields for fragments that are synthons in a virtual library. Products whose corresponding fields are similar enough to those of a query are stored within "bitsets" (see ref 11 for details). Complete structures are assembled only for viewing or export.

similarity searches, using diverse structures from recent publications as examples of medicinally interesting queries. To provide objective evidence of the relevance of the structures retrieved, we assessed the shape similarity of some of these hits by independent and much more rigorous computations, using the FLEXS program.[10] We have also considered the other structure/activity data within most of these publications, as the only experimental data available, in order to demonstrate that compounds already known to be potent are preferentially selected by this methodology.

## Experimental Section

Most studies were carried out using ChemSpace,[11] a large proprietary system developed for building and searching virtual libraries, which uses SYBYL, Unity, Oracle, and CONCORD as key modules, linked together by Java, C, SPL, and Unix shell scripts.

**Overview of Topomer Shape Searching.** This approach differs from other shape comparison methods in several fundamental ways:

- Topomer shape comparisons consider all atoms in a structure, in contrast to the pharmacophore-based "3D searching" approaches widely used, where "shape" comparison focuses on a small subset of atom-like features.
- Shapes are compared as a combination of fragment-to-fragment differences, rather than by performing operations on complete structures.
- Shape comparisons involve only one rule-derived "topomer" *conformation* for each fragment, rather than some form of optimization among the indefinitely large variety of conformations that most fragments are capable of achieving.
- Shape comparisons involve only one rule-based *orientation*, specifically superposition of the attachment bond(s) necessarily associated with any fragment, again, rather than some optimization process. (For polyvalent fragments, shape comparisons are performed and averaged over all such attachment bonds, and the relative positions of attachment bonds contribute to the shape difference.)
- Shapes resemble one another mostly to the extent that the same spatial volume or "field" elements are occupied or avoided by the corresponding fragments (in their topomer conformations),[12] not to the extent that achievable geometries among assumed critical features are shared.
- Shape comparison optimization *is* performed over all possible fragment-to-fragment mappings, the result being considered as the minimum found among all the possible mappings.

The process of topomer searching is summarized in Figure 1, flowing from entry of the query structure at the upper left to inspection of the search results at the lower right. The upper right quadrant shows the construction of the database to be searched. The central calculation in topomer similarity searching, comparing the fields of topomer conformations, requires, as input, the representations of both the database and query structures as the fields (or other properties) of a set of connected fragments. Conversion of fragments into topomers and then fields is necessarily the same process for both database and query structures, as represented schematically in Figure 1 by the scroll icons labeled "Topomer Field Generation". However, when performing a search, explicit fragmentation is necessary only for the query structure, since the database structures within a virtual combinatorial library have

**Table 1.** Composition of the Research Virtual Library (RVL)

| Library Name | Site | Reagent Class Name | Reagent Query SLN[a] | # Synthons | #Var@ Site[b] | Library Size[c] |
|---|---|---|---|---|---|---|
| **1** N derivatives | | | | | | |
| | R1 | Amines | HcN[charge=0;is=N*C;not=N*C=#Any,N*Het]\(H){Hc:H,C} | 22460 | 22460 | |
| | R2 | Isocyanates | RN=C=O | 333 | | |
| | | Chloroformates | ClC(=O)OR | 76 | | |
| | | Epoxides | C[1]H2OC@1 | 325 | | |
| | | Aldehydes/ketones | HcC(=O)C[not=C=#Any]{Hc:H,C[not=C=#Any]} | 18602 | | |
| | | Acid chlorides | RC(=O)Cl | 638 | | |
| | | Carboxylic acids | HOC(=O)C | 17916 | | |
| | | Carboxylate salts | O[F]C(=O)C | 824 | | |
| | | Sulfonyl chlorides | ClS(=O)(=O)C | 242 | | |
| | | Activated aryl fluorides | FAny:Any:Any:AnyWdraw{Wdraw:Any=#Het} | 1024 | | |
| | | Azahalides | N:CHal | 2526 | | |
| | | Benzyl/allyl halides | HalCHC:=Any | 2740 | | |
| | | Amines (=> ureas) | HcN(H)C[not=C=#Any]{Hc:H,C[not=C=#Any]} | 23688 | | |
| | R2 sum | | | | 68934 | |
| | | | | | | 1.55E+09 |
| **2** O derivatives | | | | | | |
| | R1 | Alcohols/phenols | HOC[not=C=Any] | 21183 | 21183 | |
| | R2 | Isocyanates | RN=C=O | 333 | | |
| | | Chloroformates | ClC(=O)OR | 76 | | |
| | | Acid Chlorides | RC(=O)Cl | 638 | | |
| | | Thioisocyanates | RN=C=S | 476 | | |
| | | Activated aryl fluorides | FAny:Any:Any:AnyWdraw{Wdraw:Any=#Het} | 1024 | | |
| | | Azahalides | N:CHal | 2526 | | |
| | | Alkyl chlorides | ClCH | 2527 | | |
| | | Alkyl bromides | BrCH | 133 | | |
| | R2 sum | | | | 7733 | |
| | | | | | | 1.64E+08 |
| **3** S derivatives | | | | | | |
| | R1 | Thiols | HSC[not=C=Any] | 1640 | 1640 | |
| | R2 | (same as for O derivatives, excluding Chloroformates and Acid Chlorides but including Sulfonyl Chlorides) | | 8043 | 8043 | |
| | | | | | | 1.32E+07 |
| **4** Grignard | | | | | | |
| | R1 | Bromides | BrC:-C | 9669 | 9669 | |
| | R2 | Aldehydes/ketones | HcC(=O)C[not=C=#Any]{Hc:H,C[not=C=#Any]} | 17245 | | |
| | | Nitriles | N[F]#CC | 12182 | | |
| | R2 sum | | | | 29427 | |
| | | | | | | 2.85E+08 |
| **5** Diamines | | | | | | |
| | Core | Diamines | Hc-:N[is=N*C;not=N*C=#Any,N*Het](H)~X~N\[is=N*C;not=N*C=#Any,N*Het](H)-:Hc{Hc:H,C} | 4145 | 4145 | |
| | R1 | (same as R2 list for N derivatives) | | | 68934 | |
| | R2 | (same as R2 list for N derivatives) | | | 68934 | |
| | | | | | | 1.97E+13 |
| **6** Dihalides | | | | | | |
| | Core | Dihalides | HalCel~X~CelHal{Cel:C[is=C*:=Het,\C*H2,C*C=:Any,C*C=:Cany=#Het,C*C:C:C:Cany=#Het;not=HalC*Hal} | 198 | 198 | |
| | | Secondary N | CNHC | 18079 | | |
| | | Alcohols/phenols | HOC[not=C=Any] | 9486 | | |
| | | Thiols | HSC[not=C=Any] | 642 | | |
| | R1 sum | | | | 28207 | |
| | R2 | (same as R1 list for dihalides) | | | 28207 | |
| | | | | | | 1.58E+11 |
| **7** Amino Acids | | | | | | |
| | Core | Amino Acid | Hc-:N[is=N*C;not=N*C=#Any,N*Het](H)XC(=O)OH{Hc:H,C} | | 4200 | |
| | R1 | (same as R2 list for N derivatives) | | | 68934 | |
| | R2 | Amines | HcN(H)C[not=C=#Any]{Hc:H,C[not=C=#Any]} | | 22460 | |
| | | | | | | 6.50E+12 |
| **TOTAL Library Size** | | | | | | 2.64E+13 |

[a] Sybyl Line Notation (SLN) used to identify reagents having the named reactive functionality. Reagent searches also automatically excluded reagents containing more than one such functionality. (See ref 16.) [b] The total number of variations at a site (R1, R2, or core). When no SLN is given, this value is the sum of the #Synthons values. [c] The size of the library is calculated as the product of the values in the #Var@Site column.

already been fragmented and stored during the library construction process. The strategy for fragmenting query structures represents a compromise between search completeness and computational cost. As suggested by the three examples of query fragmentations in Figure 1, each additional fragmentation provides a distinct set of shapes possibly yielding very

different matching structures, but with the cost of performing another completely independent search of the database.

**Domain (Database) Construction.** Seven combinatorial reactions were chosen, as generalizations of various combinatorial syntheses in general use. Each of the corresponding virtual libraries was constructed in three stages. The first stage was substructure searching of the Available Chemicals Directory[13] (ACD) for all commercially offered reagents that contain exactly one (or one combination) of the reactive functionalities needed. Then each of the individual reagents found was "clipped" to form the corresponding synthon (the fragment that a reagent would contribute to any of its product structures). Finally various physical properties of each synthon were generated and stored, in particular the steric field of the topomeric conformation, the descriptor used in topomer shape searching. Details of the procedures for calculating the topomeric descriptor have been described previously.[5,7,11]

The exact composition of the resulting Research Virtual Library (RVL) is given in Table 1. There are four "two-piece" libraries (for example, sulfonamide formation, one of the 12 sublibraries within the "N derivatization" library) and three "three-piece" (or "scaffold-based") libraries (for example, amide linkages to both the carboxyl and amine moieties of an amino acid, represented by three of the 12 sublibraries within the "Amino Acid" library). Appropriate multiplications and additions yield the total number of structures represented by the RVL as $2.6 \times 10^{13}$ (about 26 trillion), surely one of the largest collections of "small molecule" structures ever assembled.[14] Every one of these structures is accessible in one or two formal synthetic steps,[15] by combining only those reagents that have recently been offered commercially.

**Selection of Query Structures.** Query structures were taken from publications appearing in three recent issues of the *Journal of Medicinal Chemistry*. Each article in those issues that described the synthesis and testing of novel structures contributed one such query structure (excepting linear oligomers such as peptides, which were deemed inappropriate queries for this RVL). The structure selected to be the query was either the one featured in the abstract or else the one reported to be most potent in the primary assay. The 34 query structures resulting are shown as the second column of Figure 3. The first column provides an identifier, the page number of the associated article. Additional information in Table 2 confirms the diversity of the biological targets, therapeutic potentials, and institutional origins for each of these 34 query structures.

**Fragmentation of the Query Structures.** The composition of the RVL suggested how the query structures should be fragmented. All possible acyclic single bonds were each broken (yielding two fragments), as were all possible pairings of acyclic single bonds (yielding three fragments). Further restrictions on the acyclic single bonds to be broken were as follows:

- The bond must connect two heavy atoms.
- At least one of those heavy atoms must either be in a ring, or be attached to a ring, or be attached to at least two other heavy atoms, or be a heteroatom.
- All fragments resulting must include at least two heavy atoms.
- These requirements were expressed as search patterns in Sybyl Line Notation (SLN)[16] and automatically applied to the 34 query structures. From one to over 50 fragmentation subqueries and searches resulted for the various individual query structures, depending on the number and locations of the appropriate acyclic bonds.

**Validation of the Topomer Shape Searching Process.** To assess the reliability of the searching software and the integrity of a database, our practice is to extensively "self-search" the database. From each of its sublibraries (194 within this RVL), a structure is taken randomly and used to query the entire database. A very small search radius (acceptable shape difference) of 5 units[17] ensures that only one hit—the query structure itself—should usually be retrieved. When no matching structure is found, a search has failed, so the cause of failure must be identified and, if practical, remedied. Over the last year, fixes to algorithms and to the stored fragment data have raised this self-search test success rate for the RVL from about 30% to over 90%. All results reported here were obtained with the most recent and robust versions of the ChemSpace technology and the RVL database.

**Selection of Search Radii.** Search radii (maximally acceptable shape differences between the query and a hit) always present a tradeoff. Larger radii yield increasing numbers of candidate structures, but each then is expected to be less likely to share the biological properties of the query structure than would hits from searches using smaller radii. In previous work,[6,7] statistically significant enhancements in the frequency of biological activity have been found with search radii of 90 and of 120 topomeric shape units. Also, for combinatorial synthesis, arrays of target structures are needed rather than "singletons", and many alternative arrays of targets are desirable since difficulties such as failed validations or unavailable reagents will make many routes unrealizable. However, searches are seldom performed at radii lower than 80 units because the number of structures retrieved is usually so small.

