The design of combinatorial libraries using properties and 3D pharmacophore fingerprints

Brett R. Beno and Jonathan S. Mason

Molecular diversity and similarity methods relevant to drug–receptor interactions are key for the design of combinatorial libraries for lead discovery and optimization. BCUT chemistry-space values for ligands have been used for many diversity-related applications and incorporate receptor-relevant properties such as hydrogen bonding and polarizability. Three-dimensional (3D) multiple-point pharmacophore descriptors (fingerprints) quantify diversity (or similarity) in terms of combinations of three or four functional groups associated with non-covalent ligand–receptor binding. BCUTs and pharmacophore fingerprints have been effectively utilized to design diversity libraries, and also show promise for focused library design.

***Brett R. Beno**
Computer Assisted Drug Design
Structural Biology and Modeling
Department of Macromolecular Structure and Biopharmaceuticals
Bristol-Myers Squibb Pharmaceutical Research Institute
Wallingford, CT-06492, USA
*tel: +1 203 677 7812
fax: +1 203 677 7702
e-mail: brett.beno@bms.com
Jonathan S. Mason
Structural Biology and Modeling
Department of Macromolecular Structure and Biopharmaceuticals
Bristol-Myers Squibb Pharmaceutical Research Institute
Princeton, NJ, USA

▼ Combinatorial chemistry allows rapid increases in the size of compound collections to support HTS programs, and rapid exploration of the structure–activity relationships (SAR) around chemotypes of interest in medicinal chemistry programs. To ensure that the products of a combinatorial library possess the desired characteristics, substantial effort is required in the library design process. In view of the large number of potential products that can be synthesized from combinations of available reagents, computational methods able to rapidly evaluate molecular similarity and diversity, calculate physicochemical properties, and select the most suitable reagents based upon the structures of either combinatorial reagents or their products are essential.

This review will focus on combinatorial reagent selection methods that utilize DiverseSolutions™ (DVS; Tripos, St Louis, MO, USA) BCUT descriptors[1–3] and three-dimensional (3D) multiple-point pharmacophore descriptors (fingerprints)[4,5]. These descriptors incorporate information relevant to ligand–receptor binding and are potentially complementary. Following a brief discussion of some general concepts related to combinatorial library design, an explanation of each descriptor will be presented. Examples of the application of BCUTs and 3D multiple-point pharmacophore fingerprints in the context of combinatorial library design will then be provided. For information concerning combinatorial library design using other methods and descriptors, the reader is referred to several recent reviews[6–13].

Diversity and focused combinatorial libraries

Combinatorial libraries can be divided into two general classes: diversity libraries and focused libraries[6–13]. Diversity libraries are intended to yield products that cover an area of some 'diversity-space' as uniformly as possible, and are generally produced for lead discovery purposes. Focused libraries are intended to yield products that interact with a single biological receptor, or possibly a family of related receptors. These libraries might be designed based on information obtained from the structures of known active ligands, X-ray structures of target receptors, or both. Focused libraries are useful for both lead optimization and lead discovery purposes.

Combinatorial reagent selection based on product or reagent structure

Combinatorial reagents can be selected based on the structures and properties of the reagents or the combinatorial products. For example, reagents for a library of 100 amides formed from carboxylic acids and amines can

1359-6446/01/$ – see front matter ©2001 Elsevier Science Ltd. All rights reserved. PII: S1359-6446(00)01665-2

be selected by choosing ten diverse acids and ten diverse amines from the available pools of each reagent type. Alternatively, a large virtual library of amides can be constructed from the available acids and amines, and the most diverse 100 products identified. The reagents required to generate these products can then be determined based on the product structures. This 'cherry-picking' procedure is unlikely to provide a combinatorially efficient set of reagents (e.g. ten amines and ten acids). However, the approach can be modified to select a diverse set of amides within the constraint of combinatorial efficiency[14].

The reagent-based procedure has the advantage of speed and simplicity, requiring far fewer expensive descriptor/property calculations. However, the added complexity and resource requirements of product-based reagent selection are justified in the design of focused libraries because the shape and electrostatic properties of whole molecules dictate their interactions with biological receptors[15]. Application of product-based reagent-selection methods to the design of diversity libraries is supported by studies suggesting that they provide sets of combinatorial products that are more diverse than those generated by reagent-based methods in some instances[14,16].

Partition-based methodology

In partition-based diversity or similarity methods[1,17,18], a 'chemistry space' is initially defined. Each dimension of the chemistry space represents a range of values for some descriptor, and an individual compound is located within the chemistry space on the basis of its value for each descriptor. Once the dimensions are determined, each is divided into bins. This binning further defines cells within the chemistry space. For example, a 3D chemistry space in which each dimension is divided into ten bins contains 1000 cells. Compounds that occupy the same cell are similar with respect to their values for the descriptors that define the chemistry space, whereas compounds that occupy distant cells are different.

