Metric Validation and the Receptor-Relevant Subspace Concept

R. S. Pearlman* and K. M. Smith

Laboratory for Molecular Graphics and Theoretical Modeling College of Pharmacy, University of Texas, Austin, Texas 78712

Received August 3, 1998

Following brief comments regarding the advantages of cell-based diversity algorithms and the selection of low-dimensional chemistry-space metrics needed to implement such algorithms, the notion of metric validation is discussed. Activity-seeded, structure-based clustering is presented as an ideal approach for the validation of either high- or low-dimensional chemistry-space metrics when validation by computer-graphic visualization is not possible. Whereas typical methods for reducing the dimensionality of chemistry-space inevitably discard potentially important information, we present a simple yet novel algorithm for reducing dimensionality by identifying which axes (metrics) convey information related to affinity for a given receptor and which axes can be safely discarded as being irrelevant to the given receptor. This algorithm often reveals a three- or two-dimensional subspace of a (typically six-dimensional) BCUT chemistry-space and, thus, enables computer graphic visualization of the actual coordinates of active compounds and combinatorial libraries. Most significantly, we illustrate the importance of using receptor-relevant distances for identifying near neighbors of lead compounds, comparing libraries, and other diversity-related tasks.

1. INTRODUCTION

Although the concept of chemical diversity has been considered intuitively by chemists for many years, the advent of combinatorial chemistry and high-throughput screening have focused unprecedented attention on the need for efficient software tools to address a variety of diversity-related tasks. The most fundamental task related to chemical diversity is that of selecting a diverse subset of compounds from a much larger population (library or database) of compounds. One obvious objective of that task is to identify a subset which best represents the full range of chemical diversity present in the larger population to avoid the time and expense of screening "redundant" compounds. Ironically, recent results (see, for example, an article by Jilek et al.[1]) suggest that the utility of this most fundamental task may also be the most controversial. However, in addition to simple subset selection, practical experience has revealed other, probably more important diversity-related tasks which must also be addressed in pharmaceutical and agrochemical industry. These include identifying diversity voids ("missing diversity") within a population of compounds, identifying compounds from secondary populations to fill diversity voids found in the first, comparing both the range and diversity of compounds in two or more populations, selecting property-biased and/or reagent-biased subsets of synthesizable compounds to facilitate optimal library design, and choosing chemistry-space metrics which best represent the diversity of a given population of compounds.

Pearlman and Smith[2−4] have indicated that distance-based algorithms are quite satisfactory for simple subset selection but, lacking reference to absolute position in chemistry-space, are considerably less useful for all other diversity-related tasks. In contrast, they have demonstrated that cell-based algorithms (based on partitioning chemistry-space into hypercubic cells) enable reference to both intercompound distance and absolute position in space and, thereby, are ideally suited for all chemical diversity-related tasks.

Practical implementations of cell-based algorithms require the definition of a low-dimensional chemistry-space (in contrast to the ∼1000-dimensional "fingerprint"-based chemistry-spaces usually used with distance-based algorithms). Pearlman and Smith[2−4] observed that traditional whole-molecular descriptors such as logP, surface area, pK_a, dipole moment, HOMO/LUMO gap, etc. make rather poor chemistry-space metrics and suggested three reasons for this observation:

1. Many of the "traditional" descriptors are highly correlated; the axes of a vector-space should be orthogonal (uncorrelated).

2. Some traditional descriptors (e.g., logP and pK_a) are strongly related to drug transport or pharmacokinetics but are very weakly related to receptor affinity or activity as measured in most screening-based drug discovery efforts.

3. The traditional descriptors are whole-molecule descriptors which convey very little information about the details of molecular substructural differences which are the basis of structural diversity.