Therefore, for each query structure, the objective was to establish the smallest search radius value (divisible by 10) which yielded more than 1000 hits or produced hits from at least five sublibraries. The initial search radius was 120 shape units, and searches were then repeated, expanding or contracting the radius in steps of 10 or 20 units, until the desired outcome was obtained. The results of these 34 searches are summarized in Table 3.

**Inspection of the Hit Lists.** The quality of the individual hit structures will clearly be the most critical information when evaluating any process for structure selection, no matter how subjective any such assessment of quality may seem to be. Therefore, to illustrate each of the 34 hit lists, up to four individual structures were selected by hand, as shown in Figure 2, totaling 122 examples. Factors considered in making these selections were as follows:

- Are the structures unlikely to have been suggested by rationale other than shape similarity?
- Are the structures plausibly active, assuming only the knowledge that the query structure was active but no other SAR information? (Particular attention was given to matching any positive charge in the query structure.)
- Is the intended synthesis likely to succeed? (The bonds to be formed are pointed to by wedge bonds and the general synthetic route is indicated by the library name under each structure.)
- Are the synthetic routes, the underlying query fragmentations, and the building blocks all different for each of the hit structures shown?

The structures shown are divided into two groups, with the leftmost two columns showing the "more plausible" hits, particularly in their expected ease of synthesis, and the rightmost two columns showing "less plausible" but more structurally diverse, yet shape-similar, structures.

**Figure 2.** (On the following 4 pages.) Representative hit structures from each of 34 topomeric shape similarity searches. From left to right are given the page number of the article as a short identifier; the query structure; the search radius, in shape units, for which the structures shown were retrieved; and up to four hit structures. The "wedge bonds" point to the one or two bonds that would be formed by combinatorial synthesis. Appearing below each of the hit structures is one of seven library names, implying the route of synthesis, and the total number of other structures within the same sublibrary that are at least as shape-similar to the query. The presence of a footnote indicates a structure for which the best FLEXS overlay was not equivalent to the topomer similarity. For more information, see the text.