Partition-based methods are especially useful for comparing different compound populations and for identifying diversity voids[1,17,18]. Two sets of compounds can be partitioned into the same chemistry space, and their overlap determined by identifying cells that contain compounds from each population. In a similar fashion, regions of the chemistry space that are occupied by compounds from one population, can be distinguished from those occupied by another population. Diversity voids are cells in a chemistry space that contain less than a threshold number of compounds. These can be targeted in the library design process to provide a more complete coverage of the chemistry space.

Molecular descriptors incorporating 'receptor-relevant' information

DiverseSolutions™ BCUT descriptors

BCUT descriptors[1–3], developed by Pearlman and incorporated into the DVS software package, combine physicochemical properties relevant to ligand–receptor binding with topological or distance information, such as the number of bonds (or Euclidean distance) between atoms in a molecule. In effect, each BCUT condenses a large amount of molecular structure and property information into a single number. The properties evaluated include tabulated atomic polarizability, atomic charge, and atomic hydrogen-bond donor and acceptor ability. BCUT descriptors can be calculated either from two-dimensional (2D) connection tables or 3D low-energy conformations generated with CONCORD™ software (Tripos). BCUTs are the highest and lowest eigenvalues of square matrices, including property information in the diagonal elements and distance-related information in the off-diagonal elements. Various scaling factors are also incorporated for both the diagonal and off-diagonal components.

Generally, many BCUT descriptors are calculated for a set of compounds (e.g. a virtual combinatorial library), and the subset of BCUTs that provides the best separation between the compounds is selected using a chi-squared algorithm integral to DVS[1,2]. This subset, usually 4–6 BCUT descriptors, defines a low-dimensional BCUT chemistry space for the compound population. The chemistry-space is then divided into cells. This step determines the resolution at which the virtual library will be examined, and binning, such that 12%–15% of the cells are occupied, has been recommended[2,19].

The similarity or diversity of a population of compounds is evaluated based on cell occupancy. Reagents for diversity libraries can be selected so that their products occupy the maximum number of different chemistry-space cells (Fig. 1). Reagents for focused libraries can be selected to yield products that occupy the same chemistry-space cells as active compounds. Compounds that are active against biological targets have been shown to cluster within 2-D and 3D subsets of larger BCUT chemistry spaces[3,20]. The existence of these 'receptor-relevant subspaces' suggests that BCUT descriptors incorporate information related to affinity at biological receptors.

The physical interpretation of individual BCUT descriptors is not readily apparent[10]. However, BCUTs have been used as descriptors in a quantitative SAR (QSAR) study of dihydrofolate reductase inhibitors[21], and partial least squares (PLS) discriminant models, utilizing BCUT descriptors, were developed to classify kinase inhibitors based on their targets[22]. Both studies found that BCUTs encode

information related to molecular properties, and complement the information encoded in other descriptors.

3D multiple-point pharmacophores and pharmacophore fingerprints

A three-point pharmacophore is defined by a set of three molecular 'features' (selected from: acids, basic nitrogen atoms, hydrophobes, aromatic ring centroids, hydrogen bond donors and hydrogen bond acceptors) and the three inter-feature distances[4] (Fig. 2). Four features and six distances define a four-point pharmacophore[5]. The set of six different feature types was designed to include functionality that is expected to contribute to ligand–receptor non-covalent binding.

Inter-feature distances are assigned to bins corresponding to distance ranges, rather than stored as discrete distances. Typically, seven or ten distance bins are used[5]. With six feature types and seven distance ranges, there are approximately 9000 geometrically possible three-point pharmacophores and 2.3 million four-point pharmacophores.

Pharmacophores are calculated within the ChemDiverse™ module of the Chem-X™ software package (Oxford Molecular, Oxford, UK). An extensive set of parameters is used to identify pharmacophoric features in molecules, and a rapid rule-based conformational searching routine that performs either systematic or random searches is employed. Thus, all possible three- or four-point pharmacophores in a large set of conformations for any given molecule can be identified. This process generally requires 1–10 central processing unit (CPU) seconds per compound on an R10000™ processor (MIPS Technologies, Mountain View, CA, USA).

Bit strings, or fingerprints, are used to denote the presence or absence of individual pharmacophores in a molecule. Each bit in the fingerprint corresponds to a pharmacophore, and the value of the bit is set to one, if the pharmacophore is present in a molecule, or zero if it is absent. Fingerprints can be computed for individual molecules or sets of molecules, and the similarity of two molecules can be evaluated using a Tanimoto score[23] based on the number of pharmacophores they have in common[5].

