DIVSEL and COMPLIB - Strategies for the Design and Comparison of Combinatorial Libraries using Pharmacophoric Descriptors

S. D. Pickett,*,† C. Luttmann, V. Guerin, A. Laoui, and E. James

Rhône-Poulenc Rorer S. A., Centre de Recherche de Vitry-Alfortville, 13 quai Jules Guesde, B. P. 14-94403 Vitry sur Seine, Cedex, France

Received August 4, 1997

Screening synthetic combinatorial libraries may facilitate rapid drug lead discovery by substantially increasing the number of molecules tested. Drug discovery efficiency and productivity can be further improved by designing libraries to maximize their molecular diversity or by comparing them to existing collections of compounds and/or libraries to select those that complement the properties already well represented. In this paper we describe two strategies to aid in the design and comparison of combinatorial libraries. The methods employ multi-pharmacophore three-dimensional (3D) descriptors in combination with two recent proposals for dissimilarity-based compound selection and library comparison. This method allows the design to be performed in product space and library comparison to consider all pair-wise intermolecular contributions to the diversity.

1. INTRODUCTION

Combinatorial chemistry coupled with high-throughput screening offers the possibility of rapidly synthesizing and testing large numbers of structures.[1] For reasons of screening efficiency and limited resources, it is not possible or desirable to synthesize and test all structures that may be generated by a given chemistry. In addition, as with more traditional approaches to drug discovery, it is still necessary to set clear objectives when choosing the reagents, core groups, etc., that will make up the final library. In particular, when the library is to be used for general screening, design is needed to ensure that the proposed library covers the property space of potential products and also that the library complements properties already well represented in available screening sets. The challenge facing computational chemists is to aid chemists in rationally designing the libraries to meet these objectives. The descriptors necessary to achieve the design should represent key aspects of intermolecular interactions and should be applicable to a wide range of tasks, such as designing libraries focused to a particular receptor or a set of disparate leads. Several different methods for describing property space have been suggested for both library design and compound selection from large databases (see below). The idea of a pharmacophore is widely used to rationalize and aid in the drug design process, and pharmacophore searching has achieved many successes in lead generation.[2,3] Recently, methods enabling a systematic analysis of all potential pharmacophores in a molecule have been developed,[4,5] which opens the way to using such descriptors in a general way for library design. In this paper, we explore the use of these pharmacophore descriptors in combination with two recent proposals for efficient maximum dissimilarity based compound selection and library comparison. We show, through examples from our own work, that pharmacophore descriptors provide a possible means for addressing some of the objectives just outlined.

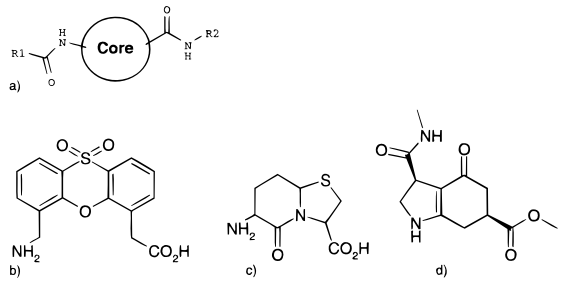

Figure 1. (a) General scheme for libraries for a peptidomimetic approach. (b−d) Example core structures: (b) and (c) are β-turn mimics [Feigel,[31] Nagai & Sato[32]]; (d) is an α-helix nucleator [Seto & Bartlett[33]].

2. LIBRARY DESIGN STRATEGIES

The peptidomimetic approach involves the rigidification or replacement of backbone amide units using scaffolds which, in general, favor a limited number of local conformations; for example, β-turn or α-helix. Such an approach has been used with some success in-house.[6,7] Figure 1a represents schematically a simple two-component library based on amide chemistry that may be used in peptidomimetic design. The core could represent any of a range of aminocarboxylic scaffolds such as those shown in Figure 1b. Synthesizing such libraries using amino acid R-groups provides a pool of potential peptidomimetic compounds. However, ultimately the goal must be to move away from the peptidic nature of these compounds. With this objective in mind, the strategies described in this paper were developed for two purposes. Firstly, to aid in the selection of a set of non-α-aminocarboxylic acids for use with a number of proprietary aminocarboxylic acid scaffolds. Secondly, to be able to select between potential scaffolds using the information from libraries that already exist in-house.

* To whom all correspondence should be addressed. E-mail stephen.pickett@rp-rorer.co.uk.

† Current address: Rhône-Poulenc Rorer, Dagenham Research Centre, Rainham Road South, Dagenham, Essex, RM10 7XS, U.K.

S0095-2338(97)00060-7 CCC: $15.00
Published on Web 01/30/1998

The design of a library is dependent first and foremost on the descriptors used in the design. A number of descriptors have been proposed for this purpose including two-dimensional (2D) fragment based descriptors,[8] physicochemical properties,[9] BCUTs,[10] steric fields,[11] combinations of descriptors[12] analyzed using factor analysis, and three-dimensional (3D) similarity between R-groups based on weighted superpositions of conformers.[13] Where 3D properties have been used they have generally been limited to a single or small ensemble of conformations. An exception is the recent work of Kick et al.[14] on the design and synthesis of libraries directed towards finding inhibitors of cathepsin D, where the crystal structure of a bound inhibitor was used as the starting point for design.