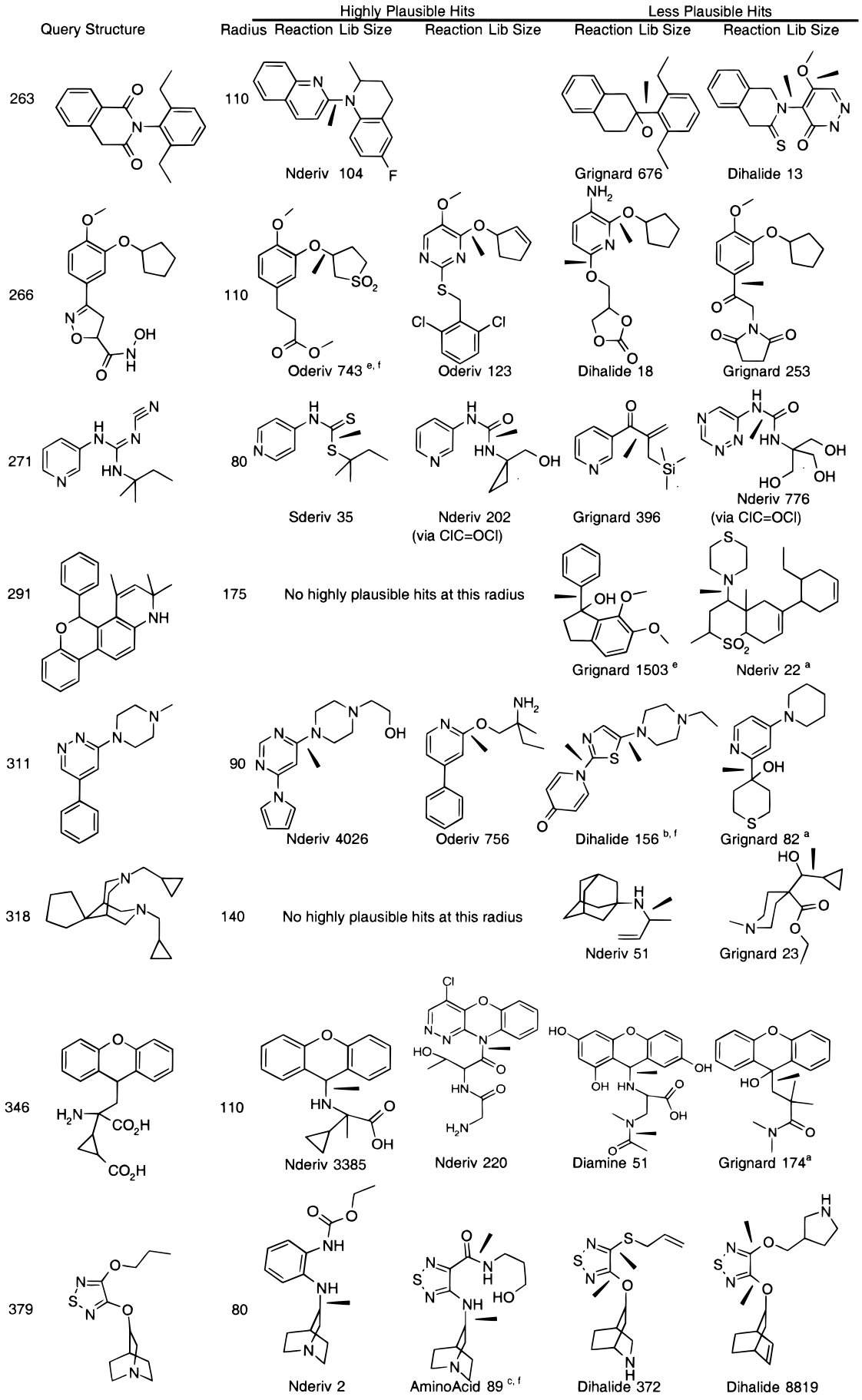



**Figure 2.** (Continued)

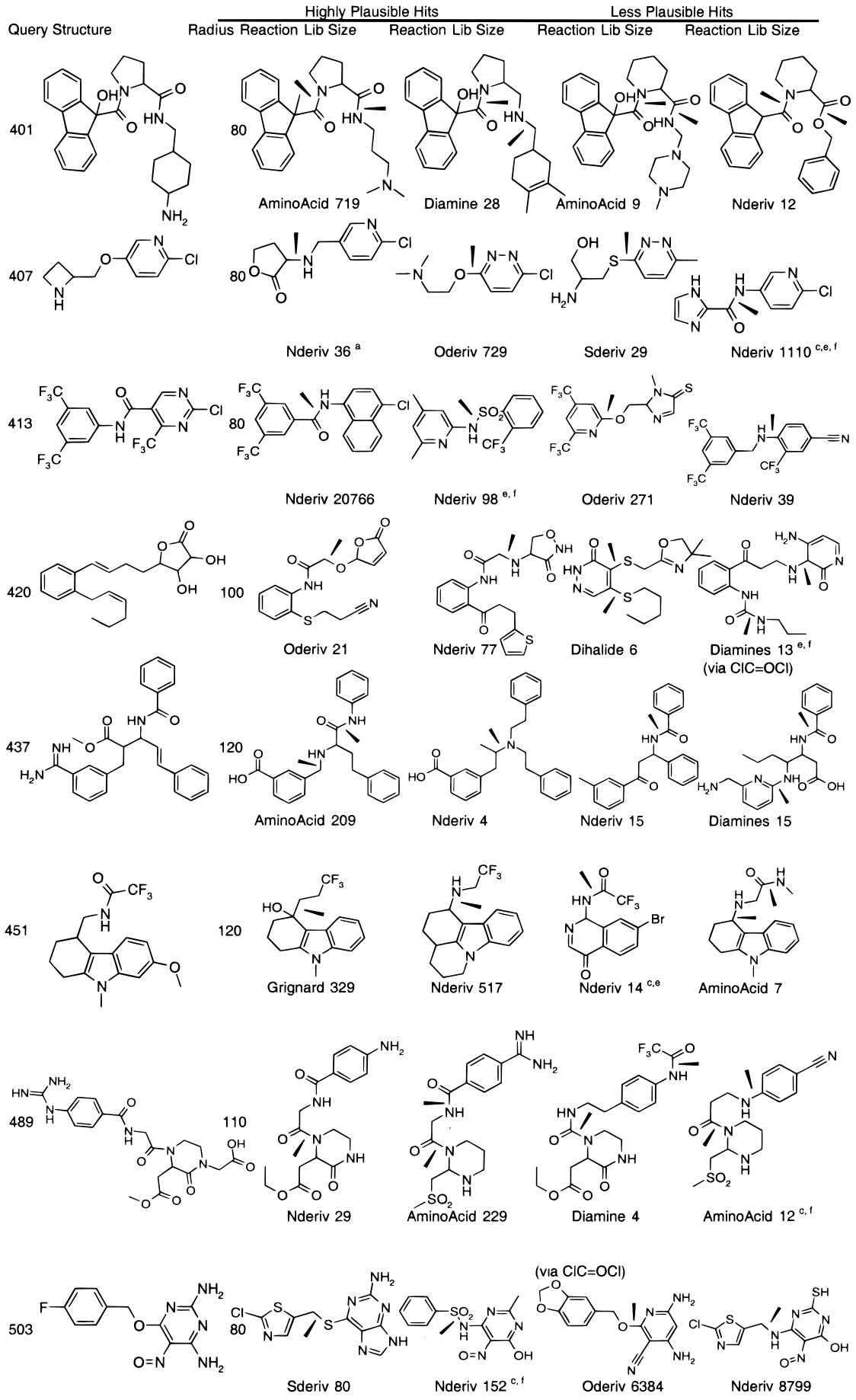



**Figure 2.** (Continued)

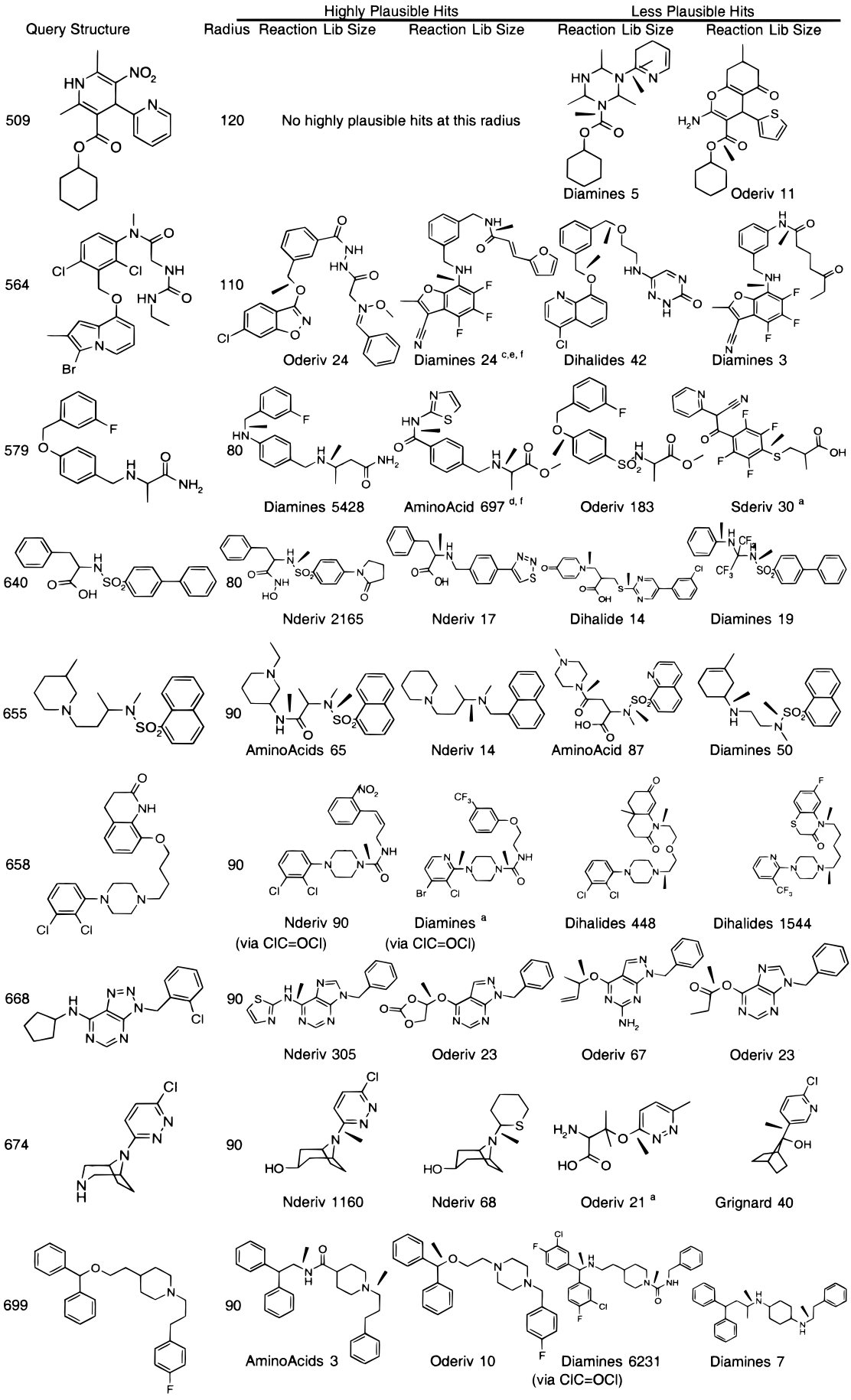
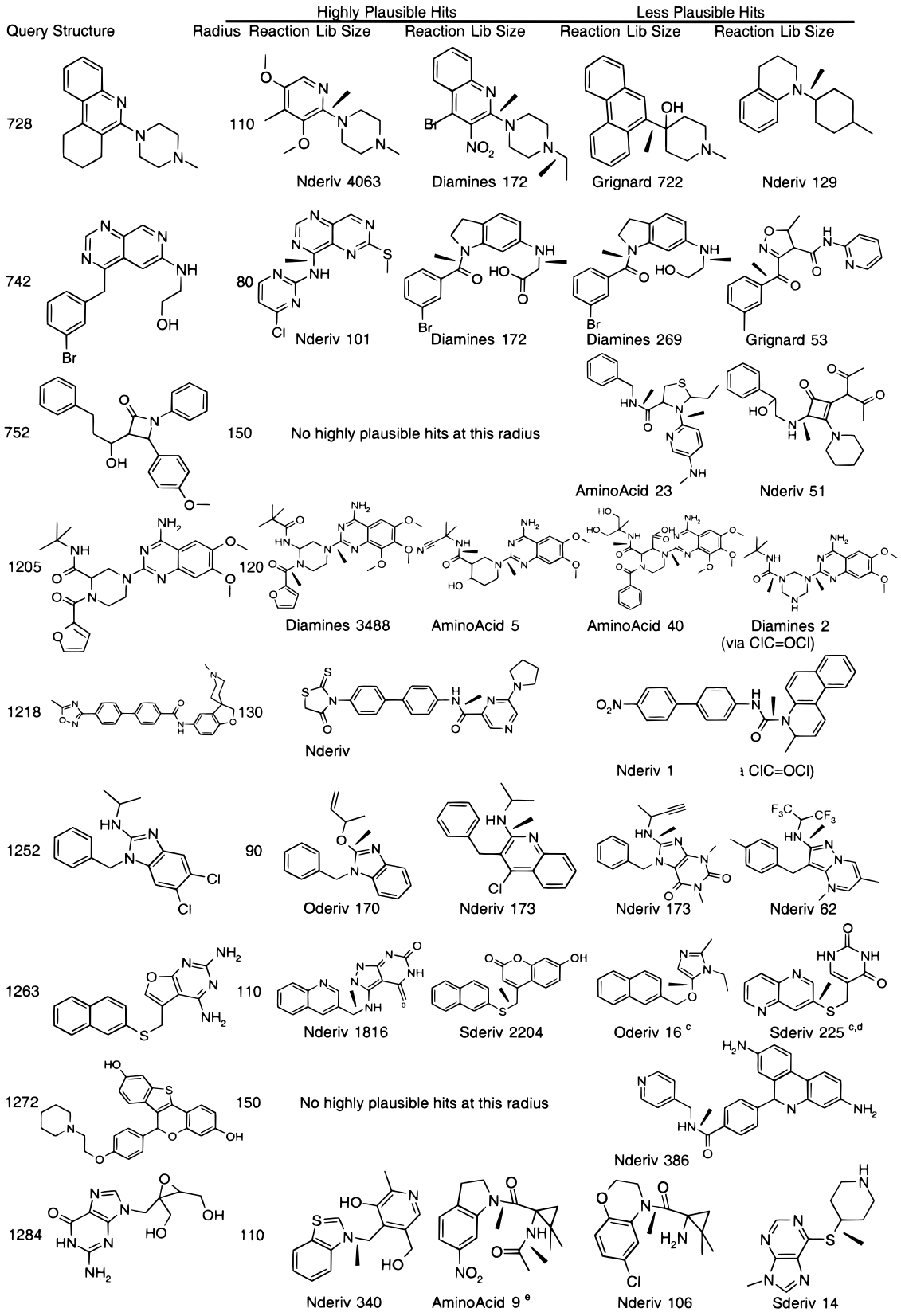



**Figure 2.** (Continued)




[a]FlexS did not produce any matches

[b]Best FlexS match differs, seemingly because of a difference in the relative position of core attachment bonds

[c]Best FlexS match differs, in seeming to prefer a pharmacophore match to maximal steric overlap

[d]Best FlexS match differs, an attachment bond being "frame shifted" by one bond length.

[e]Best FlexS match differs, with an inverted matchup of "side chain" synthons.

[f]The TOTAL score for the best FlexS match is poorer than that for any other structure in this row.

**Figure 2.** (Continued)

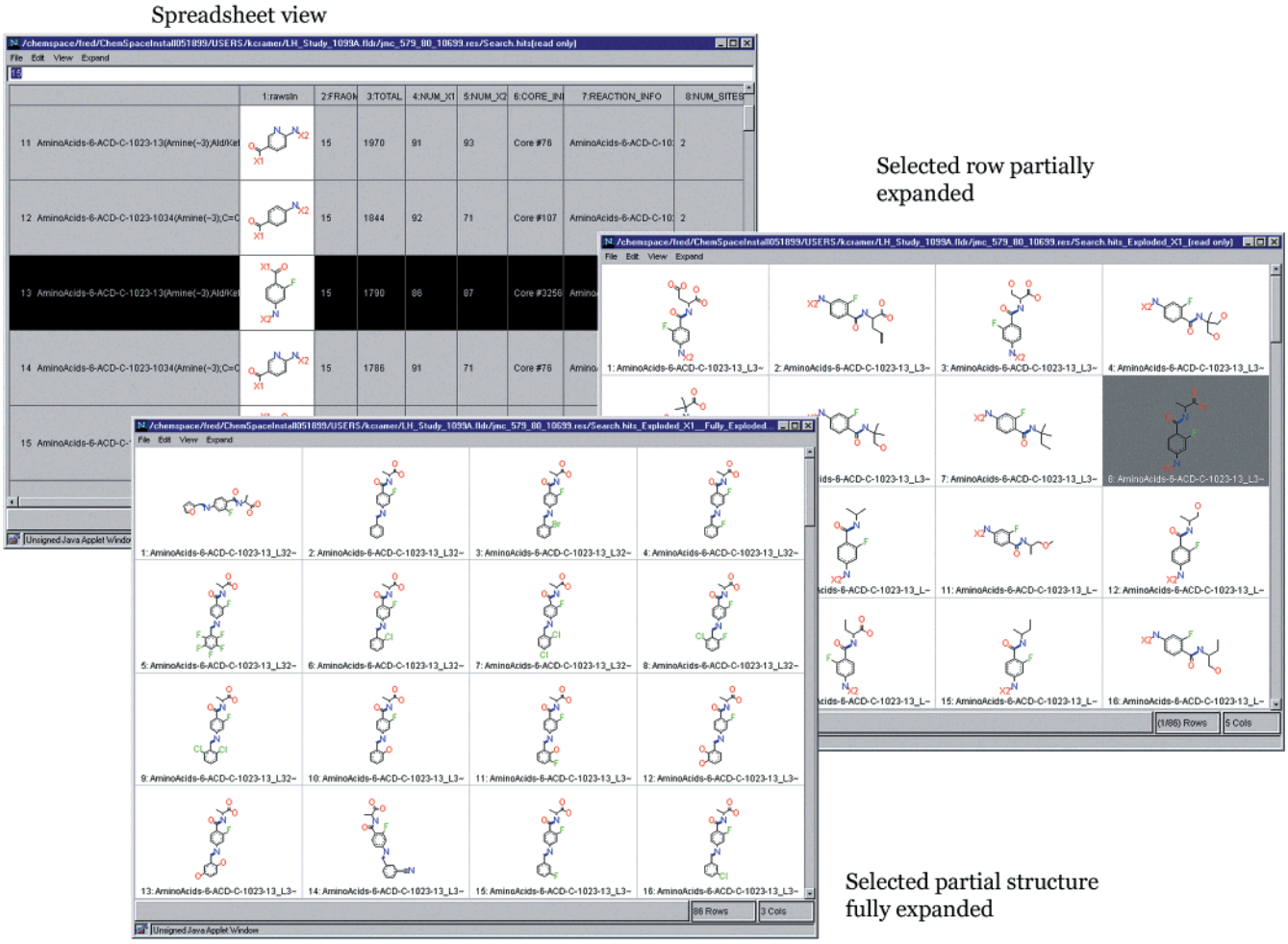



**Figure 3.** Screen captures showing the tool used for selecting the structures displayed in Figure 2. One of the spreadsheet rows representing a particular sublibrary (a particular combination of query fragmentation, specific reaction, and core structure) has been partially expanded to display the structural variations at one side chain position. A highlighted partially expanded structure has then been fully expanded to generate the front-most display of complete structures.

Selection of these few structures from such large hit lists was greatly aided by the tabular organization of search results afforded by the hit list viewer. As illustrated by Figure 3, each row of this "spreadsheet" display references a sublibrary, a particular combination of synthetic route, query fragmentation, and core (scaffold) synthon. Sorting and selection tools help in navigating among the possibly hundreds of sublibraries. Any sublibrary row can be expanded to yield either the underlying structures or another "sub-spreadsheet", most usefully by one variation site at a time, as illustrated in Figure 3.

Using this hit list viewer, a typical procedure for selecting the structures shown in Figure 2 was as follows:

- The initial hit list view was sorted by fragmentation pattern and then by decreasing number of hits.
- An individual row was selected, containing a large number of products from a synthetically attractive path, but preferably differing from any previous selections in synthetic route and in fragmentation pattern.
- The selected row was expanded at the less numerous of the two possible variation sites, to show all the shape-acceptable synthon variations.
- The display of partially expanded structures was sorted by decreasing numbers of final products. (Relatively large numbers of final products indicate the synthons that are relatively shape-similar to the matching fragments of the query structure. Conversely, smaller numbers of final products must include only the most shape-similar of the synthons at the still unexpanded variation.)
- Final structures were usually chosen from the bottom of this sorted list, only for the convenience of having fewer structures to scan.