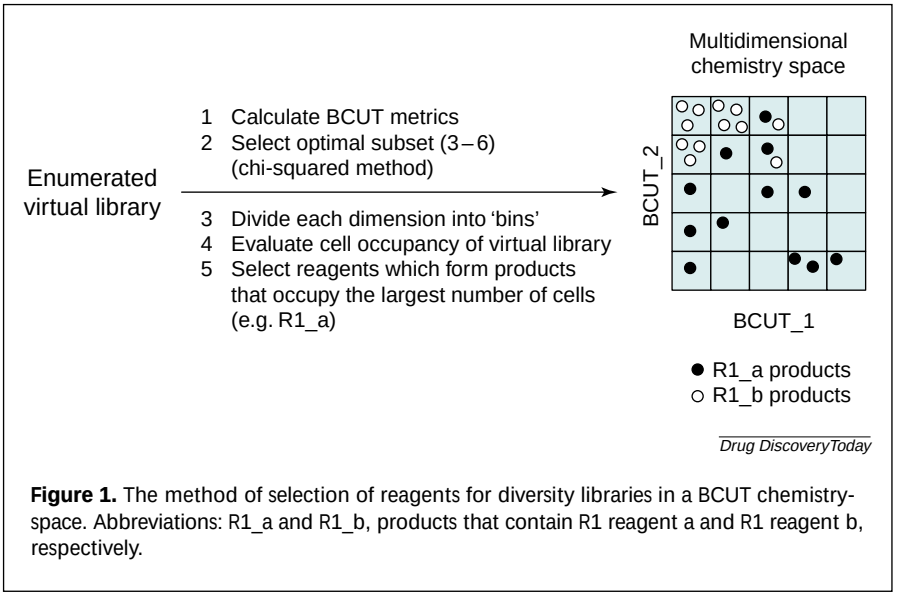


Figure 1. The method of selection of reagents for diversity libraries in a BCUT chemistry-space. Abbreviations: R1_a and R1_b, products that contain R1 reagent a and R1 reagent b, respectively.

BCUT and 3D multiple-point pharmacophore fingerprints in combinatorial library design

BCUT descriptors

Several properties of BCUT descriptors make them well suited for library design purposes. These include low computational cost, their ability to encode 3D information to a limited extent, and the ease with which they can be used for partition-based selections.

BCUTs are rapidly calculated. This is especially important when product-based reagent selection methods are used because descriptors for all products in potentially large virtual libraries must be computed. A set of 18 2D BCUT descriptors for a virtual library containing four million compounds can be calculated in a single day on a multi-processor server. Although 2D BCUTs are probably sufficient for most diversity-related tasks, for combinatorial library design problems, where the overall diversity of the virtual library is limited, the use of 3D BCUTs could provide improved discrimination[2].

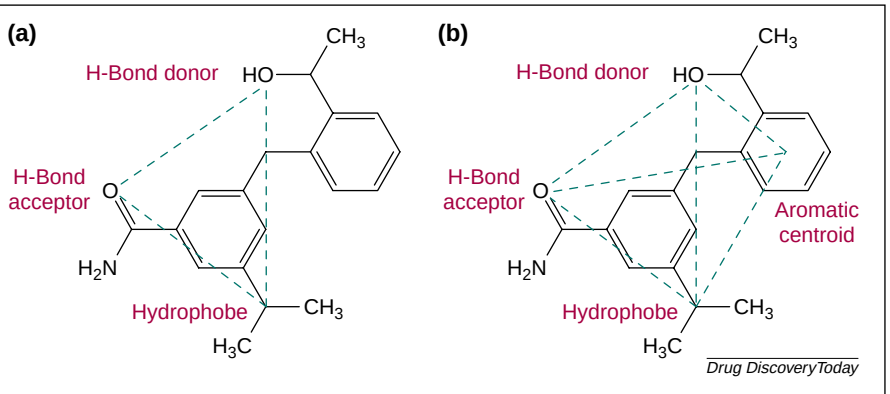


Figure 2. Three-point **(a)** and four-point **(b)** pharmacophores.

The similarity or diversity of compounds within BCUT chemistry spaces can be evaluated with cell-based partitioning methods, which are especially powerful when used in the context of combinatorial library design. One of their strengths is the ease with which diversity voids can be identified[1,17,18]. Cells that are occupied at a level below some chosen threshold in a reference population (e.g. HTS deck) can be identified, and combinatorial libraries can be designed to fill them.

Examples of BCUT descriptors in library design

The DVS software package contains an algorithm for selecting subsets of reagents that provide diverse sets of products within a BCUT chemistry space, subject to plate layout constraints (Pearlman, R.S. and Smith, K.M. Novel algorithms for the design of diverse and focused combinatorial libraries. *217th ACS National Meeting*, 21–25 March 1999, Anaheim, CA, USA, abstract COMP-197). Products of reagents chosen with this method show excellent coverage of the areas of BCUT chemistry spaces occupied by the products of the virtual libraries from which they were selected. The algorithm can also be used to select reagents for focused libraries. In this case, reagents giving products that are as similar as possible to active compounds in the same BCUT chemistry space are selected.