Defined in a manner which incorporates both connectivity information (based on actual bonding or interatomic distances) and atomic properties relevant to intermolecular interaction (i.e., atomic charge, polarizability, H-bond donor- and acceptor-abilities), BCUT metrics[2−4] have repeatedly been shown to be quite useful as metrics of a low-dimensional chemistry-space and have enabled implementation[4] of cell-based algorithms for all diversity-related tasks. Given the number of choices of connectivity information, atomic information, and scaling factors which control the relative balance of these two types of information, it is clear that *many* BCUT values can be generated for each chemical structure and that some algorithm is needed to decide which

10.1021/ci980137x CCC: $18.00
Published on Web 01/05/1999

BCUT values should be used as the axes of a chemistry-space. Pearlman and Smith[2−5] introduced the χ-squared-based "auto-choose" algorithm which accomplishes this task. Rather than being a problem, they demonstrated that the wide choice of BCUT values coupled with the auto-choose algorithm provides an ideal method for tailoring a chemistry-space to best represent the diversity of a given population. This is particularly desirable when considering the limited diversity within a focused population such as a combinatorial library.

It should also be noted that this χ-squared-based, auto-choose algorithm can be applied not only to combinations of BCUT-values but also to combinations involving any other low-dimensional chemistry-space metrics. In essence, the algorithm forms all possible combinations of metrics and evaluates the resulting chemistry-spaces in terms of their orthogonality and the extent to which they distinguish whatever structural differences exist between the compounds in a given population. The auto-choose algorithm typically identifies five- or six-dimensional BCUT-based chemistry-spaces. Although experience to-date strongly supports the use of BCUT-values as metrics, the DiverseSolutions software developed by Pearlman and Smith[5] encourages the user to consider his own metrics in addition to BCUT-values. However, this must be done with extreme caution for the following somewhat ironic reason. Imagine assigning random numbers to each compound of a large population and then considering those numbers as a potential axis of a chemistry-space. Since the random numbers would be uniformly distributed over the population of compounds, the χ-squared approach would perceive this "metric" (and other similarly random "metrics") as good choices as axes of a chemistry-space. This brings us, rather dramatically, to the need to validate the choice of metrics used to define a chemistry-space.

2. VALIDATION OF CHEMISTRY-SPACE METRICS

The fundamental question to be addressed by any approach to the validation of metrics used to address structural diversity is to what extent do the metrics distinguish "dissimilar" compounds and position "similar" structures near each other in chemistry-space. This question is extremely difficult—perhaps impossible—to answer in the most general sense because the concept of "similarity" depends to a very great extent on the metrics themselves and on observed structure-dependent properties which we consciously or subconsciously use in our own subjective definitions of similarity and diversity. Thus, a more practical and tractable question to be addressed is to what extent do the metrics place compounds exhibiting similar properties of interest near each other in chemistry-space. For drug and agrochemical discovery purposes, this amounts to repeated assessment of the extent to which compounds showing affinity for a given receptor are clustered near each other in chemistry-space. In other words, do the metrics reveal a structure-affinity/activity relationship (SAR) for compounds showing affinity for a given receptor?

It is important to note that valid metrics *should* be expected to reveal an SAR but *should not* be expected to support a QSAR for the same compounds. In general, chemistry-space metrics are *not* QSAR descriptors. This is easy to understand once one recalls that two drug-sized compounds differing by just a single methylene unit can have remarkably different affinities for a given receptor despite the fact that any set of general purpose metrics (fingerprints or low-dimensional metrics) based on "total molecular structure" would regard these two compounds as very similar and place them very near each other in chemistry-space. Indeed, the same two compounds might have essentially identical affinities for some other receptor. The quantitative relationship between affinity and position in chemistry-space would be terrible for the first receptor yet excellent for the second despite the fact the very same metrics were being used for (tested by) the QSAR. Thus, QSAR-related methods are of uncertain reliability for metric validation purposes.