**Validating the Shape Similarity of Topomer Searching Hits with FLEXS.** To provide objective evidence of the effectiveness of topomer searching in recognizing shape similarity, the sample hits shown in Figure 2 were evaluated by the much more rigorous and very different shape similarity algorithms embedded in the FLEXS program. The methodology of FLEXS was designed and calibrated to reproduce experimentally determined structures of ligand-bound structures.[18] Some basic properties of the topomer and FLEXS shape similarity algorithms are contrasted in Table 4.

To evaluate the shape similarity of a sample hit to its query structure with FLEXS, 3D conformations of both query and hit structure were generated with CONCORD,[19] and partial atomic charges were obtained by the SYBYL implementation of the Gasteiger/Marsili algorithm. Each hit structure was then flexibly superposed onto the rigid query structure, using the FLEXS program with all its parameter settings at their default values. The one "best scoring" superposition, of the many possibilities that usually resulted, was displayed within SYBYL, allowing detection of any major inconsistencies by visual comparison with the query structure.

We then sought an objective measure of "similarity of superpositions". Because overall topomer shape similarity is defined by combining the similarities of fragments, and because the bonds connecting those fragments are features that can be objectively identified in both the query and any

**Table 2.** Reference Keys, Target Receptors, Therapeutic Areas, and Organizational Affiliations of the 34 Query Structures

| Page | Ref | Target Receptors | Potential Therapeutic Areas | Organizational Affiliations |
|---|---|---|---|---|
| 263 | 27 | Aminopeptidase N Inhibitor | metastatic cascade inhibition | University of Tokyo; Ishihara Sangyo Kaisha |
| 266 | 28 | Phosphodiesterase Type 4 | anti-asthmatic, anti-inflammatory | Pfizer |
| 271 | 29 | ATP-Sensitive Potassium Channel Openers | hypertension, asthma, ischemia, hair growth | Bristol-Myers Squibb |
| 291 | 30 | Progesterone Receptor Agonists | hormone replacement, birth control, gynecological disorders | Ligand |
| 311 | 31 | 5-HT3 Antagonists | CNS | Universite Louis Pasteur |
| 318 | 32 | Bradycardic antianginal agents | Antiarrhythmic, blood pressure | Solvay |
| 346 | 33 | Glutamate Receptor Antagonists | CNS signalling, cellular cascade (GPCR) | Lilly |
| 379 | 34 | M1 Muscarinic Agonists | Alzheimer's, synaptic transmission | Lilly |
| 401 | 35 | Thrombin Inhibitor | blood coagulation | Merck |
| 407 | 36 | nicotinic acetylcholine receptor | non-opiate analgesics | Abbott |
| 413 | 37 | Transcription factors: nuclear factor-kappa binding (NF-kB) and activator protein-1 (AP-1) | autoimmune disease, inflammation | Signal |
| 420 | 38 | antioxidant arachidonic acid mimics | antioxidant-based therapies (replacing some anti-inflammatory therapies) | OXIS |
| 437 | 39 | Factor Xa Inhibitors | Blood coagulation | RPR, Hoffman La Roche, BMS |
| 451 | 40 | Melatonin Agonists and Antagonists | CNS | University College, London |
| 489 | 41 | Fibrinogen receptor anatagonists | platelet aggregation | Takeda |
| 503 | 42 | O-Alkylguanine-DNA transferase inhibitors | chemotherapy | Nayoga City University |
| 509 | 43 | calcium channel modulators | congestive heart failure | University of Alberta |
| 564 | 44 | Bradykinin B2 receptor antagonists | inflammatory response and diseases | Fujisawa |
| 579 | 45 | anticonvulsants | epilepsy | Pharmacia and Upjohn; Thomas Jefferson |
| 640 | 46 | Matrix metalloproteinase (MMP-9 and MMP-2) | antitumor drugs, rheumatoid arthritis, and multiple sclerosis | Shionogi |
| 655 | 47 | 5-HT7 Antagonist | unknown - sleep disorders, depression, schizophrenia | SmithKline Beecham |
| 658 | 48 | Dopamine autoreceptor agonists | antipsychotic | Otsuko |
| 668 | 49 | Adenosine Receptor Agonists | cognitive disease, renal failure, Alzheimers, Parkinsons, cardiac arrhythmia, Huntington's chorea, schizophrenia, myasthenia gravis | Universita di Pisa |
| 674 | 50 | Epibetadine-type Analgesics | Analgesia, mu-opioid receptor modulators | Abbott; several academic collaborators |
| 699 | 51 | Dopamine Transporter Ligands | Dopamine Transport, Cocaine addiction, antipsychotic | Organix |
| 728 | 52 | 5-HT3 antagonists | CNS, antiemetic | University of Innsbruck |
| 742 | 53 | Tyrosine Kinase Inhibitor | cancer, chemotherapy | Parke-Davis; University of Auckland |
| 752 | 54 | cholesterol absorption inhibitors | serum cholesterol reduction, coronary heart disease | Schering-Plough, Duke |
| 1205 | 55 | alpha1 adrenergic receptor antagonist | antihypertensive drugs | Merck |
| 1218 | 56 | 5-HT1B Inverse Agonist | serotonergic, GCPR, second messenger pathway disruption | SmithKline Beecham |
| 1252 | 57 | Nonnucleoside Antiviral Agents | antiviral for immunocompromised patients | University of Michigan, Wake Forest University |
| 1263 | 58 | DHFR Inhibitor | antimicrobial for immunocompromised patients | Duquesne University and collaborators |
| 1272 | 59 | Estrogen receptor | osteoporosis, breast cancer | Lilly |
| 1284 | 60 | nucleoside analogs to block specific viral kinases | antiviral | Ajinomoto |

**Table 3.** Counts of Hits (Structures Retrieved) and of Sublibraries Yielding Hits, Using the Smallest "Topomeric Radius" That Yielded Sufficient Hits, for Each of the 34 Query Structures

| | | *"2-piece" Virtual Libraries* | | | | | | | | *"3-piece" Virtual Libraries* | | | | | | **RVL Totals @** | | **RVL Totals @** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Query ID**[a] | **Topmi Rad**[b] | **Nderiv** #Hits[c] | SL[d] | **Oderiv** #Hits[c] | SL[d] | **Sderiv** #Hits[c] | SL[d] | **Grignard** #Hits[c] | SL[d] | **diamine** #Hits[c] | SL[d] | **dihalide** #Hits[c] | SL[d] | **amino acid** #Hits[c] | SL[d] | **Topm Rad**[b] #Hits[c] | SL[d] | **Topm Rad +10**[e] #Hits[c] | SL[d] |
| 263 | 110 | 159 | 3 | 5 | 1 | 0 | | 742 | 2 | 0 | | 78 | 5 | 0 | | 984 | 11 | 7050 | 13 |
| 266 | 110 | 2515 | 10 | 1207 | 7 | 4 | 1 | 907 | 2 | 13 | 5 | 312 | 6 | 74 | 8 | 5032 | 39 | 36735 | 62 |
| 271 | 80 | 15977 | 10 | 2828 | 7 | 180 | 3 | 1820 | 2 | 0 | | 50 | 1 | 0 | | 20855 | 23 | 89089 | 34 |
| 291 | 175 | 22 | 1 | 0 | | 0 | | 1503 | 1 | 0 | | 0 | | 0 | | 1525 | 2 | 7761 | 11 |
| 311 | 90 | 423 | 1 | 0 | | 0 | | 32 | 1 | 210 | 14 | 1988 | 1 | 0 | | 2653 | 19 | 30911 | 27 |
| 318 | 140 | 167 | 7 | 51 | 2 | 3 | 1 | 301 | 2 | 0 | | 0 | | 0 | | 522 | 12 | ND | ND |
| 346 | 110 | 3781 | 7 | 0 | | 0 | | 174 | 1 | 538 | 8 | 0 | | 15 | 2 | 4508 | 18 | 21677 | 37 |
| 379 | 80 | 2 | 1 | 0 | | 0 | | 0 | | 4165 | 19 | 13921 | 5 | 430 | 1 | 18518 | 26 | 59794 | 39 |
| 401 | 90 | 12 | 1 | 0 | | 0 | | 0 | | 118 | 6 | 0 | | 1112 | 1 | 1242 | 8 | 10855 | 14 |
| 407 | 80 | 6016 | 12 | 7022 | 6 | 279 | 5 | 3080 | 2 | 0 | | 0 | | 0 | | 16397 | 25 | 90791 | 26 |
| 413 | 80 | 26020 | 9 | 2357 | 4 | 0 | | 801 | 1 | 0 | | 0 | | 0 | | 29178 | 14 | 101587 | 18 |
| 420 | 100 | 5579 | 10 | 3356 | 7 | 0 | | 732 | 2 | 18 | 4 | 12 | 3 | 0 | | 9697 | 26 | 10483 | 38 |
| 437 | 120 | 468 | 9 | 147 | 1 | 0 | | 0 | | 0 | | 0 | | 849 | 2 | 1464 | 12 | 22209 | 34 |
| 451 | 120 | 1518 | 6 | 12 | 2 | 0 | | 910 | 1 | 548 | 27 | 0 | | 106 | 1 | 3094 | 31 | 12062 | 41 |
| 489 | 110 | 30 | 2 | 0 | | 0 | | 0 | | 45 | 7 | 0 | | 340 | 5 | 415 | 14 | 1358 | 43 |
| 503 | 80 | 19420 | 7 | 11094 | 3 | 80 | 1 | 650 | 1 | 0 | | 2 | 1 | 0 | | 31246 | 13 | 106665 | 20 |
| 509 | 120 | 59 | 5 | 16 | 3 | 12 | 2 | 1581 | 2 | 316 | 6 | 0 | | 0 | | 1984 | 18 | 12749 | 42 |
| 564 | 110 | 1 | 1 | 25 | 2 | 0 | | 4 | 2 | 407 | 20 | 44 | 3 | 0 | | 481 | 28 | 12172 | 66 |
| 579 | 80 | 15835 | 9 | 5227 | 4 | 62 | 3 | 5118 | 2 | 47344 | 40 | 251 | 7 | 17394 | 2 | 91231 | 66 | 500000 | 104 |
| 640 | 80 | 3158 | 5 | 11 | 2 | 25 | 1 | 635 | 2 | 300 | 6 | 139 | 3 | 98 | 1 | 4366 | 20 | 12312 | 29 |
| 655 | 90 | 56 | 6 | 7 | 1 | 0 | | 2 | 1 | 321 | 4 | 858 | 4 | 1705 | 5 | 2949 | 22 | 15383 | 50 |
| 658 | 90 | 134 | 3 | 0 | | 0 | | 0 | | 2 | 2 | 3838 | 1 | 0 | | 3974 | 6 | 23290 | 34 |
| 668 | 90 | 307 | 1 | 496 | 6 | 49 | 1 | 0 | | 0 | | 0 | | 0 | | 852 | 8 | 1707 | 12 |
| 674 | 90 | 1354 | 7 | 21 | 1 | 0 | | 40 | 1 | 0 | | 0 | | 0 | | 1415 | 9 | 6413 | 14 |
| 699 | 90 | 0 | | 10 | 1 | 0 | | 0 | | 29799 | 28 | 0 | | 10 | 3 | 29819 | 32 | 95115 | 55 |
| 728 | 110 | 4323 | 4 | 269 | 1 | 0 | | 746 | 2 | 2251 | 31 | 0 | | 0 | | 7589 | 38 | 33429 | 76 |
| 742 | 80 | 101 | 1 | 49 | 1 | 0 | | 53 | 1 | 763 | 10 | 0 | | 0 | | 966 | 13 | 8380 | 32 |
| 752 | 150 | 137 | 6 | 88 | 4 | 1 | 1 | 318 | 2 | 241 | 7 | 0 | | 3071 | 2 | 3856 | 22 | ND | ND |
| 1205 | 110 | 0 | | 0 | | 0 | | 0 | | 4505 | 24 | 0 | | 608 | 1 | 5113 | 25 | 11248 | 35 |
| 1218 | 130 | 103 | 3 | 3 | 1 | 0 | | 25 | 2 | 0 | | 0 | | 0 | | 131 | 6 | 1347 | 16 |
| 1252 | 90 | 313 | 9 | 170 | 1 | 12 | 1 | 24 | 1 | 0 | | 0 | | 0 | | 519 | 12 | 1610 | 16 |
| 1263 | 110 | 4193 | 11 | 633 | 3 | 3020 | 4 | 509 | 2 | 0 | | 0 | | 0 | | 8355 | 20 | 59191 | 23 |
| 1272 | 150 | 416 | 6 | 26 | 2 | 0 | | 240 | 2 | 0 | | 14 | 2 | 0 | | 696 | 12 | 3942 | 25 |
| 1284 | 110 | 8068 | 12 | 244 | 6 | 43 | 4 | 5017 | 2 | 0 | | 364 | 4 | 31 | 3 | 13767 | 31 | 78520 | 189 |
| Totals | | 120667 | | 35374 | | 3770 | | 25964 | | 91904 | | 21871 | | 25843 | | | | | |
| Log (Freq of hits)[f] | | -4.12 | | -3.67 | | -3.54 | | -4.04 | | -8.33 | | -6.86 | | -8.40 | | | | | |

[a] The Query ID is the page number of the article, within volume 41 of the *Journal of Medicinal Chemistry*. [b] The topomer radius (maximum shape difference) used to produce the results shown. [c] The number of structures retrieved from the library. [d] The number of sublibraries from which any structures were retrieved. (Sublibraries are distinguished by reagent type, query fragmentation, or core structure). [e] Topm Rad + 10 results were those obtained by repeating the search with a topomer radius equal to the sum of 10 and the value shown in the Topm Rad Column. [f] Logarithm of the ratio of the corresponding value in the Totals row to the number of compounds in that library (from Table 1).