The second choice that needs to be made is whether to design on the individual reagents or the final products. Most current methods have tended towards the first approach whereas ideally the second appears more intuitive. It is true that certain properties will be fairly additive across R-groups; however, this is certainly not the case for 3D properties, particularly when conformational flexibility is taken into account. Selecting individual reagents out of context carries with it the implicit assumption that all compounds in a library will bind in the same way with respect to the core structure. The trade off here is in computational complexity, the number of products scaling as the product of the number of reagents at each position. In addition, inclusion of 3D properties requires the consideration of conformational flexibility. The advantage of methods based on the products is that they are applicable to other compound collections; for example, evaluating the similarity between libraries and designing focused libraries towards protein structures.

Several methods have been proposed for reagent selection. For example, Martin et al.[12] have used a D-optimal design strategy in reagent property space. Maximum dissimilarity based selection methods have also been proposed.[15] Partition based methods[9,10,16] select one or more representatives from each occupied cell. Similarly, clustering selects representatives from each cluster (usually the centroid). In principle, such methods can be applied to either reagent or product selection, but in the latter case the combinatorial nature of the design needs to be considered.

In this paper we explore novel strategies for library design that are based on the use of maximum dissimilarity based selection procedures for the final products in combination with the recently proposed multi-pharmacophore descriptors.[4,5] These methods were selected to satisfy the design objectives stated at the start of this section. Reagent selection is product based and the pharmacophore descriptors encode important 3D information about the molecules. The accessible pharmacophores for a molecule are calculated from a systematic conformational analysis, aspects of shape are implicit in the descriptor, and the pharmacophore group definitions reflect the crucial molecular properties in ligand−receptor binding; acids, bases, aromatic rings, H-bond donors and acceptors, and hydrophobic groups. In addition, pharmacophore space provides a consistent frame of reference for all types of design work, such as combinatorial libraries or subset selection from corporate databases. The identification of missing or poorly occupied regions of diversity space is straightforward and so is the comparison between different compound collections (at least at a preliminary level).

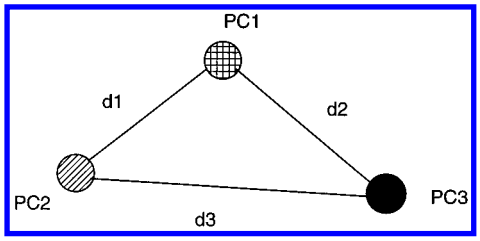

Figure 2. Illustration of pharmacophore profiling with three pharmacophore points: PC1, PC2, and PC3 are the pharmacophore points and d1, d2, and d3 are the respective distances between them. Each point represents one of H-bond donor, H-bond acceptor, H-bond acceptor & donor (e.g., hydroxyl), aromatic center, hydrophobe center, acid, basic nitrogen. The distances are binned into 20 ranges from 0 to 24 Å. The pharmacophore key bit string represents all valid pharmacophore point and distance combinations and marks the presence or absence of a pharmacophore in a molecule.

Focusing of libraries, for example towards pharmacophores commonly expressed by a particular class of inhibitors,[16,17] can also be envisaged. A novel procedure for the preclassification of reagents using general functional group counts has been implemented. This simple procedure ensures that the "simplest" reagents (as defined later) can be considered first and is easily incorporated into a maximum dissimilarity based selection algorithm.

Comparing libraries to existing compound collections is important to ensure that the libraries are complementing existing screening sets. The pharmacophore descriptors are essentially partitioning the space and so, in principle, comparing two populations is straightforward (i.e., does a set of compounds add additional pharmacophores to those already covered?). This process ignores contributions to the pharmacophore space from individual molecules as well as the pharmacophore frequency count; that is, how many times each pharmacophore is hit in a population. Turner et al.[18] published one possible method for evaluating molecular diversity between different compound libraries based on the rapid calculation of intermolecular similarities. The method includes the count information implicitly, and we describe the use of this measure to choose between a number of potential scaffolds based on the pharmacophoric similarity of products.

In the following section we describe in more detail the methods used in this work. Following this, two specific problems from our group are used to demonstrate their application. In the first of these the goal was to select a set of acids to be used in general libraries. The second example shows how interlibrary comparison can be used to choose between potential core structures.

3. METHODS

Calculation of Pharmacophore Descriptors. The ChemDiverse module of Chem-X[19] was used to calculate the three-point pharmacophore descriptors used in this work. A schematic illustration of the method is given in Figure 2. A total of seven pharmacophore features have been considered (hydrogen bond donor, hydrogen bond acceptor, hydrogen bond donors and acceptors - imidazole N/NH etc., acid, base, aromatic center, and hydrophobe) when defining the pharmacophores. These features are assigned within Chem-X using a powerful parameterization technique whereby each

atom is parameterized according to its immediate environment. Dummy atoms are placed at the center of all aromatic rings and dummy atoms representing hydrophobic groups are placed based on electronegativity differences between bonded atoms. The features as defined have been modified from the defaults available in Chem-X and aspects of our own in-house customization of this parameterization have been given elsewhere.[4] The pharmacophore descriptor is a binary string where each bit represents a pharmacophore generated by combining any three of seven features using up to 32 distance ranges. This gives a total of 892 038 theoretically accessible pharmacophores (bits). Once calculated, each molecule descriptor (pharmacophore key) can be saved on disk. In-house routines have been written using ChemLib to reformat these keys for use with the in-house programs described below.