How, then, can one reveal metric-based SARs and validate metrics used for diversity-related purposes? Pearlman and Smith[2−5] have presented a simple, direct yet novel approach to metric validation which they refer to as *activity-seeded, structure-based clustering*. Unlike typical clustering algorithms (based on structure alone) which can be used for a variety of purposes, this algorithm requires activity data (preferably, quantitative data) for a set of compounds and is intended *only* for the diversity-related task of validating chemistry-space metrics. Given a set of active compounds which all bind to a given receptor in the same way, it is certainly reasonable to expect that those active compounds should be positioned near each other in a small region of chemistry-space *if* the chemistry-space metrics are valid. If the dimensionality of chemistry-space were three or less, computer-graphic display of the coordinates of active compounds would enable visual assessment of the extent to which the active compounds are clustered. The activity-seeded, structure-based clustering algorithm provides a method for directly testing that expectation in the typical case in which the chemistry-space dimensionality is greater than three and, thus, simple visual inspection of the distribution of active compounds is difficult or impossible. The algorithm consists of the following procedure:

1. Choose a unit-cluster radius: a very small distance in the chemistry-space to be validated.
2. Center a sphere of that radius on the most active compound in the validation test set.
3. Assign other active compounds located within that sphere to that "unit-cluster".
4. Center another sphere on the next most active compound not already assigned to some unit-cluster.
5. Repeat steps 3 and 4 until all active compounds have been assigned to some unit-cluster.
6. "Coalesce" adjoining (overlapping) unit-clusters and record both the number of unit-cluster spheres per coalesced-cluster and the minimum inter-unit-cluster distance between each pair of coalesced clusters.

The algorithm can be implemented as an **O**(N) process and, thus, is extremely fast. More significantly, the algorithm can be used to validate all types of chemistry-space definitions including other (non-BCUT) low-dimensional chemistry-spaces and those based on high-dimensional fingerprints as well. When used in a cell-based context, the unit-cluster radius is typically chosen to yield a tiny hypersphere of volume equal to that of a single hypercubic cell reflecting the "resolution" corresponding to a user-specified number of bins/axis. In any case, the total number of unit-cluster

metric 1: bcut_gastchrg_S_invdist_1.00_R_L
metric 2: bcut_hdonor_S_invdist_0.30_R_H
metric 3: bcut_haccept_S_invdist_0.50_R_H
metric 4: bcut_gastchrg_S_invdist2_0.08_R_H
metric 5: bcut_tabpolar_S_invdist2_1.00_R_L
metric 6: bcut_tabpolar_S_invdist_0.70_R_H

Figure 1. The six BCUT metrics which best represent the structural diversity of compounds contained in the MDDR database. The three highlighted metrics comprise the receptor-relevant subspace identified for the ACE receptor and are consistent with the published binding models depicted in Figure 2.

spheres contained in all coalesced-clusters provides an upper bound on the volume of chemistry-space required to contain all the active compounds.

Using the activity-seeded, structure-based clustering algorithm, Pearlman and Deanda[6] have performed a number of validation studies. For example, after auto-choosing the BCUT chemistry-space which best represents the diversity of compounds in MDL's MDDR database,[7] they computed the positions of 191 diverse ACE inhibitors in that chemistry-space. The 191 inhibitors were culled from the primary literature.[8−15] Measured activities ($-\log IC_{50}$) were reported for all compounds and spanned the range 5.24−9.64. The 74 most active compounds (top 40%) had activities in the range 7.85−9.64 and were identified as "highly active" compounds. If the BCUT-values used as chemistry-space metrics were random numbers or quantities unrelated to structure and intermolecular interaction, the active compounds would be randomly distributed throughout chemistry-space. However, using the activity-seeded, structure-based clustering algorithm, they found that the 74 "highly active" compounds are all contained by just three coalesced clusters occupying less than 0.02% of the entire chemistry-space and less than 0.19% of occupied chemistry-space. Significantly, the three clusters were close to each other; the largest inter-cluster distance being just 3.2R where R is the unit-cluster radius.