**Table 4.** Differences between the Shape Similarity Algorithms of Topomer Searching and of FLEXS

| Aspect of Shape Similarity | Topomer Shape Similarity | FLEXS Shape Similarity |
|---|---|---|
| Characteristics of "shape" | Steric shape overlap only | Emphasizes matches in pharmacophoric character and partial charge, as well as steric overlap |
| Molecular shape similarity | Root sum of several squared fragment similarities | Overall shape similarity |
| Conformational variability | Single topomer conformations compared | Multiple conformational possibilities considered for all ring systems and acyclic bonds |

topomer-similar hit structures, we concentrated on the relative locations of those bonds. The shortest of the four distances among the endpoints of a pair of such bonds was used to identify one pair of atoms. If the distance between the two most distant atoms remaining was greater than 2.0 Å, we considered the FLEXS overlay of those structures to disagree with the topomer shape similarity result. (When there were two fragmentation bonds, e.g., in structures from the "AminoAcid" library, both bonds had to pass this test.) Among these 122 examples, this objective index was found to be an infallible marker for all such visually recognizable discrepancies as conformational or stereoisomeric state differences, end-to-end flips, and relative displacements along chains.

**Determining That Structures Reported To Be Active in the Query-Containing Articles Are Retrieved.** In general, an object will be retrieved from a database if and only

if the object exists within the database and the searching procedure correctly identifies the object. In this particular case, a structure already known to be active will be retrieved if and only if it can be formed from commercially offered reagents by one of the seven reaction classes, and the shape differences between its corresponding synthons and those of the query structure are small enough.

Therefore the first step was to determine which of the structures in the 34 articles might be contained within the RVL. Roughly 10 of those 34 articles described structures amenable to synthesis by one of the seven underlying reaction classes. Within each of these 10 articles, the tables of structures were scanned to determine whether the invariant portion of the structure either was itself a synthon derived from a commercially available building block (queries 311 and 413) or else could have been synthesized from two such synthons by one of the seven reaction classes (query 379). Finally the variant synthons in the acceptable tables were also checked for commercial availability. This process identified 28 of the several hundred compounds described in the original 34 publications, from the three queries 311, 379, and 413, that could also be found among the tens of trillions of structures in the RVL.

To determine as efficiently as possible how many of these 28 would then be identified by topomer searching, it was noted that each of these 28 structures were substantially identical to the corresponding query structures. Thus the overall steric difference between each structure and the corresponding query would simply be equal to the topomeric difference in the structurally variant portions. These 28 topomer difference values were then computed directly, using the same underlying code that is used to build and search the RVL.

## Results

Topomer shape similarity searching of simple combinatorial chemistry products appears to be effective for a very high proportion of "typical medicinally interesting structures", according to the search results summarized in Table 3. The critical search radius, within which topomer shape similarity to an active structure clearly and significantly enhances the frequency of similar biological activity, has taken on values of 90 units[5] and of 120 units.[7] Of the 34 queries shown in Table 3, a search radius of 120 shape units or less yielded the useful results shown in 29 searches (85%). (Here a "useful result" implies that at least one combinatorial synthesis matrix could be designed; hence fully acceptable hit lists contained at least 400 shape-similar structures from at least six different virtual sublibraries, including structures plausible enough to be shown in Figure 2.) Note further that most of the retrieved structures suggest a possible "lead hop", a "chemotype" different from the query structure, if only because of being synthesized by a route different from that of the query structure. Even at the rather conservative cutoff of 90 shape units or less, 16 of the queries (47%) found useful results. This very positive outcome is the key result of our investigation.

Inspection of the query structures (first column of Figure 2) shows that the major feature distinguishing the five less successful topomer queries was an exotic cyclic system. Search 291 contains a tetracyclic system moderately reminiscent of an unusually substituted steroid; search 318, a tricyclic system including a spiro junction grafted onto a bridge; search 752, a four-membered ring; search 1218, a tricyclic system with a spiro junction; and search 1272, another steroid-variant tetracycle. Such oddly shaped ring systems are seldom encountered among commercially offered intermediates. Since these virtual libraries are limited to combinatorial products formed by making one or two bonds between those intermediates, a lower degree of success in shape similarity searching with these five queries is only to be expected.

The data in Table 3 also address two subsidiary questions. How rapidly do the number of structures retrieved and the number of libraries included increase with an increase in the topomer search radius? Are there particular virtual libraries (synthetic routes) that are more likely to yield hits in shape similarity searching?

The dependence of search output on the search radius can be derived from the values in the Topm Rad and Topm Rad+10 columns of Table 3. Increasing the search radius by 10 units increases the number of hits 5-fold and doubles the number of sublibraries yielding hits (based on the geometric mean ratios, each calculated as the antilog of the mean logarithms of the ratios of the values shown).

The "yield frequency" of the seven individual virtual libraries greatly depended on the query structure. Every one of the seven libraries was a major contributor of results to at least one of the 34 searches. For example, study of Table 3 shows that six of the seven libraries produced the most hits for at least one of the 34 queries (and the seventh was once a runner-up). Nevertheless it is useful to define an "overall yield frequency" as the total number of hits found for all 34 searches divided by the size of the libraries (shown in Table 1). To better illuminate overall trends, and to minimize the biases from individual search results, this ratio is expressed as its logarithm, forming the bottom row of data in Table 3. Here there emerges an unmistakable trend. Although there are important exceptions (searches 379, 401, 489, 564, 655, 658, 699, 1205), in general the structures in the "2-piece" virtual libraries were from one to ten thousand times more likely to be shape-similar to a query structure than are structures in the "3-piece virtual libraries.[20] Yet, since the "3-piece" libraries are also one to ten thousand times larger than the "2-piece" libraries (Table 1), the numbers of hits per library are rather similar. (The Totals at the bottom of Table 3 vary by less than an order of magnitude among the seven libraries, excepting only the Sderiv library.)

The individual examples of hit structures shown in Figure 2 are also an important result. They were chosen deliberately to show how topomer shape searching often detects structural similarities among readily accessible compounds that will not immediately occur to every medicinal chemist. The route of synthesis for each structure is indicated by wedges pointing to the bond(s) to be formed and by the library name, enough information for most synthetic chemists to recognize the intermediates and reactions that would be involved.[21] The number to the right of each library name is the reaction size, the total number of structures, at least as shape-similar, that could be produced in the same combinatorial reaction from other commercially offered building blocks. The two left-hand columns of Figure 2 contain structures that were deemed more "plausible" than those in the two right-hand columns. "Plausibility" was determined mainly by synthetic accessibility, with the more plausible compounds eliciting decreased con-

cerns about protection group chemistry, sluggish reaction, or competing functionality. Thus Nderiv, Oderiv, Sderiv, and AminoAcid products dominate the "more plausible" columns, while Diamine, Grignard, and Dihalide products are usually "less plausible".

Opinions will surely vary about which of the structures shown in Figure 2 are "nonobvious". Some will no doubt seem naïve because of additional SAR knowledge that of course was not considered in topomer searching. (In actual library design, such information is used to filter the topomer search results.) For example, the two hits shown from query structure 509, though clearly somewhat shape-similar, do not include the dihydropyridine moiety itself and so seem unlikely to share its antihypertensive activity. On the other hand, the strong tendency to attribute promising activity, a rare outcome, to the presence of a rare structural feature can be misleading. An accidental but excellent example is the hydroxamic moiety in query 266. Every organic chemist (including one of the authors) who so far has evaluated the hits shown for this query has noted disapprovingly the absence of the hydroxamic acid group. But in fact this group is not necessary, nor perhaps even desirable, for PDE4 inhibition. Instead a feature which is shared by many PDE4 inhibitors used clinically is the more subtle 3-cyclopentyloxy-4-methoxyphenyl assembly.

Some generalizations are suggested by the structures collected in Figure 2.

• The vast majority of the hit structures undoubtedly represent potential "lead hops" or different "chemotypes" from the query structure, as mentioned above. This assertion follows from their differences in synthetic route and/or intermediate structures from the query structure and applies to all hits referenced in Table 2, not just the few structures shown in Figure 2. Clear exceptions to this generalization are found among queries 401 and 413 (the only two originally synthesized combinatorially), while arguable exceptions are included among queries 266, 271, 311, 379, 451, 503, 579, 640, 655, 668, 699, 1205, and 1252.

• Certain topologies are not easily formed by combinatorial chemistry, more precisely by short reaction sequences assembling commercially available building blocks. Exotic ring systems have already been mentioned. Another challenging yet frequent structural feature is an acyclic bond connecting two rings, especially if one or both of the rings is aromatic or heteroaromatic. (Such a feature is found, for example, in queries 263, 266, 291, 311, 346, 509, 640, 658, 674, 728, 752, 1205, 1218, and 1272, over 40% of the total.) Whenever in Figure 2 the formation of such a bond is indicated, most often the reaction is (hetero)aromatic nucleophilic displacement by a saturated heterocycle, with the occasional reductive amination or Grignard.[22] Unless the ring-to-ring bond in the query structure was also a heterocycle-to-heteroaromatic bond, the resulting structure is thus constrained to be quite different from the query. In contrast, chains of one or especially two atoms between rings are very easy to achieve combinatorially, in many different ways.

• Commercially available reagents often lack the additional hydrogen bond accepting and (especially) donating groups that are important in so many biologically active structures, no doubt because of the synthetic complications these groups introduce. Thus the number of candidate structures containing some such possibly critical query feature, such as a basic amine, may be small. (Of course such features, when known to be important within individual queries, can readily be designed into custom virtual libraries, as additional synthetic steps.)