The 3D structures of products (as MACCS SDfiles[20]) were generated by CONCORD[21] from the SMILES[22] and loaded into a Chem-X database. In-house programs have been developed to generate SMILES for the products from the individual R-groups and core structures.[23] Parameterization takes place as the SDfiles are loaded into the Chem-X database. A pharmacophore key is calculated for each product molecule using a systematic conformational analysis to explore the conformational space. The following bond increments are used by default, though more rigorous sampling can be used if required: sp^3-sp^3, 3 points; sp^3-sp^2, 4 points; conjugated bonds, 2 points. A CPK (3/5 VDW radius) bump-check is used to screen out undesirable conformations. For very flexible molecules, which would not complete a systematic analysis within a given time, a random conformational analysis is performed using the same torsional increments. These calculations are controlled using scripts written in-house.[24] Chem-X also provides the possibility to apply "volume" and accessibility checks to the pharmacophores. The volume check removes from consideration smaller pharmacophores within large molecules, for example the two aromatic nitrogens and aromatic centroid of a substituted pyrimidine. The accessibility check ensures that H-bond donor and acceptor groups have their hydrogens or lone pairs in an orientation away from the center of the pharmacophore (i.e., less likely to be sterically hindered by the rest of the molecule for receptor interaction). For both these checks, default settings in Chem-X have been used.

Interlibrary Comparisons. Turner et al.[18] have suggested a method for evaluating the overall similarity between two libraries based on the sum of intermolecular similarities. This method uses an *O(N)* centroid algorithm for calculation of the intermolecular similarities. The I^{th} element of centroid *Ac* for a collection of molecules is given by eq 1:

$$Ac(I) = \sum_{J=1}^{N(A)} W(J)xM(J,I) \qquad (1)$$

where $M(J,I)$ is the I^{th} element of the vector (here pharmacophore bit string) for molecule J with $W(J)$ being the weight of the vector J. If the weights are set according to eq 2:

$$W(J) = 1/\sqrt{\sum_{I=1}^{C} M(J,I)^2} \qquad (2)$$

then it can be shown that the dot product of the centroids for two molecule collections A and B is given by eq 3, the sum of the individual cosine coefficients:

$$Ac \cdot Bc = \sum_{J=1}^{N(A)} \sum_{K=1}^{N(B)} COS(J,K) \qquad (3)$$

In other words, if the cosine coefficient is used to estimate similarity, then the sum of all $N(A) \times N(B)$ intermolecular similarities can be calculated from the dot product of the two vector centroids. Now, defining the internal diversity of A, $D(A)$ as $1 - Ac \cdot Ac/N(A)^2$, then Turner et al. suggest that the change in diversity of A, $\delta(A)$, as a result of adding a set of compounds X is given by:

$$\delta(A) = D(AX) - D(A) = Ac \cdot Ac/N(A)^2 - AXc \cdot AXc/(N(A) + N(X))^2 \qquad (4)$$

with $AXc \cdot AXc$ given by:

$$AXc \cdot AXc = Ac \cdot Ac + Xc \cdot Xc + 2xAc \cdot Xc \qquad (5)$$

We have implemented this approach in an in-house program COMPLIB, which uses the pharmacophore bit strings as descriptors. The use of such a measure permits the evaluation of a number of proposed libraries against libraries that have been synthesised previously by calculating the change in diversity, $\delta(A)$, on adding a library to those already synthesized. The results obtained with this measure have been compared with the similarity calculated from the pharmacophore OR key for each of the libraries.

Maximum Dissimilarity Based Selection. Holliday et al.[15] have recently published an efficient algorithm for dissimilarity based selection that uses the same centroid algorithm as that just described to calculate the total intermolecular similarities. The selected subset is seeded with the molecule that is most dissimilar to all the others in the dataset. The next molecule chosen is that which is most dissimilar to the selected subset and so on until the required number of molecules has been selected. This method has been implemented in an in-house program DIVSEL, which uses the pharmacophore keys as the descriptor for similarity calculations. There is an option to select molecules only if they have a similarity less than a user-defined value with each of the molecules previously selected. The total number of pharmacophores covered by the subset at each step and the number of new pharmacophores added by each molecule are also output. This aspect aids in the final selection of reagents, starting from the diverse ordered list. It is also possible to input a starting list of reagents that is used to seed the initial selected subset.