It is instructive to consider the analogous results obtained using all 191 compounds (including the 117 "poorly active" compounds). Once again, the active compounds were all clustered relatively near each other, but they occupied a somewhat larger volume of chemistry-space than that occupied by just the 74 "highly active" compounds. This result is entirely consistent with expectations. There can be many different structures which exhibit poor to modest activities. In contrast, there are relatively fewer structures which exhibit high activities. This fact may be easier to appreciate by considering the notion of making structural modifications of a very highly active compound. There may be a few modifications which preserve high activity but there are far more modifications which reduce or destroy activity.

3. DIMENSIONAL REDUCTION: RECEPTOR-RELEVANT SUBSPACES

Pictures *can* be worth "a thousand words". This fact has driven some researchers to resort to cartoon illustrations depicting "islands" of active compounds or combinatorial libraries in abstract, wishfully contrived two-dimensional "chemistry-spaces". Obviously, these cartoons are merely intended to illustrate broad concepts but, frankly, may often present a very misleading impression of how *real* compounds or libraries are *really* positioned in a real chemistry-space.

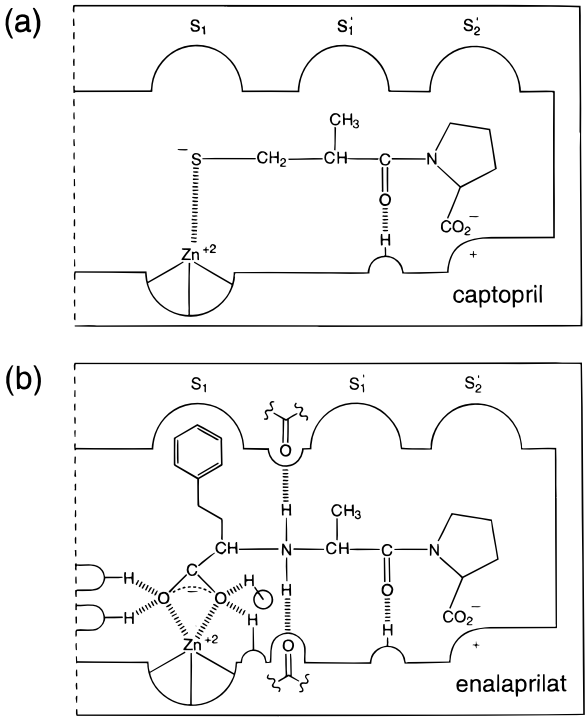

Figure 2. Cartoon representations of (a) captopril and (b) enalaprilat binding to the ACE receptor as proposed by Wyvratt and Patchett.[11] Note that the importance of negative charge and H-bond donor- and acceptor-abilities in the published binding models are consistent with the receptor-relevant axes identified in Figure 1.

Other researchers have resorted to applying various mathematical techniques (e.g., Sammon maps, Kohonen maps, principle components, etc.) to reduce the dimensionality of higher dimensional spaces down to two or three dimensions just so they can display computer graphic images. However, there is a very serious problem associated with *any* such mathematically based approach to dimensional reduction: the inevitable loss of potentially important information. That is, if 1000 bits are required to adequately describe structural diversity in a high-dimensional fingerprint space or if six BCUTs are required to adequately describe structural diversity in a low-dimensional space, the mathematical transformation of 1000 or six dimensions down to two or three must, inevitably, discard information which was previously deemed necessary for the adequate description of the structural diversity of those compounds.