Optimization of product diversity in BCUT chemistry spaces concurrently with other properties has been achieved with both genetic algorithms (GA)[24] and simulated annealing (SA)[25,26] protocols. A reagent selection program based on a GA that concurrently optimizes the number of BCUT chemistry-space cells occupied by the products of the selected set of reagents, and the drug-like index scores for the reagents, has been developed (S.J. Cho, unpublished). The drug-like index score (M.A. Hermsmeier, unpublished) was developed to rank combinatorial reagents based on how frequently their constituent fragments occur in compounds present in the MDL Drug Data Report (MDDR; MDL Information Systems, San Leandro, CA, USA) and World Drug Index (WDI; Derwent Information, http://www.derwent.com/worlddrugindex/index.html).

Mason and Beno reported an SA procedure for combinatorial reagent selection that concurrently optimizes product diversity in a BCUT chemistry space, as well as in terms of four-point pharmacophores[27]. Examples of reagent selection based on product diversity in a BCUT chemistry space alone were also provided. From a virtual library of 86,140 amides formed from 292 carboxylic acids and 295 amines, they were able to choose subsets of 20 acids and 20 amines (400 products) that occupied 400 cells in the five-dimensional (5D) BCUT chemistry space derived for the entire virtual library. The products of the optimized reagents occupied approximately 23% more cells than those of randomly selected reagents.

3D multiple-point pharmacophore fingerprints

Like BCUT descriptors, pharmacophore fingerprints allow facile identification of diversity voids. Any zero bits in the fingerprint calculated for a database of compounds (e.g. HTS deck) signify pharmacophores that are not represented in the database. Combinatorial libraries can be designed to yield products enriched in the missing pharmacophores at any desired level of coverage.

Reagents for diversity libraries can be selected to yield products in which the total number of different pharmacophores represented is maximized. However, not all of the 9000 three-point pharmacophores or 2.3 million possible four-point pharmacophores are likely to represent interactions that actually contribute to binding of active compounds to their receptors. Alternatively, coverage of 'biologically important' pharmacophores can be maximized. For example, the pharmacophores present in a large database of drug-like molecules, such as the MDDR, can be identified, and libraries could be designed to provide good coverage of that set of pharmacophores[28].

Libraries can be designed to maximize the pharmacophoric similarity of the products with a single active compound[29]. Alternatively, a pharmacophore fingerprint can be derived for a set of ligands that are active against a particular receptor. Each of the pharmacophores present in the fingerprint represents a binding hypothesis, and focused libraries can be designed to produce compounds that cover as many of these hypotheses as possible, at any desired level of coverage[5].

This approach can be extended by including a frequency count for each pharmacophore present in the set of ligands[30,31]. A set of compounds active against a common target might have many pharmacophores in common. Those that occur frequently are more likely to correspond to the actual pharmacophore(s) responsible for binding (assuming a common binding mode). In this case, libraries can be designed to cover pharmacophores at levels proportional to their frequency of occurrence in the set of actives.

When a protein X-ray structure is available, and the location of the binding site is known, pharmacophore fingerprints can be calculated complementary to the binding site [Design in Receptor (DiR) Chem-X™ module; Oxford Molecular][27,32]. This involves mapping the functional groups present in the binding site, and placing complementary functional groups within the binding site in chemically reasonable positions. A pharmacophore fingerprint can then be derived to include all of the pharmacophores formed from the complementary site points.

Each pharmacophore thus identified represents a binding hypothesis, and combinatorial libraries can be designed to cover each of the hypotheses as many times as desired.

Diversity library design using 3D multiple-point pharmacophore fingerprints

Several groups have utilized three- or four-point pharmacophores for library design purposes. Davies and Briant proposed a procedure for selecting reagents, based on their frequency in a set of combinatorial products chosen to maximize the number of different three-point pharmacophores covered (presentation by K. Davies, and C. Briant, available online at http://www.netsci.org/Science/Combichem/feature05.html).

The DIVSEL program was developed by Pickett and coworkers for maximum dissimilarity selection, and has been used for combinatorial reagent selection using three-point pharmacophores to quantify diversity[33]. The algorithm starts by selecting the compound most dissimilar to the others in the set, and then iteratively selects compounds most dissimilar to those already selected. DIVSEL was used to select a set of carboxylic acids for an amide library, based on the pharmacophoric diversity of the products. Initially, a virtual library of 12,100 amides was constructed from 11 amines (selected for maximum pharmacophoric diversity) and 1100 carboxylic acids. Pharmacophore fingerprints for the 1100 products formed from the reaction of each amine with all 1100 acids were calculated and used as input for the DIVSEL program. The products of 100 acids selected with DIVSEL, together with the 11 amines, covered 85% of the three-point pharmacophores represented by the entire 12,100 compound virtual library.