Reagent Functional Group Classification. Reagents were identified from the ACD version 95.1[25] or in-house databases using 2D searching with ISIS/Base.[26] Routines to classify reagents according to structural/chemical complexity were written using GENIE,[27] where the SMARTS language is used to define queries for searching the reagent SMILES. These routines are interpreted using programs written with the Daylight Toolkit. Previous work in-house had used such procedures to produce hydrogen bond donor and acceptor counts.[9] These scripts were modified and extended for our purpose. Counts per reagent are output

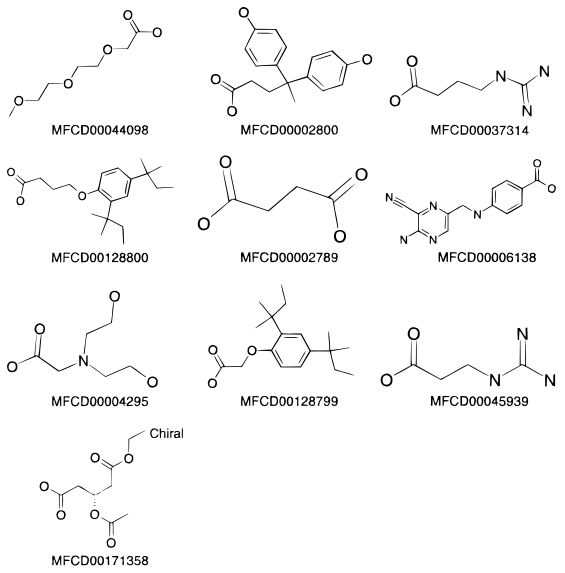

Figure 3. First 10 acids selected by DIVSEL run 1.

for general functional group definitions including, amongst others, acids, aromatic rings, nitrogen-containing aromatics, "simple" hydrogen-bonding groups like carbonyl oxygen (excluding amides), number of four-substituted carbons, and bi- or trifunctional groups, like amides, ureas, etc. This represents a generalized 2D functional group count for each molecule that is subsequently analyzed to classify the reagents into sets of (generally) increasing "complexity" for example, those reagents containing only one functional group class (one of simple hydrogen bond acceptor, base, multifunctional group, etc.) in addition to the reactive acid, those reagents containing more than one acid function, those reagents containing more than one basic function, and molecules with no more than two aromatic rings. These classifications are used to prioritize the reagents for selection.

4. RESULTS

Product-Based Reagent Selection. A set of amines for R2 in Figure 1 had been selected previously. To reduce the size of the virtual combinatorial library, each amine was attached to a representative scaffold and the pharmacophore key was calculated. A pharmacophorically diverse set of 11 amines was selected using DIVSEL. Then, 1100 monocarboxylic acids were identified from 2D searching, and the pharmacophore keys were calculated for the 1100 subgroups involving the products of each acid in combination with the 11 amines (a total of 12 100 products). It is these 1100 subgroup keys that are used in the selection by DIVSEL. During the selection a similarity cutoff was applied so that a key would not be selected if it had a similarity >0.85 to a key already selected. In addition, once the required number of selections has been made, the similarity with the remaining keys can be calculated, providing potential lists of replacement compounds should a selected compound be unavailable or be incompatible with the reaction. After 100 rounds of selection, 85% of the pharmacophore space of all 12 100 products had been covered. Figure 3 shows the first 10 acids selected. Although the molecules are diverse, it is clear that some would not be suitable for a general library. This observation led us to develop a stepwise strategy where the reagents were first prioritized by structural/chemical com-

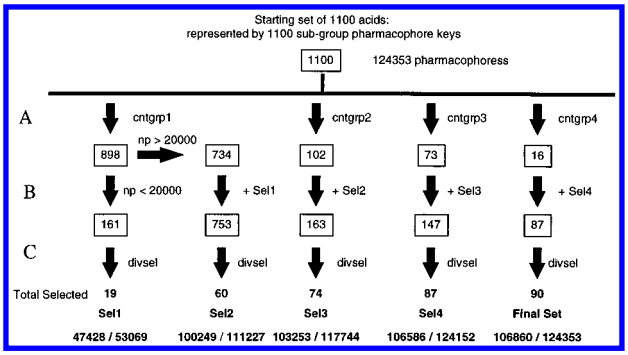

Figure 4. Schematic of DIVSEL run 2. Step A: the 1100 acids are classified into different sets (cntgrp1, cntgrp2, etc.) of increasing structural complexity (see text). cntgrp1 is further divided into two groups based on the number of pharmacophores (*np*) in the pharmacophore key representing each acid in combination with the 11 amines at R2. Step B: Any previously selected are added to the working set to constitute the starting list. Step C: A diverse set is selected using DIVSEL. The two numbers at the bottom of each column give the number of pharmacophores in the selected set and those in all molecules considered to date, respectively.