Pearlman and Smith[3−5] have described a novel yet simple approach to dimensional reduction for specific purposes which minimizes the aforementioned loss of potentially important information. This approach is best appreciated by reference to a simple analogy. Consider the list of car-descriptors needed to describe all the possible features which might be important to car-buyers. The list would include descriptors such as brand, price, color, two-door vs four-door, length, engine size, transmission type, cup-holder location, etc. All of these descriptors would be of potential interest to some or all car-buyers. However, a particular buyer might care about only three of those descriptors: e.g., color, size, and cup-holder location. Those descriptors would be "relevant" to that particular car-buyer. The other descriptors would be irrelevant to that particular buyer but, obviously, might be relevant to some other car-buyer. If, for example, the particular car-buyer prefers silverish cars, all cars for

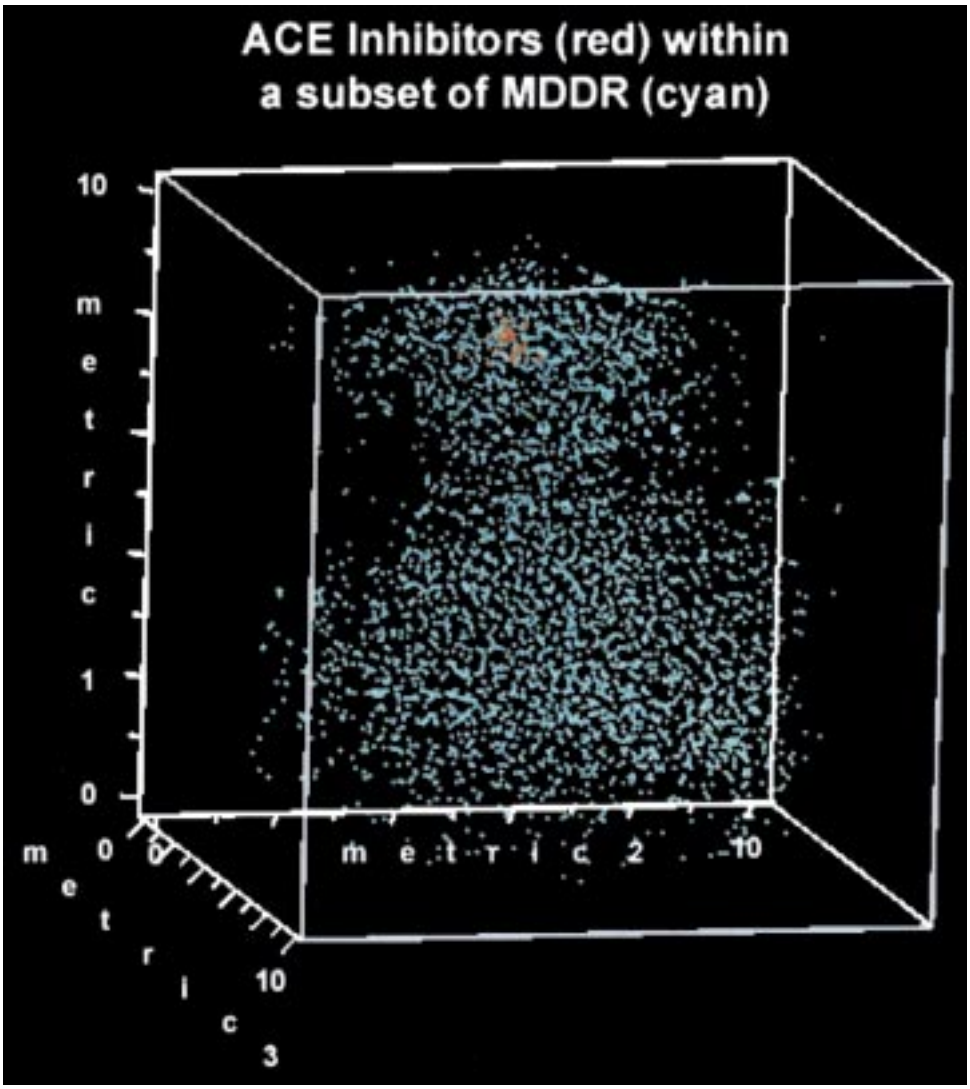

Figure 3. 74 ACE inhibitors tightly clustered within the three-dimensional ACE-receptor-relevant subspace of the MDDR (imaginary corporate database) chemistry-space. Only a 5% diverse subset of the total MDDR population is shown so that the actives can be seen within the three-dimensional "cloud" of MDDR compounds. Please note that in this and subsequent figures, the metric numbers appearing as labels of axes correspond to the ordered list of metrics presented in Figure 1.

which he or she "shows affinity" will be very similar (tightly clustered) with respect to color but could be dissimilar (poorly clustered) with respect to brand.