Several groups have developed full-matrix optimization approaches for combinatorial reagent selection, based on the pharmacophoric diversity of the combinatorial products. These methods couple various pharmacophore diversity-based scoring functions with SA[25,26] or GA[24] optimization routines. Lewis and coworkers developed SA and GA versions of the program Rpick, to perform full-matrix optimizations based on three-point pharmacophore diversity as well as alternative descriptors[34]. A simple pharmacophore diversity function, based on the ratio of the number of different pharmacophores found in the fingerprint for the selected subset of products, to the total number of different pharmacophores found in the fingerprint for the entire virtual library, was used. Using the GA version of Rpick, they were able to select a set of $4 \times 4 \times 3 \times 2$ reagents for a 4-dimensional (4D) benzodiazepine library (96 products) from a virtual library of 1320 compounds, such that the products of the selected reagents covered 83% of the pharmacophores represented by the entire virtual library.

Good and Lewis developed the HARPick program that uses SA to select reagents based on product pharmacophore diversity[30]. The scoring function in HARPick is highly flexible, allowing inclusion of several different terms. The primary function used is the number of unique three-point pharmacophores represented in the selected subset of the virtual library. Also available are partition-scoring function terms for the number of heavy atoms, largest triangle perimeter and largest triangle area for all pharmacophores. The scoring function forces an even distribution of these properties in the selected virtual library subset, and ensures that reagents are selected that will yield products that are diverse and representative of the entire virtual library. An additional scoring function term controls the flexibility of the selected products. Importantly, the HARPick scoring function also allows optimization to distributions of pharmacophores. This feature facilitates pharmacophore void-filling.

Beno and Mason reported a procedure for reagent selection based on an SA algorithm that also incorporates multiple terms into the scoring function[27]. This was designed to concurrently optimize four-point pharmacophore and BCUT chemistry-space diversity. Unlike the HARPick program, this procedure calculates four-point pharmcophores during the optimization procedure ('on the fly') from pre-enumerated virtual libraries to allow exploration of virtual libraries where precalculation of four-point pharmacophores for all of the products would be impossible. The program was used to select 20 carboxylic acids and 20 amines from a virtual library of 86,140 amides formed from 292 acids and 295 amines. Selection was performed on the basis of the diversity of the product subset within a 5D BCUT chemistry space derived for the entire virtual library, and in terms of four-point pharmacophore diversity. In two optimization trials, the products of the optimized reagents occupied 20%–23% more BCUT chemistry-space cells than the products of the initial randomly selected reagents. The number of unique four-point pharmacophores represented by the products of the selected reagents increased by 1.8–2.6-fold during the optimization. This program can select reagents that generate products that are diverse in terms of four-point pharmacophores and BCUT chemistry-space cell occupation, while enforcing reasonable physical property distributions. With no added constraints, pharmacophore-based reagent selection algorithms, which attempt to maximize numbers of pharmacophores, tend to select highly functionalized, flexible, high-molecular-weight reagents[33], because these have the potential to form products with large numbers of pharmacophores. The 'on the fly' pharmacophore calculation approach employed by Mason and Beno allows exploration

of large virtual libraries. However, the authors noted that the current implementation is too slow for the selection of library subsets of more than a few hundred compounds.

The reagent selection methods described previously, utilize three- or four-point pharmacophore descriptors directly in procedures designed to optimize diversity. McGregor and Muskal used a different approach with Affymax (Santa Clara, CA, USA) PharmPrint™ fingerprints[35], which are similar to Chem-X/ChemDiverse three-point pharmacophores. The authors applied principal component analysis (PCA) to derive a low-dimensional space from the PharmPrint fingerprints of 9104 compounds representing 152 MDDR activity classes[28]. Combinatorial reagent selection for libraries formed from eight scaffolds and the set of amino acids was performed using a Monte Carlo algorithm and a scoring function that maximized overlap between the selected library subset, and the 9104 MDDR compounds within the first three components of the MDDR/PCA space. An average overlap with the MDDR compounds of 29.7% (ten simulations) was found for random reagent selections, and the optimization procedure increased this to 52.6%.

Focused library design using 3D multiple-point pharmacophore fingerprints

Pharmacophore fingerprints can be used to design focused libraries based on the structures of either single lead compounds or sets of active compounds. Pickett and coworkers reported a product-based reagent selection procedure that optimizes the average similarity of a library subset with lead compounds, based on the contribution of each library product[29].

Another approach to focused library design using pharmacophore descriptors is based on the distributions of pharmacophores in sets of active compounds. Initially, a pharmacophore fingerprint for a set of compounds active against a particular target is calculated. The frequency with which each pharmacophore in the fingerprint occurs is also determined. The resultant histogram of pharmacophore counts and identities can then be used with optimization routines to select sets of combinatorial products that best match the reference pharmacophore distribution. Reagents are then chosen based on their frequency of occurrence in the set of selected products.