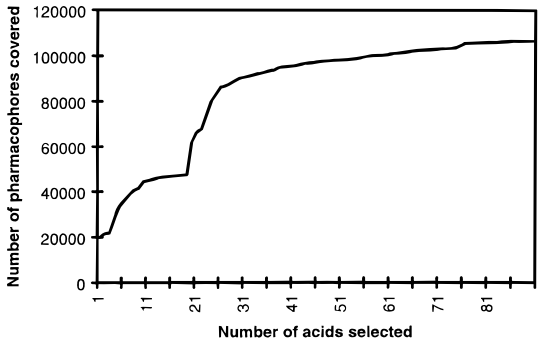

Figure 5. Pharmacophore coverage as a function of the number of acids selected, using the procedure described in the text.

plexity as described in the *Methods Section*. The overall process is shown in Figure 4. In step A, the reagents are preclassified, using the functional group counts, into different working sets of increasing "structural complexity" (cntgrp1 to cntgrp4). The initial set, cntgrp1, was defined as those reagents containing one functional group class (one of base, simple acceptor, multifunctional group, etc.) in addition to the reactive acid. In addition, this first set was split into two according to the number of pharmacophores in the product subgroup key. In step B, the starting list of previously selected compounds are added to the working set (two reagents were identified to start the design, acetic acid and benzoic acid, and hence are added to cntgrp1). In step C, DIVSEL is run on the working set using the starting list to seed the initial selection. At each stage, the use of the previously selected reagents to define a starting list ensures that keys are selected that are diverse with respect to those previously selected. Figure 5 shows how the total pharmacophore coverage of the selected acids increases in stepwise fashion as each reagent set is considered. The final set of 90 acids covers $>85\%$ of the pharmacophore space of all 12 100 products (note that the pharmacophore key and hence the count is calculated for each acid in combination with 11 amines at R2).

Library Comparison. For any given chemistry it is generally possible to purchase or synthesize a number of potential scaffolds or core groups, in our case aminocarboxylic acids. Our choice for evaluating the potential for

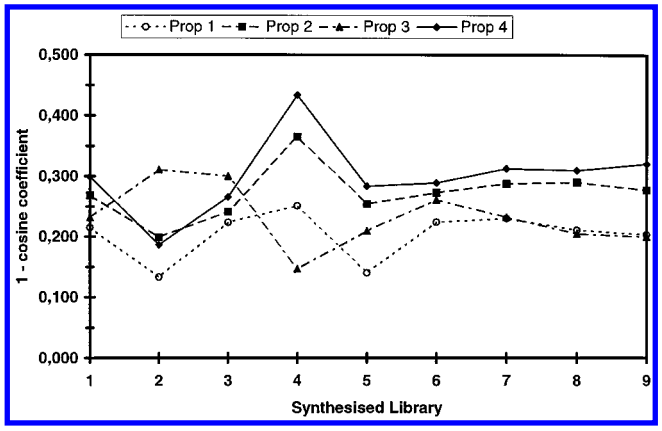

Figure 6. (1 - cosine coefficient) calculated from the library pharmacophore bit strings for a number of propositions with a set of libraries previously synthesized.

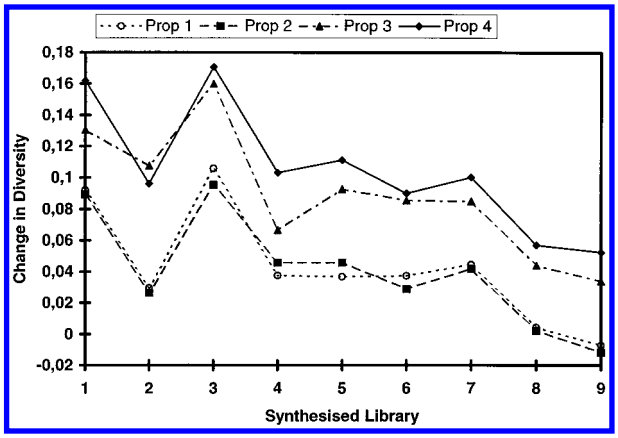

Figure 7. Change in diversity on adding a proposed library to one already synthesized, calculated using the measure of Holliday et al.[15]

new scaffolds has been to compare them with those already in use, by analyzing the influence of the scaffolds on the pharmacophore space covered when combined with a set of reagents. This can be done in several ways. At a library level, the total pharmacophore space covered by the libraries can be compared using the OR keys for each library. Such a comparison is shown in Figure 6 where the product OR keys calculated for four potential scaffolds have been compared with nine libraries already synthesized in-house. The cosine coefficient calculated from the library OR keys is plotted for each proposed library against a synthesized library. At a molecule level, the measure suggested by Turner et al.[15] can be used, based on the intermolecular similarities. The results for the same set of potential scaffolds are shown in Figure 7. For these calculations, the pharmacophore keys were calculated for each product and the change in diversity on adding each proposition to an already synthesized library calculated using the centroid algorithm (eq 4). With this measure, the more positive the *y*-value, the more interesting is the proposition with respect to an already synthesized library. From the example it can be seen that proposition 4 would be the most interesting to synthesize.

5. DISCUSSION

The results just presented show how it is possible to use pharmacophore-based descriptors to select reagents for combinatorial libraries while taking account of the pharmacophores that exist in the final products. This feature is important as it is the final products that will be tested. The more specific design objectives for a general library are twofold; firstly, to maximize the pharmacophore coverage of the library (i.e., coverage of property space), and secondly, to minimize the redundancy between compounds. It will generally be possible to identify a subset of flexible or highly functionalized molecules that cover the majority of pharmacophore space of the library, but such "promiscuous" molecules could lower the binding compared with a compound exhibiting the desired pharmacophore more succinctly. The second design objective tends to focus the design towards molecules that cover fewer number of pharmacophores. The use of the "volume" option is also important to ensure that smaller pharmacophore distances are not covered by larger molecules.