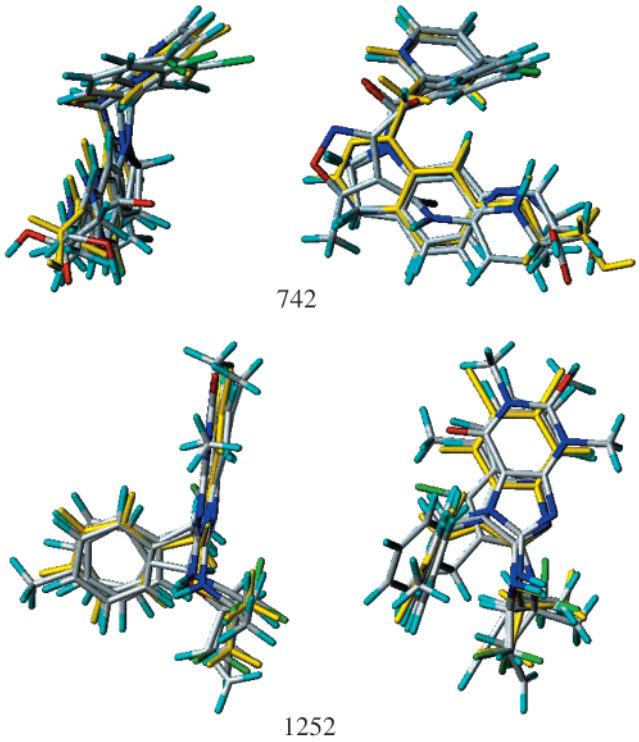



**Figure 4.** Orthogonal views of the overlays generated by FLEXS for the four representative hit structures shown in Figure 2 for queries 742 and 1252. In yellow is the CONCORD-generated structure for the query. The other structures are the "best matching" overlays generated by FLEXS, one for each of the four hit structures.

**Validation of Shape Similarity by FLEXS.** Shape similarities of all 122 hit structures in Figure 2 to their corresponding query structures were calculated by FLEXS as described above. In most cases, visual inspection of the best scoring FLEXS overlay showed a striking agreement of shape between query and hit. Eight such overlays are shown in Figure 4. Those structures showing any discrepancy in their FLEXS results are marked with a footnote in Figure 2 (next to the sublibrary hit count). There are 22 such structures, yielding a minimal rate for "agreement between FLEXS and topomeric shape similarity" of 100/122 or 82%. However, in the seven cases with a footnote of "a", the discrepancy was that FLEXS did not report that any shape similarity existed. With these cases excluded as indeterminant, the agreement rate would improve to 100/115 or 87%.

The remaining discrepancies result from a distance of greater than 2.0 Å in the endpoints of one or more fragmentation bonds. If it is inferred, from the seven no-result cases, that FLEXS can fail to find a shape similarity which actually exists, then the possibility also arises that the FLEXS scoring system might have preferred the ChemSpace-implied overlap to the best overlap that FLEXS itself could find. Indeed, in 10 of the 14 possible comparisons,[23] footnoted with an "f" in

Figure 2, the FLEXS TOTAL score for the discrepant hit was poorer than the FLEXS TOTAL score for any of the other hits for the same query structure. If these 10 suspect FLEXS overlap cases are also dropped from consideration, the rate of agreement between best FLEXS overlap and the topomer searching result becomes 100/105 or 95%.

All 16 discrepant overlays were examined visually for four anticipated sources of disagreement between the FLEXS and topomer alignments, recorded as footnotes "b" through "e" in Figure 2. The first of these arises because the topomer methodology approximates overall structure differences as a combination of fragment differences. This approximation is weakest for cores, since changing the position of two large side chains to a central phenyl core from, say, *para* to *ortho* has a minimal effect on the overlap volume of the core and no effect at all on the side chain volumes, but an enormous effect on the overlap volumes of the completed structure. To compensate for this insensitivity, the topomer difference value for a core is increased by any differences in the endpoints of its attachment bonds, weighted relative to volume overlap by 10.0 (adjustable). In Figure 2, the poor relative alignment for the structure marked with "b" suggests that this default weighting value should be increased.[24] A second anticipated source of discrepancy was the treatment of pharmacophoric centers (hydrogen bond donor and acceptors; charged atoms). These play a central role in generating FLEXS alignments but are completely ignored in topomer shape differences. And indeed, a preference by FLEXS for matching pharmacophoric points rather than maximizing overlap volume seemed visually to contribute to at least 7 of the 16 discrepant overlays, marked by "c" in Figure 2. A third source of discrepancy "d" could arise whenever the compared structures differ in their longest dimension. A structure can then "slide along" that dimension, relative to the other, without much change in the overlap volume. There were two examples where such a displacement amounted to at least a bond length. The fourth source of discrepancy "e" is most likely with structures that are roughly centrosymmetric, such that the overlap volume does not change much when a rotation or "flip" about that center interchanges two groups. This was the most frequently observed of the four anticipated sources of discrepancy, in 9e of the 14 discrepant overlays.

**Determination That Structures Reported To Be Active Are Retrieved.** At a search radius of 90, 16 of the 28 published structures that are also contained within the RVL would have been retrieved. Of these 16 structures, 14 (87%) were reported to have "high" potencies (within 1.0 log units of the query structure). Of the other 12 structures, less shape-similar to the query structures, just six (50%) were reported to have "high" potency. This particular increase in frequency of activity with increasing shape similarity to the most interesting (query) structure has marginal statistical significance ($p < 0.1$).[25]

## Discussion

The combinatorial chemistry paradigm has engendered widespread interest in virtual libraries of structures, formed by assembling readily available building blocks with short reaction sequences. Our results confirm the enormous impact that virtual libraries can have on the compound selection activities so critical for successful drug discovery. Searching such virtual libraries using typical medicinally interesting structures as queries indeed produced a large variety of structures that are not obvious query analogues, that are readily accessible synthetically, and that are likely to share the activity of the query structure. The few query structures that were not productive could have readily been identified by the oddly shaped ring systems that each contained.

It is possible that technologies other than topomer shape similarity searching can be found that will be as effective in selecting structural targets within very large virtual libraries. However, to achieve this feat, such a technology would need several key attributes:

**Extremely High Searching Speeds.** Our current implementation searches the RVL at rates of $10^{12}$ to $10^{13}$ structures per hour—or over a thousand times the total of structures ever registered by Chemical Abstracts every second! The results we report depended on this high level of underlying performance, which allowed the large variety and number of query structures, of query fragmentations (required to produce the great structural variation shown among the hits), and of query radii. Such an extraordinary search speed can be obtained in an exhaustive search only by a highly restrictive "divide-and-conquer" algorithm, applicable only to fragment descriptors that can be combined strictly independently to effect the overall structure selections.

**A Selection Criterion with Experimentally Validated Success in Identifying Active Structures, over a Wide Variety of Query Structures and Biological Targets.** For example, we report here that the experimental frequency of high activity was higher (14/16 or 87% compared to 6/12 or 50%) for the structures that were the most topomerically shape-similar to the query and that were also to be found within the RVL. But this result is merely an increment to a consistent accumulation of findings "validating" topomeric shape as the most powerful and general similarity descriptor known for various biological activities. Similar results had been obtained retrospectively for 20 series, with the topomer shape similarity criterion being the best among 11 classes of molecular descriptor.[5] In a prospective study involving independent synthesis and testing of a 425-compound library, the seven structures found to be highly active structures were all among the 63 that were shape-similar to a query by 120 units or less.[7]

**Structural Variety Among the Selected Target Molecules.** For many, an ideal selection method would report only structures that are both "novel" (representing "lead hops" or new "chemotypes") and biologically promising. Others would object, either believing this goal to inherently be both subjective and somewhat self-contradictory, or else being sufficiently untrusting of both human intuition and computer legerdemain as to want to see considerable "obvious" structural output as well. Regardless, any experienced researcher knows how important the promise of an attractive and immediately accessible "lead hop" can be when adverse biological data or a competitive patent unexpectedly appears. We

believe the structural examples shown in Figure 2 strongly support the proposition that shape similarity searching is an extraordinarily powerful tool for selecting the most promising "lead hops" from the enormity of readily accessible structures. A wide variety of potentially relevant chemistries (as evidenced for topomer searching by the results in Table 3), and a flexible library construction facility which simplifies the entry of more chemistries, are both very important to the overall goal of shortening make-and-test cycles.

**Minimal Query Information Needed.** Similarity to a "known active" structure seems the most widely applicable compound selection strategy. As one contrasting example, if a receptor target structure had been required, as in "combinatorial docking", then compound selection would have been impossible for over 30 of the 34 queries used here.

Other valuable attributes of a methodology for selecting synthetic targets from a virtual library include:

**Adaptability of the Underlying Physical Model.** Usually there is more information than a single active structure to help guide compound selection, especially on second and subsequent rounds of synthesis and testing. Such emerging information can often be blended readily with shape similarity selection, in predictive formulations such as "shape increases (or decreases) for this group of atoms seem good, not bad", much as in the established CoMFA method. Or, a pharmacophore hypothesis, a different 3D shape based selection method requiring complete structures, can be applied to a set of, say, 100 000 structures initially chosen by such vastly faster methods as topomeric shape similarity. Of course, if a reliable 3D structure of the target exists, a variety of docking methods may also be applied to the initial hits.

**Independently Verifiable Search Criterion.** The very fast methods needed to search vast virtual libraries will necessarily involve approximations. For example, topomer searching expresses overall molecular shape difference as the root of the sum of squared steric differences between single conformations. While synthesis and testing provide the ultimate validation of such approximations, more immediate feedback can provide useful encouragement. Thus we are pleased to be able to report that the shape similarities of the 122 hand-selected diverse structures were strongly confirmed by the much subtler similarity criteria of the algorithmically disjoint FLEXS program.

**Integration into an Automated Discovery Process.** Surely the goal in compound selection is to optimize the ratio of information returned to experimental time and cost. However such major determinants of synthesis time and cost as reagent availability and outcomes of synthetic validation experiments will often emerge very late in the intended make-and-test cycle. Therefore an ability to rapidly accept new constraints and redesign target libraries can be one of the most important qualities of a practical compound selection methodology.

Some of the structures suggested in Figure 2 will appear naïve to some observers, especially to those who are the most familiar with the cumulative SAR associated with any particular query structure. Indeed, we emphasize that some difference between the outcomes of any similarity searching and of any specific SAR analysis is both inevitable and desirable, even if the molecular descriptors are similar, because the underlying selection strategies differ. When similarity searching, *every* structural difference is *equally* bad. But the very essence of SAR knowledge is that structural changes have unequal effects, with some small changes found to be unacceptable and (hopefully) other relatively large changes found to be highly desirable. Thus the structures in Figure 2 are proposed as early stage selections, most appropriate for rapidly and inexpensively following up screening hits or "lead-hopping" around competitive structures, rather than for later stage lead optimization.

That said, however, we believe that SAR-based selection strategies may often be overemphasized or prematurely applied, relative to similarity strategies. The possible interactions between a ligand and a receptor site are so numerous, complex, dynamic, and interrelated, as to make simple SAR generalizations suspect without testing and to make complex SAR rules too localized to be very useful. For these reasons we believe that any SAR information or hypotheses should be blended with, rather than replacing, similarity-based selection. With a similarity selection method as powerful as topomer searching, we are currently achieving such a blend by using SAR approaches such as pharmacophores or CoMFA models to filter the initial topomer searching output. (This initial output can of course be made as voluminous as desired by increasing the search radius.)

The physical simplifications underlying the topomer similarity calculation, considering only assemblies of steric[26] differences between single conformations of fragments, perhaps need some remarks. Originally this approach was pragmatic, to include only those differences of shape that can be calculated with greatest certainty and compared with greatest speed. Its strongest support has been empirical, in its consistently superior record so far in identifying biologically active structures and in its appeal to many practicing chemists. We now add that the "best" shape similarities computed by the full structure, conformationally flexible, pharmacophore-centric comparison algorithms of FLEXS are only very seldom different from those found by topomer searching. Since programs such as FLEXS generate many acceptable candidate overlays differing little in their score, it is truly remarkable that the single best-scoring overlay so often agrees with the implicit topomer alignment. To determine whether the agreement would be as good for the many possible binding conformations of each query structure, in addition to the single CONCORD-generated conformation considered here, extensive additional work would be necessary.

It may be surprising that as few as 28 of the other structures described in the 34 articles were included among the 26 trillion structures referenced by the RVL. But of course the other hundreds of structures in those publications were assembled in more than two steps and/or their components were not commercially available.