Each of the (typically) six BCUT metrics chosen by the χ-squared-based auto-choose algorithm represent descriptors which might be relevant to some receptor. However, a particular receptor might care about just two or three (or more) of those descriptors. Given a number of compounds for which a particular receptor has significant affinity, we can identify the *receptor-relevant subspace* for that receptor by identifying the axes (metrics) along which the active compounds are tightly clustered. Axes (metrics) which are irrelevant to this particular receptor will show poor clustering of the active compounds. The algorithm for identifying receptor-relevant metrics has recently been implemented in the DiverseSolutions software.[5] Relevant axes are identified based upon a cluster-breadth normalized value of χ-squared computed either from the simple count of active compounds per bin along each axis or from an activity-weighted count of those active compounds. The algorithm also accounts for the possibility of multiple receptor-binding modes by allowing more than one cluster of actives per relevant axis.

To illustrate the utility of this receptor-relevant subspace concept, we have used the aforementioned 191 diverse ACE inhibitors to identify a three-dimensional ACE-receptor-relevant subspace within the six-dimensional chemistry-space which best represents the structural diversity of all compounds in MDL's MDDR database of drugs and druglike compounds. Figure 1 lists the six BCUT metrics which comprise the chemistry-space which best represents the diversity of those MDDR compounds. By determining which of those metrics best cluster the 191 active ACE inhibitors, it was found that three of the six metrics—those related to negative charge, H-bond donor-ability, and H-bond acceptor-ability—are apparently relevant to the ACE receptor, while the other three are not. Figure 1 lists the metrics in order of their ACE-receptor-relevance. Figure 2 depicts a cartoon representation of ACE inhibitor binding as proposed by Wyvratt and Patchett.[11] Note that the three ACE-receptor-relevant metrics chosen by the algorithm appear to be consistent with the published binding model.

Figure 3 shows the 74 highly active ACE inhibitors tightly clustered within the three-dimensional ACE-receptor-relevant subspace of the MDDR chemistry-space. To see the actives