As stated in the *Introduction*, any design is critically dependent on the descriptors used. A pharmacophore represents key aspects of ligand−receptor interactions, encoding for hydrogen bond donor/acceptor ability, for example, and to some extent shape. This latter aspect could be better represented by moving from the use of three-point to four-point pharmacophores (a plane to a tetrahedron) and, indeed, recent work has shown improvements on using a four-point pharmacophore.[16] The definition of the pharmacophore key (distance ranges between pharmacophore points) is to some extent dependent on the conformational sampling and this is another area of current research, larger bin ranges being preferable at longer distances.[16] The advantages of pharmacophore key descriptors over other 3D measures are that no molecular alignment is necessary and a very large number of accessible conformations are probed. However, this means that the pharmacophore key descriptor pools information from all accepted conformations; thus, there is a need for design objectives, as set out in the previous paragraph; which focus on partitioning rather than clustering with the pharmacophore keys.

The design was improved by the sensible preclassification of reagents using fairly simple functional group type descriptors. It should be noted that this is not a problem with the pharmacophore descriptors themselves, but rather a reflection of the large structural diversity of the available reagents. Indeed, similar behavior has recently been observed by other workers using real-valued calculated molecular properties in combination with a maximum dissimilarity method to select compounds from large databases.[28] In the strategy proposed here, a complex reagent is not rejected, it is simply considered later in the design process and could still be chosen if it adds to the available pharmacophore space. The ability to proceed in stages is a useful feature of the maximum dissimilarity based selection methods and the additional output, such as number of pharmacophores, added by a particular product (or subset of products) can be important in making the final selection. Thus, the final set of reagents is not chosen by just selecting the top *n* reagents from the output. Graphs such as Figure 5 are inspected to see when additional reagents are adding only very few pharmacophores (the graph levels off) and the individual contributions of molecules are examined. In effect, this method is combining the partitioning aspects ofthe pharma-

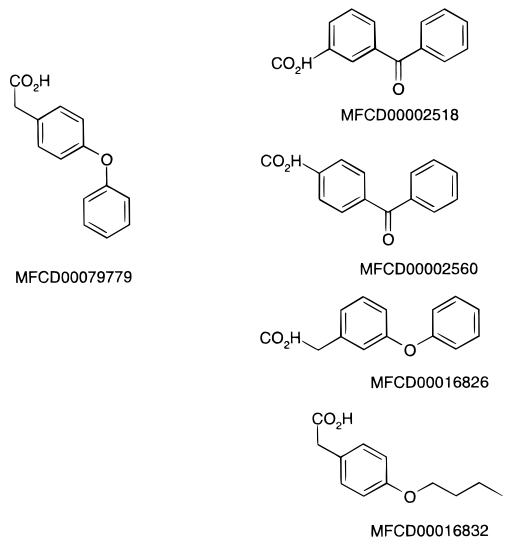

Figure 8. Selected reagent and potential replacements based on pharmacophore similarity.

cophore descriptors with a distance-based selection procedure.

It is sometimes the case that a particular reagent is not available or does not react sufficiently. In this case it is necessary to suggest replacements. A feature has been added to DIVSEL that automatically calculates the similarity between each selected reagent and those that have not been selected (remembering that in this case this means calculating the similarity between pharmacophore keys representing each acid combined with the 11 amines). No clustering is carried out on this data so that at a particular level of similarity a reagent may be similar to more than one selected reagent. Similarly, the reagents in the replacement list need not be similar to each other at the selected cutoff. These lists provide a means for the chemist to select an alternative without greatly altering the design of the final library. Figure 8 shows a selected reagent and four similar reagents. These similarities have been calculated from the product pharmacophore keys from the 11 representative amines and so are only relevant within the context of this particular chemistry/library. There is some degree of structural variation within the set, in particular the alkyl ether. This variation can be understood because the aromatic rings are also deemed as hydrophobic; hence, the considerable pharmacophore overlap between this structure and the structure selected is not surprising.

In the example just given, there are two R-group positions, one of which has been predefined; thus, the combinatorial nature of the library has been considered by using supermolecule pharmacophore keys where each key represents an ensemble of products at one of the positions. This procedure is, of course, not always feasible where significant numbers of reagents need to be considered at each position. In this instance, the methods just described could be useful to reduce the number of reagents, for example, by profiling "product-like" reagents (e.g., adding the core) and selecting from these keys. We have also investigated iterative procedures for product-based reagent selection whereby several cycles of selection are made. However, more rigorous approaches using simulated annealing or genetic algorithms are currently under development for use with the pharmacophore descriptors[29] that could be used on the reduced reagent sets. Such methods allow for the inclusion of additional properties in combination with pharmacophore descriptors.[30]

The comparison of proposed libraries to those already synthesized is an important aspect that can not be ignored when designing libraries. We have explored a recent proposal by Turner et al.[15] to tackle this problem. Comparing Figures 6 and 7, some similarities but also important differences are seen that highlight the need to consider the contribution of individual molecules to the diversity. In other words, two libraries could be covering similar regions of pharmacophore space but the individual molecules would be expressing the pharmacophores in different combinations. In generating these graphs, only the core has been changed and the same set of R-groups were used in each case. The differences between the different libraries highlights the need to consider the whole molecule properties rather than just reagents in the design.