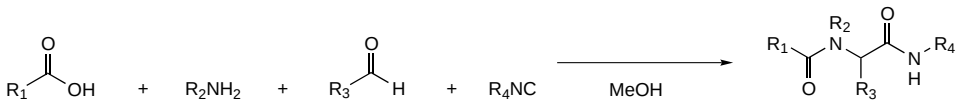
A similar, although more sophisticated, approach was developed by Bradley and coworkers, and was used to design a library enriched in α_1-adrenergic antagonists[31]. Each two-, three- and four-point pharmacophore found in training sets of active and inactive compounds was assigned an 'information content' score, based on how often it occurred in the active and inactive compounds. Pharmacophores that occurred frequently in the active compounds but infrequently in the inactive compounds were given high 'information-content' scores, as were pharmacophores that occurred frequently in the inactives and infrequently in the actives. The set of pharmacophores with the highest 'information-content' scores was then used as a filter for differentiating actives from inactives. From a virtual library of 4000 compounds, including three active α_1-adrenergic receptor antagonists, the authors were able to select a set of 639 compounds that passed the filter. A 160-compound library was designed to yield as many of those compounds as possible. Two of the three active compounds in the virtual library of 4000 compounds were also present in this small library.

As a final example of 3D multiple-point pharmacophores applied to focused library design, the previously mentioned Design in Receptor™ (DiR Chem-X™ module; Oxford Molecular) method is discussed[27,32,36]. DiR utilizes both pharmacophore and shape information obtained from X-ray structures of proteins. The method involves several steps. First, locations for site points complementary to features (e.g. hydrogen bond donors, acids, bases) within the binding site are determined using the GRID™ (Molecular Discovery Limited, Oxford, UK; Ref. 37) program, which performs energetic surveys of the site using various probe atoms. A pharmacophore fingerprint based on these complementary features is then calculated. Each pharmacophore present in the fingerprint represents a binding hypothesis. Compounds are docked into the binding site, and the docked orientations are scored, based on the number of pharmacophore hypotheses they match. Docked orientations that have unfavorable steric contacts with the protein are rejected. Mason used DiR to rank combinatorial reagents for a library based on the Ugi condensation reaction[38] by their potential to yield products that match pharmacophores complementary to the factor Xa binding site[27,39]. The study showed that isocyanate reagents containing *meta*-benzamidine moieties generated products that matched more site pharmacophores than those containing *para*-benzamidines. The study also demonstrated the value of including methylene chains of increasing length between the isocyanate and benzamidine moieties in terms of the number of site-pharmacophore hypotheses matched (Table 1).

Future directions

BCUTs have been utilized for a variety of diversity-related purposes including reagent selection for diversity libraries and diverse subset selection. Although DVS contains an algorithm allowing BCUTs to be used for reagent selection for focused libraries, examples of this have yet to appear in

Table 1. Factor Xa binding-site four-point pharmacophore hypotheses matched by products of Ugi condensation library[a]



R4	R1	R2	R3	Hypotheses matched
p-benzamidine	Ph	H	Et	4
p-benzamidine	CH_2Ph	H	Et	4
m-benzamidine	Ph	H	Et	20
m-benzamidine	CH_2Ph	H	Et	17
-(CH_2)-*m*-benzamidine	CH_2Ph	H	Et	35
-$(CH_2)_2$-*m*-benzamidine	CH_2Ph	H	Et	44
-$(CH_2)_3$-*m*-benzamidine	CH_2Ph	H	Et	64

[a]Adapted from Ref. 27.

the literature. Reports of the successful application of BCUTs as QSAR descriptors[21,22], and especially the observation that compounds active against common targets can cluster in low-dimensional BCUT chemistry spaces, suggest the potential for focused library-design tasks[3,20]. Studies to validate BCUT descriptors for this purpose are required. Another area of interest is the objective comparison of 2D and 3D BCUTs for library design purposes. Although 3D BCUTs have the potential to provide increased discrimination of compounds in virtual libraries where diversity is limited[2], this has not yet been demonstrated.

Improvements in two areas of 3D multiple-point pharmacophore technology will increase the utility of the descriptors for combinatorial library design tasks. The first of these is computational speed. Molecular conformer generation is the most time-consuming step in pharmacophore fingerprint calculations, and virtual libraries of more than ~500,000 compounds are impractical. This is not a substantial limitation for focused library design, because there is often enough information available to reduce the size of the reagent pool based on other criteria. However, consideration of large virtual libraries is important for diversity library design. If the atom-typing methods currently used for pharmacophore perception can be coupled with faster conformational searching routines, 3D multiple-point pharmacophore fingerprints will be even more effective for combinatorial library design.

Additional work in the area of noise reduction is also necessary. A ligand might have hundreds or thousands of three- or four-point pharmacophores. However, only a few of these are likely to be important for receptor binding. Current methods including examination of pharmacophore distributions[5,28,30,31] and trend vector analysis[40,41] can be used when sets of ligands that bind to common targets are available. However, a systematic study of the pharmacophores represented in known drugs could provide additional insights.