In summary, similarity searching of large virtual libraries, at least by topomer shape similarity, appears

very useful for selecting structures likely to share the biological properties of the query structure. No other tool has been described which promises such speed, such demonstrated predictive performance, and such nonobvious yet plausible hit structures. The potential value of this approach for enhancing the overall efficiency of early stage drug discovery is further strengthened.

**Acknowledgment.** Special thanks are due to Farhad Soltanshahi (Tripos) for the design of the GUI depicted in Figure 3 and able programming support overall and to synthetic chemists Tony Cooper and Julian Smith (Tripos Receptor Research) for continuing and constructive feedback. We also thank Stefan Guessregen, David Patterson, Mike Lawless, and Bob Clark for useful discussions, Dave Larson and Jon Swanson for technical support, and Don Felley and Ed Hodgkin for helping to move this project forward.

## References

(1) *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Chapman and Hall: London, 1995. Maggiora, G. M.; Johnson, M. A. *Concepts and Applications of Molecular Similarity*; John Wiley and Sons: New York, 1990.

(2) A leading reference is: Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.

(3) A leading reference is: Wang, T.; Zhou, J. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 71−77.

(4) For comparison, this work involves exhaustive selection of structures from a million-fold larger domain, more than $10^{13}$ structures, the majority undoubtedly both "drug-like" and "readily synthesizable" (assuming actual availability of commercially offered reagents). Although $10^7$ is a reasonable threshold if only compounds physically on hand are to be chosen, any computational limitation to 0.000001 of the actual synthetic candidates seems a serious restriction. The structural variety reported here indicates the benefit of as large a candidate pool as possible.

(5) Cramer, R. D.; Clark, R. D.; Patterson, D. E.; Ferguson, A. M. Bioisosterism as a molecular diversity descriptor: steric fields of single topomeric conformers. *J. Med. Chem.* **1996**, *39*, 3060−3069.

(6) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: a useful concept for validation of molecular diversity descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059. Other important comparative studies, but considering different descriptors, are the following: (a) Matter, H. Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **1997**, *40*, 1219−1229. (b) Brown, R. D.; Martin, Y. C. Use of structure−activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.

(7) Cramer, R. D.; Poss, M. A.; Hermsmeier, M. A.; Caulfield, T. J.; Kowala, M. C.; Valentine, M. T. Prospective Identification of Biologically Active Structures by Topomer Shape Similarity Searching. *J. Med. Chem.* **1999**, *42*, 3919−3933.

(8) Hodgkin, E.; Andrews-Cramer, K. Knowledge-based Drug Discovery. *Modern Drug Discovery* **2000**, *3*, 55−60.

(9) Throughout this article, the term "hit" refers to results of a computer search, not to results of a laboratory assay experiment.

(10) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem.* **1998**, *41*, 4502−4520. The FLEXS program is distributed by Tripos, Inc.

(11) Cramer, R. D.; Patterson, D. E.; Clark, R. D.; Soltanshahi, F.; Lawless, M. S. Virtual libraries: a new approach to decision making in molecular discovery research. *J. Chem. Inf. Comput. Sci.* **1998**, *6*, 1010−1023.

(12) Almost exactly as in a CoMFA steric field, except that the field contributions of each individual atom are attenuated as more rotatable bonds separate it from a fragment attachment bond. 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. Moreover, there is nothing about the topomer alignment procedure that requires the final similarity comparison to be steric shape only. Thus an extension of topomer similarity, "feature matching" á la conventional 3D pharmacophoric searching, has yielded excellent results in neighbor validation experiments (per ref 6) and is currently being implemented.

(13) This compendium of all commercially offered compounds may be obtained from MDL Information Systems, Inc., 140 Catalina Street, San Leandro, CA 94577.

(14) For comparison, this RVL contains more than one million times the cumulative number of structures registered by Chemical Abstracts.

(15) In practice, particularly for the diamine library, one or two blocking/deblocking operations are highly desirable.

(16) Ash, S. M.; Cline, M. A.; Homer, R. W.; Hurst, T.; Smith, G. B. SYBYL line notation (SLN): A versatile language for chemical structure representation. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1−9.

(17) Because topomeric shape differences are computed over CoMFA steric fields, their units are formally kcal/mol. In practice there are three major factors to consider in interpreting these units. (1) Differences are combined, both per lattice point and per fragment, in root-sum-square Euclidean fashion, not by simple addition. So successive small increases in topomer radius will yield increasingly large structural variations. (2) Replacement of a hydrogen by a methyl group produces a shape difference of 60 topomer units. (3) The search radii that have produced statistically significant increases in the frequency of biological activity are in the range of 90 to 120 topomer units.

(18) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. Predicting receptor−ligand interactions by an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(19) Rusinko, A., III; Skell, J. M.; Balducci, R.; McGarity, C. M.; Pearlman, R. S. Program available from Tripos, Inc., St. Louis, MO.

(20) This strikingly lower likelihood of structures in three-piece libraries to be shape similar to a typical query structure, compared to those in two-piece libraries, seems to have several causes. Probably the most important is the additional shape constraint imposed by the second synthetic process needed to form the second bond. As an example of such a constraint, note that the products of amide bond formation can be shape similar only to queries that fragment into something shape-similar to a carbonyl. Thus an amide bond cannot be used to directly connect two rings and can form a one-atom bridge only when the amine is cyclic (e.g., piperidine), but it can form two-atom bridges between rings in the common situations where one of the bridging atoms is sufficiently shape-similar to carbonyl.

(21) However, it will usually not be certain whether an alkyl-to-N bond results from leaving group displacement (halide reagent) or from reductive amination (aldehyde or ketone reagent).

(22) Addition of virtual libraries that include a Suzuki coupling step would have helped. Indeed, another benefit of this investigation is to call attention to such missing sequences.

(23) In two other cases, there was no "matching" hit so no comparison was possible.

(24) The 39 other core-built structures in Figure 2 (Diamines, AminoAcids, Dihalides) for which FLEXS and topomer structures agree might seem to counter this judgment. However, in most of those structures there was so little difference in the attachment bond locations that changes in their weight would be immaterial.

(25) The $\chi^2$ value for this distribution was 3.07, with one degree of freedom. For a shape difference of 120 rather than 90, the corresponding cell counts are 16/20 active among structures less than 120 shape units different, and 4/8 active for the remainder. Even though these neighborhood enrichments differ little from those for the radius of 90, the increased imbalance in the underlying cell counts depresses the associated $\chi^2$ value to 1.26, which with one degree of freedom does not even reach a significance level of $p < 0.25$.

(26) Other workers have also recently stressed steric complementarity in ligand binding, as contrasted with hydrogen bonding and other factors. See Davies, A. M.; Teague, S. J. Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis. *Angew. Chem.* **1999**, 736−749.

(27) Michayi, H.; Kato, M.; Kato, F.; Hashimoto, Y. Novel Potent Nonpeptide Aminopeptidase N Inhibitors with a Cyclic Imide Skeleton. *J. Med. Chem.* **1998**, *41*, 263−265.

(28) Kleinman, E. F.; Campbell, E.; Giordano, L. A.; Cohan, V. L.; Jenkinson, T. H.; Cheng, J. B.; Shirley, J. T.; Pettipher, E. R.; Salter, E. D.; Hibbs, T. A.; DiCapua, F. M.; Bordner, J. Striking Effect of Hydroxamic Acid Substitution on the Phosphodiesterase Type 4 (PDE4) and TNFalpha Inhibitory Activity of Two Series of Rolipram Analogues: Implications for a New Active Site Model. *J. Med. Chem.* **1998**, *41*, 266−270.

(29) Atwal, K. S.; Grover, G. J.; Lodge, N. J.; Normandin, D. E.; Traeger, S. C.; Sleph, P. G.; Cohen, R. B.; Bryson, C. C.; Dickinson, K. E. J. Binding of ATP−Sensitive Potassium Channel (KATP) Openers to Cardiac Membranes: Correlation of Binding Affinities with Cardioprotective and Smooth Muscle Relaxing Potencies. *J. Med. Chem.* **1998**, *41*, 271−275

(30) Zhi, L.; Tegley, C. M.; Kallel, E. A.; Marscke, K. B.; Mais, D. E.; Gottardis, M. M.; Jones, T. K. 5-Aryl-1,2-dihydrochromeno[3,4-f]quinolines: A Novel Class of Nonsteroidal Human Progesterone Receptor Agonists. *J. Med. Chem.* **1998**, *41*, 291−302.

(31) Rival, Y.; Hoffman, R.; Didier, B.; Rybaltchenko, V.; Bourguignon, J.-J.; Wermuth, C. G. 5-HT3 Antagonists from Aminopyrazine-type Muscarinic M1 Agonists. *J. Med. Chem.* **1998**, *41*, 311−317.

(32) Schon, U.; Antel, J.; Bruckner, R.; Messinger, J. Synthesis, Pharmacological Characterization, and Quantitative Structure−Activity Relationship Analyses of 3,7,9,9-Tetraalkylbispidines: Derivatives with Specific Bradycardic Activity. *J. Med. Chem.* **1998**, *41*, 318−331.

(33) Ornstein, P. L.; Bleisch, T. J.; Arnold, M. B.; Wright, R. A.; Johnson, B. G.; Schoepp, D. D. 2-Substituted (2SR)-2-Amino-2((1SR,2SR)-2-carboxycycloprop-1-yl)glycines as Potent and Selective Antagonists of Group II Metabotropic Glutamate Receptors. 1. Effects of Aryl, Arylalkyl, and Diarylalkyl Substitution. *J. Med. Chem.* **1998**, *41*, 346−357.

(34) Ward, J. S.; Merritt, L.; Calligaro, D. O.; Bymaster, F. P.; Shannon, H. E.; Mitch, C. H.; Whitesett, C.; Brunsting, D.; Sheardown, M. J.; Olesen, P. H.; Swedberg, M. D. B.; Jeppesen, L.; Sauerberg, P. 1,2,5-Thiadiazole Analogues of Aceclidine as Potent M1 Muscarinic Agonists. *J. Med. Chem.* **1998**, *41*, 379−392.

(35) Brady, S. F.; Stauffer, K. J.; Lumma, W. C.; Smith, G. M.; Ramjit, H. G.; Lewis, S. D.; Lucas, B. J.; Gardell, S. J.; Lyle, E. A.; Appleby, S. D.; Cook, J. J.; Holahan, M. A.; Stranieri, M. T.; Lynch, J. J., Jr.; Lin, J. H.; Chen, I.-W.; Vastag, K.; Naylor-Olsen, A. M.; Vacca, J. Discovery and Development of the Novel Potent Orally Active Thrombin Inhibitor *N*-(9-Hydroxy-9-fluorenecarboxy)prolyl *trans*-4-Aminocyclohexylmethyl Amide (L-372,460): Coapplication of Structure-Based Design and Rapid Analogue Synthesis on Solid Support. *J. Med. Chem.* **1998**, *41*, 401−406.

(36) Holladay, M. W.; Wasicak, J. T.; Lin, N.-H.; He, Y.; Ryther, K. B.; Bannon, A. W.; Buckley, M. J.; Kim, D. J. B.; Decker, M. W.; Anderson, D. J.; Campbell, J. E.; Kuntzweiler, T. A.; Donnelly-Roberts, D. L.; Piattioni-Kaplan, M.; Briggs, C. A.; Williams, M.; Arneric, S. P. Identification and Initial Structure−Activity Relationships of (R)-5-(2-Azetidinylmethoxy)-2-chloropyridine (ABT-594), a Potent, Orally Active, Non-Opiate Analgesic Agent Acting via Neuronal Nicotinic Acetylcholine Receptors. *J. Med. Chem.* **1998**, *41*, 407−412.