The two approaches to library comparison are complementary to each other. The comparison of whole library pharmacophore keys to some extent preserves the partitioning aspects of the pharmacophore descriptors. The second method is a distance-based approach and implicitly includes the individual molecule information. The added value of proposition 4 is most strongly emphasized in Figure 7 where the individual molecule contributions have been compared. Thus, not only does the library show some differences to those already synthesized at the whole library level (comparison of library OR keys), new pharmacophores are being added, but the individual molecules also tend to exhibit the pharmacophores in a different way. This result may be compared with proposition 2 where the individual molecules tend to show a greater similarity to those already synthesized, even if Figure 6 suggests that from the whole library descriptor the library may be interesting.

We have noted some problems, however, with the proposed measure when comparing libraries of very different sizes. It can be seen from eq 6 that the size of the libraries enters into the denominator:

$$\delta(A) = Ac{\cdot}Ac/N(A)^2 - Ac{\cdot}Ac/(N(A) + N(X))^2 - (Xc{\cdot}Xc + 2xAc{\cdot}Xc)/(N(A) + N(X))^2 \quad (6)$$

and in the limiting case of $N(A) >> N(X)$, $\delta(A)$ will always tend to be small and/or negative because of the near cancellation of the first two terms. On the other hand, when $N(X) >> N(A)$, the first term in eq 6 will dominate. Thus, comparing between several different libraries of significantly varying sizes could give rise to misleading results. These situations could arise, for example, when comparing libraries of a few hundred structures to the corporate database of several hundred thousand. We are actively looking into other ways to address this problem. However, the method is ideally suited to the type of application presented here.

6. CONCLUSIONS

We have shown how pharmacophore-based 3D molecular descriptors can be used to design and compare combinatorial libraries, considering the properties of the final products. A maximum dissimilarity based selection method using phar-

macophore descriptors has been used to ensure a good coverage of pharmacophore space. The aim of such a library being to generate micromolar leads, we have found it useful to prioritize reagents by structural/chemical complexity for selection. This method ensures that the accessible pharmacophore space is covered where possible by relatively simple products (designs being performed in product space) more suitable for lead generation.

A recent proposed method for library comparison has also been investigated, again in combination with pharmacophore descriptors. Although some problems are noted when comparing libraries of different sizes, the method is particularly useful for selecting between potential core structures.

ACKNOWLEDGMENT

The authors thank Jon Mason for modified Chem-X parameterization and pharmacophore profiling scripts and for many fruitful discussions. Richard Lewis, Iain McLay, Paul Menard, Isabelle Morize, and François Clerc are thanked for helpful discussions.

REFERENCES AND NOTES

(1) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions. *J. Med. Chem.* **1994**, *37*, 1385−1401.

(2) Mason, J. S. Experiences with Searching for Molecular Similarity in Flexible 3D Databases. In *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Blackie Academic and Professional: Glasgow, 1995; Chapter 6, pp 138−162.

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

(4) (a) Pickett, S. D.; Mason, J. S.; McLay, I. M. Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries (PDQ). *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1214−1223; (b) Ashton, M. J.; Jaye, M. C.; Mason, J. S. New Perspectives in Lead Generation Il: Evaluating Molecular Diversity. *Drug Discovery Today* **1996**, *2*, 71−78.

(5) (a) Davies, E. K.; Briant, C. Combinatorial Library Design Using Pharmacophore Diversity. Accessible through URL http://www.awod.com/netsci/Science/Combichem/feature05.html; (b) Davies, K. In *Molecular Diversity and Combinatorial Chemistry. Libraries and Drug Discovery*, Chaiken, I. M.; Janda, K. D., Eds.; American Chemical Society: Washington, D.C., 1996; pp 309−316.

(6) Laoui, A.; Luttmann, C.; Morize, I.; Pantel, G.; Morgat, A.; Rubin-Carrez, C.; Laroche, V.; Guitton, J. D.; Gigonzac, O.; James-Surcouf, E. Molecular Mimics as Approaches for Rational Drug Design: Application to Tachykinin Antagonists. In *New Perspectives in Drug Design*; Dean, P. M.; Jolles, G.; Newton, C. G., Eds.; Academic: London, 1995; Chapter 13, pp 255−283.

(7) Burns, C. J.; Guitton, J-D.; Baudoin, B.; LeLievre, L.; Duchesne, M.; Parker, F.; Fromage, N.; Commercon, A. Novel Conformationally Extended Naphthalene-Based Inhibitors of Farnesyltransferase. *J. Med. Chem.* **1997**, *40*, 1763−1767.

(8) 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.