To ensure that hits identified from diversity or focused libraries are of sufficient quality for development, several additional parameters beyond diversity and similarity must be controlled in the library design process[42]. For example, the distributions of ADME-related physicochemical properties in the products must be constrained to acceptable ranges[43–45]. Extension of current library design tools that utilize BCUTs and/or 3D multiple-point pharmacophore fingerprints to allow concurrent optimization of these properties will facilitate the design of libraries enriched in 'drug-like' compounds.

Summary

BCUT descriptors and 3D multiple-point pharmacophores encode information relevant to ligand–receptor binding. Both descriptors could be combined with partition-based selection methods and used for product-based combinatorial reagent selection. BCUTs are rapidly calculated, allowing virtual libraries of millions of compounds to be examined. This makes them especially attractive for diversity library design. However, physical interpretation of individual BCUTs is difficult[10] because they essentially condense a large amount of structural and property information into a single number. Where compounds that bind to the same receptor cluster in a BCUT chemistry space, focused library design using BCUTs might prove effective.

Currently, the calculation of 3D multiple-point pharmacophore fingerprints requires a significant amount of CPU time per molecule, and virtual libraries of approximately 500,000 compounds represent a practical limit. This limits the utility of pharmacophores for diversity library design purposes. However, they are effective for focused library design based on structures of single active compounds and

sets of actives. In addition, the DiR method facilitates focused design, based on multiple pharmacophore binding hypotheses within the constraints of protein active sites.

Future developments will probably include studies of the utility of BCUT descriptors for focused library design and the development of more rapid methods for calculating 3D multiple-point pharmacophore fingerprints.

Acknowledgements

We thank Daniel Cheney, Andrew Good, Sung Jin Cho, Yax Sun, Keith Davies, Cathy White-Davies and Bob Pearlman.