(37) Sullivan, R. W.; Bigam, C. G.; Erdman, P. E.; Palanki, M. S. S.; Anderson, D. W.; Goldman, M. E.; Ransone, L. J.; Suto, M. J. 2-Chloro-4-(trifluoromethyl)pyrimidine-5-N-(3′, 5′-bis(trifluoromethyl)phenyl)-carboxamide: A Potent Inhibitor of NF-kB- and AP-1-Mediated Gene Expression Identified Using Solution-Phase Combinatorial Chemistry. *J. Med. Chem.* **1998**, *41*, 413−419.

(38) Hopper, A. T.; Witiak, D. T.; Ziemniak, J. Design, Synthesis, and Biological Evaluation of Conformationally Constrained aci-Reductone Mimics of Arachidonic Acid. *J. Med. Chem.* **1998**, *41*, 420−427.

(39) Klein, S. I.; Czekaj, M.; Gardner, C. J.; Guertin, K. R.; Cheney, D. L.; Spada, A. P.; Bolton, S. A.; Brown, K.; Colussi, D.; Heran, C. L.; Morgan, S. R.; Leadley, R. J.; Dunwiddie, C. T.; Perrone, M. H.; Chu, V. Identification and Initial Structure−Activity Relationships of a Novel Class of Nonpeptide Inhibitors of Blood Coagulation Factor Xa. *J. Med. Chem.* **1998**, *41*, 437−450.

(40) Davies, D. J.; Garratt, P. J.; Tocher, D. A.; Vonhoff, S.; Davies, J.; Teh, M. T.; Sugden, D. Mapping the Melatonin Receptor. 5. Melatonin Agonists and Antagonists Derived from Tetrahydrocyclopent[b]indoles, Tetrahydrocarbazoles and Hexahydrocyclohept[b]indoles. *J. Med. Chem.* **1998**, *41*, 451−467.

(41) Sugihara, H.; Fukushi, H.; Miyawaki, T.; Imai, Y.; Terashita, Z.; Kawamura, M.; Fujisawa, Y.; Kita S. Novel Non-Peptide Fibrinogen Receptor Antagonists. 1. Synthesis an Glycoprotein IIb-IIIa Antagonistic Activities of 1,3,4-Trisubstituted 2-Oxopiperazine Derivatives Incorporating Side-Chain Functions of the RGDF Peptide. *J. Med. Chem.* **1998**, *41*, 489−502.

(42) Terashima, I.; Kohda, K. Inhibition of Human O6-Alkylguanine-DNA Alkyltransferase and Potentiation of the Cytotoxicity of Chloroethylnitrosourea by 4(6)-(Benzyloxy)-2,6(4)-diamino-5-(nitro or nitroso)pyrimidine Derivatives and Analogues. *J. Med. Chem.* **1998**, *41*, 503−508.

(43) Ramesh, M.; Matowee, W. C.; Akula, M. R.; Vo, D.; Dagnino, D. V.; Li-Kwong-Ken, M. C.; Wolowyk, M. W.; Knaus, E. E. Synthesis and Calcium Channel-Modulating Effects of Alkyl (or Cycloalkyl) 1,4-Dihydro-2,6-dimethyl-3-nitro-5-pyridinecarboxylate Racemates and Enantiomers. *J. Med. Chem.* **1998**, *41*, 509−514.

(44) Abe, Y.; Kayakiri, H.; Satoh, S.; Inoue, T.; Sawada, Y.; Imai, K.; Inamura, N.; Asano, M.; Hatori, C.; Katayama, A.; Oku, T.; Tanaka, H. A Novel Class of Orally Active Non-Peptide Bradykinin B2 Receptor Antagonists. 1. Construction of Basic Framework. *J. Med. Chem.* **1998**, *41*, 564−578.

(45) Pevarello, P.; Bonsignori, A.; Dostert, P.; Heidempergher, F.; Pinciroli, V.; Colombo, M.; McArthur, R. A.; Salvati, P.; Post, C.; Fariello, R. G.; Varasi, M. Synthesis and Anticonvulsant Activity of a New Class of 2-[(Arylalkyl)amino]alkanamide Derivatives. *J. Med. Chem.* **1998**, *41*, 579−590.

(46) Tamura, Y.; Watanabe, F.; Nakatani, T.; Yasui, K.; Fuji, M.; Komurasaki, T.; Tsuzuki, H.; Maekawa, R.; Yoshioka, T.; Sugita, K.; Ohtani, M. Highly Selective and Orally Active Inhibitors of Type IV Collagenase (MMP-9 and MMP-2): N-Sulfonylamino Acid Derivatives. *J. Med. Chem.* **1998**, *41*, 640−649.

(47) Forbes, I. T.; Dabbs, S.; Duckworth, D. M.; Jennings, A. J.; King, F. D.; Lovell, P. J.; Brown, A. M.; Collin, L.; Hagan, J. J.; Middlemiss, D. N.; Riley, G. J.; Thomas, D. R.; Upton, N. (R)-3, N-Dimethyl-N-[1-methyl-3-(4-methyl-piperdin-1-yl)propyl]benzenesulfonamide: The First Selective 5-HT7 Receptor Antagonist. *J. Med. Chem.* **1998**, *41*, 655−657.

(48) Oshiro, Y.; Sato, S.; Kurahashi, N.; Tanaka, T.; Tottori, K.; Uwahodo, Y.; Nishi, T. Novel Antipsychotic Agents with Dopamine Autoreceptor Agonist Properties: Synthesis and Pharmacology of 7-[4-(4-Phenyl-1-piperazinyl)butoxyl]-3,4-dihydro-2(1H)-quinoline Derivatives. *J. Med. Chem.* **1998**, *41*, 658−667.

(49) Betti, M.; Biagi, G.; Giannaccini, G.; Giorgi, I.; Livi, O.; Lucacchini, A.; Manera, C.; Scartoni, V. Novel 3-Aralkyl-7-(amino-substituted)-1,2,3-triazolo[4,5-d]pyrimidines with High Affinity toward A1 Adenosine Receptors. *J. Med. Chem.* **1998**, *41*, 668−673.

(50) Barlocco, D.; Cignarella, D.; Tondi, D.; Vianello, P.; Villa, S.; Bartolini, A.; Ghelardini, C.; Galeotti, N.; Anderson, D. J.; Kuntzweiler, T. A.; Colombo, D.; Toma, L. Mon- and Disubstituted-3,8-diazabicyclo[3.2.1]octane Derivatives as Analgesics Structurally Related to Epibatidine: Synthesis, Activity, and Modeling. *J. Med. Chem.* **1998**, *41*, 674−681.

(51) Dutta A. K.; Coffey L. L.; Reith M. E. Potent and Selective Ligands for the Dopamine Transporter (DAT): Structure−Activity Relationship Studies of Novel 4-[2-(Diphenylmethoxy)-ethyl]-1-(3-Phenylpropyl)piperidine Analogues. *J. Med. Chem.* **1998**, *41*, 699−705.

(52) Cappelli, A.; Anzini, M.; Vomero, S.; Mennuni, L.; Makovec, F.; Doucet, E,; Hamon, M.; Bruni, G.; Romeo, M. R.; Menziani, M. C.; De Benedetti, P. G.; Langer, T. Novel Potent and Selective Central 5-HT3 Receptor Ligands Provided with Different Intrinsic Efficacy. 1. Mapping the Central 5-HT3 Receptor Binding Site by Arylpiperadine Derivatives. *J. Med. Chem.* **1998**, *41*, 728−741.

(53) Rewcastle, G. W.; Murray, D. K.; Elliot, W. L.; Fry, D. W.; Howard, C. T.; Nelson, J. M.; Roberts, B. J.; Vincent, P. W.; Showalter, H. D. H.; Winters, R. T.; Denny, W. A. Tyrosine Kinase Inhibitors. 14. Structure−Activity Relationships for Methylamino-Substituted Derivatives of 4-[(3-Bromophenyl)-amino]-6−9methylamino)-pyrido[3,4-d]pyrimidine (PD 158780), a Potent and Selective Inhibitor of the Tyrosine Kinase Activity of Receptors for the EGF Family of Growth Factors. *J. Med. Chem.* **1998**, *41*, 742−751.

(54) McKittrick, B. A.; Ma, K.; Huie, K.; Yumibe, N.; Davis, H.; Clader, J. W.; Czarniecki; McPhail, A. T. Synthesis of C3 Heteroatom-Substituted Azetidinones That Display Potent Cholesterol Absorption Inhbitory Activity. *J. Med. Chem.* **1998**, *41*, 752−759.

(55) Patane, M.; Scott, A. L.; Broten, T. P.; Chang, R. S. L.; Ransom, R. W.; DiSalvo, J.; Forray, C.; Bock, M. G. 4-Amino-2-[4-[1-(benzyloxycarbonyl)-2(S)-[[(1,1-dimethylethyl)amino]carbonyl]-piperazinyl]-6,7-dimethoxyquinazoline (L-765, 314): A Potent and Selective alpha1B Adrenergic Receptor Agonist. *J. Med. Chem.* **1998**, *41*, 1205−1208.

(56) Gaster, L. M.; Blaney, F. E.; Davies, S.; Duckworth, D. M.; Ham, P.; Jenkins, S.; Jennings, A. J.; Joiner, G. F.; King, F. D.; Mulholland, K. R.; Wyman, P. A.; Hagan, J. J.; Hatcher, J.; Jones, B. J.; Middlemiss, D. N.; Price, G. W.; Riley, G.; Roberts, C.; Routledge, C.; Selkirk, J.; Slade, P. The Selective 5-HT1B Receptor Inverse Agonist 1′-Methyl-5-[[2′-methyl-4′-(5-methyl-1,2,4-oxadiazol-3-yl)biphenyl-4-yl]carbonyl]-2,3,6,7-tetrahydrospiro[furo[2,3-f]indole-3,4′-piperidine] (SB-224289) Potently Blocks Terminal 5-HT Autoreceptor Function Both in Vitro and in Vivo. *J. Med. Chem.* **1998**, *41*, 1218−1235.

(57) Porcari, A. R.; Devivar, R. V.; Kucera, L. S.; Drach, J. C.; Townsend, L. B. Design, Synthesis, and Antiviral Evaluations of 1-(Substituted benzyl)-2-substituted-5,6-dichlorobenzimidazoles as Nonnucleoside Analogues of 2,5,6-Trichloro-1-(beta-D-ribofuranosyl)benzimidazole. *J. Med. Chem.* **1998**, *41*, 1252−1262.

(58) Gangjee, A.; Guo, X.; Queener, S. F.; Cody, V.; Galitsky, N.; Luft, J. R.; Pangborn, W. Selective *Pneumocystis carinii* Dihydrofolate Reductase Inhibitors: Design, Synthesis, and Biological Evaluation of New 2,4-Diamino-5-Substituted-furo[2,3-d]pyrimidines. *J. Med. Chem.* **1998**, *41*, 1263−1271.

(59) Grese, T. A.; Pennington, L. D.; Sluka, J. P.; Adrian, M. D.; Cole, H. W.; Fuson, T. R.; Magee, D. E.; Phillips, D. L.; Rowley, E. R.; Shetler, P. K.; Short, L. L.; Venugopalan, M.; Yang, N. N.; Sato, M.; Glasebrook, A. L.; Bryant, H. U. Synthesis and Pharmacology of Conformationally Restricted Raloxifene Analogues: Highly Potent Selective Estrogen Receptor Modulators. *J. Med. Chem.* **1998**, *41*, 1272−1283.

(60) Sekiyama, T.; Hatsuya, S.; Tanaka, Y.; Uchiyama, M.; Ono, N.; Iwayama, S.; Oikawa, M.; Suzuki, K.; Okunishi, M.; Tsuji, T. Synthesis and Antiviral Activity of Novel Acyclic Nucleosides: Discovery of a Cyclopropyl Nucleoside with a Potent Inhibitory Activity against Herpesviruses. *J. Med. Chem.* **1998**, *41*, 1284−1298.

JM000003M