(9) (a) Mason J. S.; McLay I. M.; Lewis R. A. Applications of Computer-Aided Drug Design Techniques to Lead Generation. In *New Perspectives in Drug Design*; Dean P. M.; Jones G.; Newton C. G., Eds., Academic: London, 1995; Chapter 12, pp 225−253; (b) Lewis, R. A.; Mason, J. S.; McLay, I. M. Similarity Measures for Rational Set Selection and Analysis of Combinatorial Libraries: The Diverse Property-Derived (DPD) Approach. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 599−614.

(10) (a) Pearlman, R. S.; Smith, K. M., manuscript in preparation; (b) Pearlman, R. S. Novel Software Tools for Addressing Molecular Diversity. Accessible through URL http://www.awod.com/netsci/Science/combichem/feature08.html.

(11) Cramer, R. D.; Clark, R. D.; Patterson, D. E.; Fergusson, A. M. Bioisosterism as a Molecular Diversity Descriptor: Steric Fields of Single Topomeric Conformers. *J. Med. Chem.* **1996**, *39*, 3049−3059.

(12) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. Measuring Diversity: Experimental Design of Combinatorial Libraries for Drug Discovery *J. Med. Chem.* **1995**, *38*, 1431−1436.

(13) Chapman, D. The Measurement of Molecular Diversity: A Three-Dimensional Approach *J. Comput. Aided. Mol. Design* **1996**, *10*, 501−512.

(14) Kick, E. K.; Roe, D. C.; Skillman, A. G.; Liu, G.; Ewing, T. J. A.; Sun, Y.; Kuntz, I. D.; Ellman, J. A. Structure-Based Design and Combinatorial Chemistry Yield Low Nanomolar Inhibitors of Cathepsin D. *Chem. Biol.* **1997**, *4*, 297−307.

(15) Holliday J. D.; Ranade, S. S.; Willett, P. A Fast Algorithm for Selecting Sets of Dissimilar Molecules from Large Chemical Databases *Quant. Struct.-Act. Relat.* **1995**, *14*, 501−506.

(16) Mason, J. S.; Pickett, S. D. Partition-Based Selection. In *Perspectives in Drug Discovery and Design (PD3) - Special Issue on Computational Methods for the Analysis of Molecular Diversity*; Willett, P., Ed.; Kluwer Academic, in press.

(17) Morize, I.; Mason, J. S. 213th. National Meeting of the American Chemical Society, *Abstracts of Papers*, San Francisco, April 13−17, 1997.

(18) Turner, D. B.; Tyrell, T. M.; Willett, P. Rapid Quantification of Molecular Diversity for Selective Database Acquisition. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 18−22.

(19) Chem-X developed and distributed by Chemical Design Ltd. Roundway House, Cromwell Park, Chipping Norton, Oxfordshire OX7 5SR, U.K.

(20) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244−255.

(21) Pearlman, R. S. *CDA News* **1987**, *2*, 1−7; Rusinko, A., III; Skell, J. M.; Balducci, R.; McGarity, C. M.; Pearlman, R. S. University of Texas, Austin. CONCORD, A program for the Rapid Generation of High Quality Approximate 3D Molecular Structures. Distributed by Tripos Inc., 1699 Hanley Road, Suite 303, St. Louis, MO 63144.

(22) Weininger, D. SMILES A Chemical Language and Information System. 1. Introduction to Methodology end Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31; Weininger, D.; Weininger, A.; Weininger, J. L.; SMILES. 2. Algorithm for the Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1998**, *29*, 97.

(23) Lewis, R. A., personal communication.

(24) Mason, J. S., personal communication.

(25) The Available Chemical Directory ACD is distributed by Molecular Design Ltd., San Leandro, CA.

(26) ISIS/Base software is distributed by Molecular Design Ltd., San Leandro, CA.

(27) Daylight Software Manual: Theory, Daylight Chemical Information Systems, 27401 Los Altos, Suite #370, Mission Viejo, CA 92691 and Santa Fe, NM 87501.

(28) Higgs, R. E.; Bemis, K. G.; Watson, I. A.; Wikel, J. H. Experimental Designs for Selecting Molecules from Large Chemical Databases. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 861−870.

(29) Lewis, R. A.; Good, A. C.; Pickett, S. D. Quantification of Molecular Similarity and Its Application to Combinatorial Chemistry. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H.; Testa, B.; Folkers, G., Eds.; Wiley-VCH: Basel, 1997; Chapter 9, pp 135−156.

(30) Good, A. C.; Lewis, R. A. New Methodology for Profiling Combinatorial Libraries and Screening Sets. *J. Med. Chem.* **1997**, *40*, 3926−3936.

(31) Feigel. M. 2,8-Dimethyl-4-(carboxymethyl)-6-(aminoethyl)phenoxathiin S-Dioxide: An Organic Substitute for the β-Turn in Peptides? *J. Am. Chem. Soc.* **1986**, *108*, 181−182.

(32) Nagai, U.; Sato, K. Synthesis of a Bicyclic Dipeptide with the Shape of β-Turn Central Part. *Tetrahedron Lett.* **1985**, *26*, 647−650.

(33) Seto, C. T.; Bartlett, P. A. Design and Application of an α-Helix Nucleator. In *New Advances in Peptidomimetics and Small Molecular Design; Abstracts of Papers*, Philadelphia, March 23−25, 1994.

CI970060X