References

1 Pearlman, R.S. and Smith, K.M. (1998) Novel software tools for chemical diversity. *Perspect. Drug Des. Discovery* 9, 339–353
2 Pearlman, R.S. (1997) *Diverse Solutions User's Manual*, pp. 1–44, University of Texas
3 Pearlman, R.S. and Smith, K.M. (1999) Metric validation and the receptor-relevant subspace concept. *J. Chem. Inf. Comput. Sci.* 39, 28–35
4 Pickett, S.D. *et al.* (1996) Diversity profiling and design using 3-D pharmacophores: pharmacophore-derived queries (PDQ). *J. Chem. Inf. Comput. Sci.* 36, 1214–1223
5 Mason, J.S. *et al.* (1999) New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* 42, 3251–3264
6 Bures, M.G. and Martin, Y.C. (1998) Computational methods in molecular diversity and combinatorial chemistry. *Curr. Opin. Chem. Biol.* 2, 376–380
7 Van Drie, J.H. and Lajiness, M.S. (1998) Approaches to virtual library design. *Drug Discov. Today* 3, 274–283
8 Spellmeyer, D.C. and Grootenhuis, P.D.J. (1999) Recent developments in molecular diversity: computational approaches to combinatorial chemistry. *Annu. Rep. Med. Chem.* 34, 287–296
9 Drewry, D.H. and Young, S.S. (1999) Approaches to the design of combinatorial libraries. *Chemometr. Intell. Lab. Syst.* 48, 1–20
10 Agrafiotis, D.K. *et al.* (1999) Advances in diversity profiling and combinatorial series design. In *Annual Reports in Combinatorial Chemistry and Molecular Diversity* (Vol. 2) (Pavia, M.R. and Moos, W.H., eds), pp.71–92, Kluwer Academic Publishers.
11 Leach, A.R. and Hann, M.M. (2000) The *in silico* world of virtual libraries. *Drug Discov. Today* 5, 326–336
12 Lewis, R.A. *et al.* (2000) Computer-aided molecular diversity analysis and combinatorial library design. In *Reviews in Computational Chemistry* (Vol. 16) (Lipkowitz, K.B. and Boyd, D.B., eds), pp. 1–51, Wiley-VCH
13 Willett, P. (2000) Chemoinformatics – similarity and diversity in chemical libraries. *Curr. Opin. Biotechnol.* 11, 85–88
14 Gillet, V. *et al.* (1997) The effectiveness of reagent pools for generating structurally-diverse combinatorial libraries. *J. Chem. Inf. Comput. Sci.* 37, 731–740
15 Ashton, M.J. *et al.* (1996) New perspectives in lead generation II: evaluating molecular diversity. *Drug Discov. Today* 1, 71–78
16 Jamois, E.A. *et al.* (2000) Evaluation of reagent-based and product-based strategies in the design of combinatorial library subsets. *J. Chem. Inf. Comput. Sci.* 40, 63–70
17 Mason, J.S. and Pickett, S.D. (1998) Partition-based selection. *Perspect. Drug Des. Discovery* 7/8, 85–114
18 Bayley, M.J. and Willett, P. (1999) Binning schemes for partition-based compound selection. *J. Mol. Graph. Model.* 17, 10–18
19 Menard, P.R. *et al.* (1998) Chemistry space metrics in diversity analysis, library design, and compound selection. *J. Chem. Inf. Comput. Sci.* 38, 1204–1213
20 Schnur, D. (1999) Design and analysis of large combinatorial libraries using cell-based methods. *J. Chem. Inf. Comput. Sci.* 39, 36–45
21 Stanton, D.T. (1999) Evaluation and use of BCUT descriptors in QSAR and QSPR studies. *J. Chem. Inf. Comput. Sci.* 39, 11–20
22 Pirard, B. and Pickett, S.D. (2000) Classification of kinase inhibitors using BCUT descriptors. *J. Chem. Inf. Comput. Sci.* 40, 1431–1440
23 Willett, P. *et al.* (1998) Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38, 983–996
24 Brown, R.D. and Clark, D.E. (1998) Genetic diversity: applications of evolutionary algorithms to combinatorial library design. *Expert Opin. Ther. Pat.* 8, 1447–1460
25 Agrafiotis, D.K. (1997) Stochastic algorithms for maximizing molecular diversity. *J. Chem. Inf. Comput. Sci.* 37, 841–851
26 Zheng, W. *et al.* (1999) Rational combinatorial library design. 3. Simulated annealing guided evaluation (SAGE) of molecular diversity: a novel computational tool for universal library design and database mining. *J. Chem. Inf. Comput. Sci.* 39, 738–746
27 Mason, J.S. and Beno, B.R. (2000) Library design using BCUT chemistry-space descriptors and multiple four-point pharmacophore fingerprints: simultaneous optimization and structure-based diversity. *J. Mol. Graph. Model.* 18, 438–451
28 McGregor, M.J. and Muskal, S.M. (2000) Pharmacophore fingerprinting. 2. Application to primary library design. *J. Chem. Inf. Comput. Sci.* 40, 117–125
29 Pickett, S.D. *et al.* (2000) Enhancing the hit-to-lead properties of lead optimization libraries. *J. Chem. Inf. Comput. Sci.* 40, 263–272
30 Good, A.C. and Lewis, R.A. (1997) New methodology for profiling combinatorial libraries and screening sets: cleaning up the design process with HARPick. *J. Med. Chem.* 40, 3926–3936
31 Bradley, E.K. *et al.* (2000) A rapid computational method for lead evolution: description and application to α_1-adrenergic antagonists. *J. Med. Chem.* 43, 2770–2774
32 Murray, C.M. and Cato, S.J. (1999) Design of libraries to explore receptor sites. *J. Chem. Inf. Comput. Sci.* 39, 46–50
33 Pickett, S.D. *et al.* (1998) DIVSEL and COMPLIB – strategies for the design and comparison of combinatorial libraries using pharmacophoric descriptors. *J. Chem. Inf. Comput. Sci.* 38, 144–150
34 Lewis, R.A. *et al.* (1997) Quantification of molecular similarity and its application to combinatorial chemistry. In *Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry* (van de Waterbeemd, H. *et al.*, eds), pp. 137–156, Wiley-VCH
35 McGregor, M.J. and Muskal, S.M. (1999) Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J. Chem. Inf. Comput. Sci.* 39, 569–574
36 Mason, J.S. and Cheney, D.L. (2000) Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput.* 5, 576–587
37 Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857
38 Ugi, I. and Steinbruckner, C. (1961) Isonitriles. II. Reaction of isonitriles with carbonyl compounds, amines and hydrazoic acid. *Chem. Ber.* 94, 734–742
39 Tulinsky, A. *et al.* (1993) Structure of human des9(1–45) factor Xa at 2.2 Å resolution. *J. Mol. Biol.* 232, 947–966
40 Sheridan, R.P. *et al.* (1994) Extending the trend vector: the trend matrix and sample-based partial least squares. *J. Comput.-Aided Mol. Design* 8, 323–340
41 Sheridan, R.P. and Kearsley, S.K. (1995) Using a genetic algorithm to suggest combinatorial libraries. *J. Chem. Inf. Comput. Sci.* 35, 310–320
42 Martin, E.J. and Critchlow, R.E. (1999) Beyond mere diversity: tailoring combinatorial libraries for drug discovery. *J. Comb. Chem.* 1, 32–45
43 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25
44 Blake, J.F. (2000) Chemoinformatics – predicting the physicochemical properties of 'drug-like' molecules. *Curr. Opin. Biotechnol.* 11, 104–107
45 Walters, W.P. *et al.* (1999) Recognizing molecules with drug-like properties. *Curr. Opin. Chem. Biol.* 3, 384–387