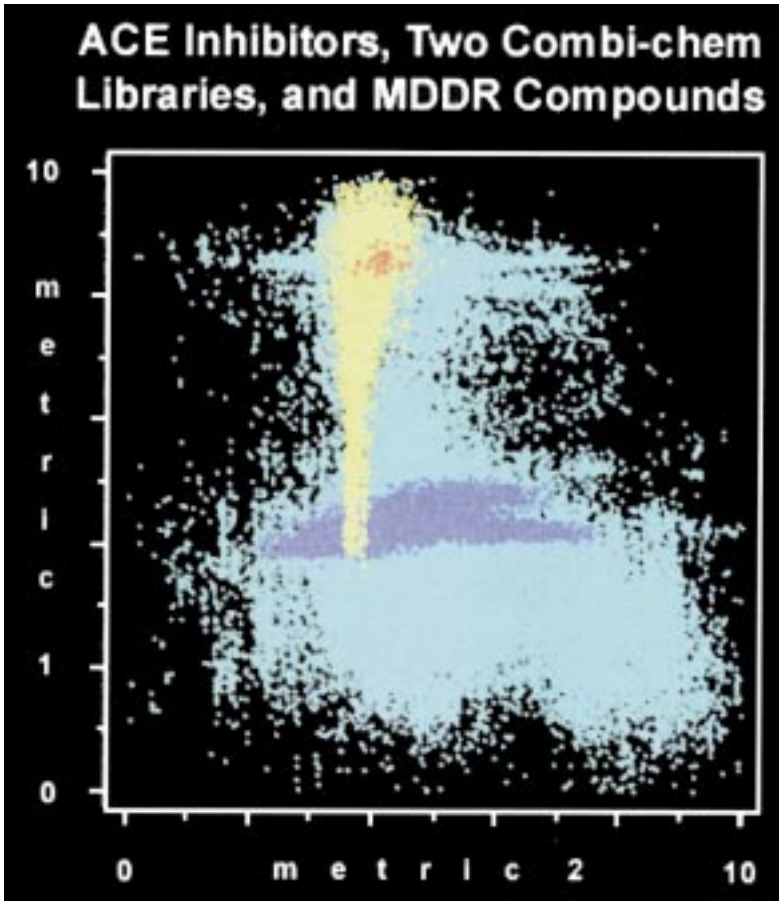

Figure 4. 74 ACE inhibitors tightly clustered within a two-dimensional projection of the three-dimensional subspace; all MDDR compounds are shown. Also shown are the regions of chemistry space covered by two hypothetical combinatorial libraries.

through the "cloud" of roughly 70 000 MDDR compounds, we have shown the positions of just a 5% diverse subset of the total MDDR population. Figure 4 shows the 74 ACE inhibitors tightly clustered within a two-dimensional projection of the three-dimensional subspace; all MDDR compounds are shown. Also shown in Figure 4 are the regions of chemistry space covered by two combinatorial libraries. The two-dimensional projection, rather than the three-dimensional space is depicted merely because the important features are easier to see in the photograph of the two-dimensional projection. Note that these figures are not cartoons! They depict the actual BCUT-based coordinates of real compounds in the real chemistry-space determined to be relevant to a real receptor.

Clearly, determining the receptor-relevant subspace enables computer graphic display which, in turn, enables *visual metric validiation*—do the metrics really cluster active compounds? If we imagine that the MDDR compounds are actually the compounds in some corporate database, then it is apparent that computer graphic display enables visual assessment of diversity—do the imaginary corporate database compounds adequately cover the region of chemistry-space containing active compounds? Visualization also enables visual comparison of compound libraries—to what extent do libraries overlap with each other, with the corporate database, and with the region of chemistry-space containing active compounds? See the article by Schnur[16] in this issue for additional examples.

Computer-graphic visualization enabled by determining the receptor-relevant subspace is clearly of tremendous utility. However, determining the receptor-relevant subspace enables something even more important. It enables the calculation of *receptor-relevant distances*. Recall the car-buyer analogy. What if, for example, the particular car-buyer had previously owned a silver Chevrolet and a silver Oldsmobile—both produced by General Motors. A friend helping this particular car-buyer choose his/her next car might think it a waste of time to look at Fords or Hondas unless the friend realized that "distance" (dissimilarity) with respect to brand is an irrelevant component of inter-car distance for this particular car-buyer. Similarly, imagine that you are helping someone choose which compounds to test (or synthesize) in a second-round of screening based on distances from the hits discovered in the first-round of screening. If the inter-compound distance includes receptor-*ir*relevant components, compounds which appear to be distant from the known-active compounds might not be considered worthy of screening, whereas if the distance was computed based only on receptor-relevant axes, some of those same compounds would appear to be close to the first-round leads and, thus, worth screening.

Figure 5 illustrates the ACE inhibitors in a three-dimensional space based on two relevant metrics (metric-1 and metric-2) and an irrelevant metric (metric-5). Note that the active compounds are tightly clustered within the respect to axes 1 and 2 but not clustered along the axis corresponding to the irrelevant metric. Imagine that the first round of screening yielded hits near the left-hand end of the roughly cylindrical region of space containing active compounds. Including the irrelevant metric-5 in the distance formula used to find near neighbors of known actives for second-round

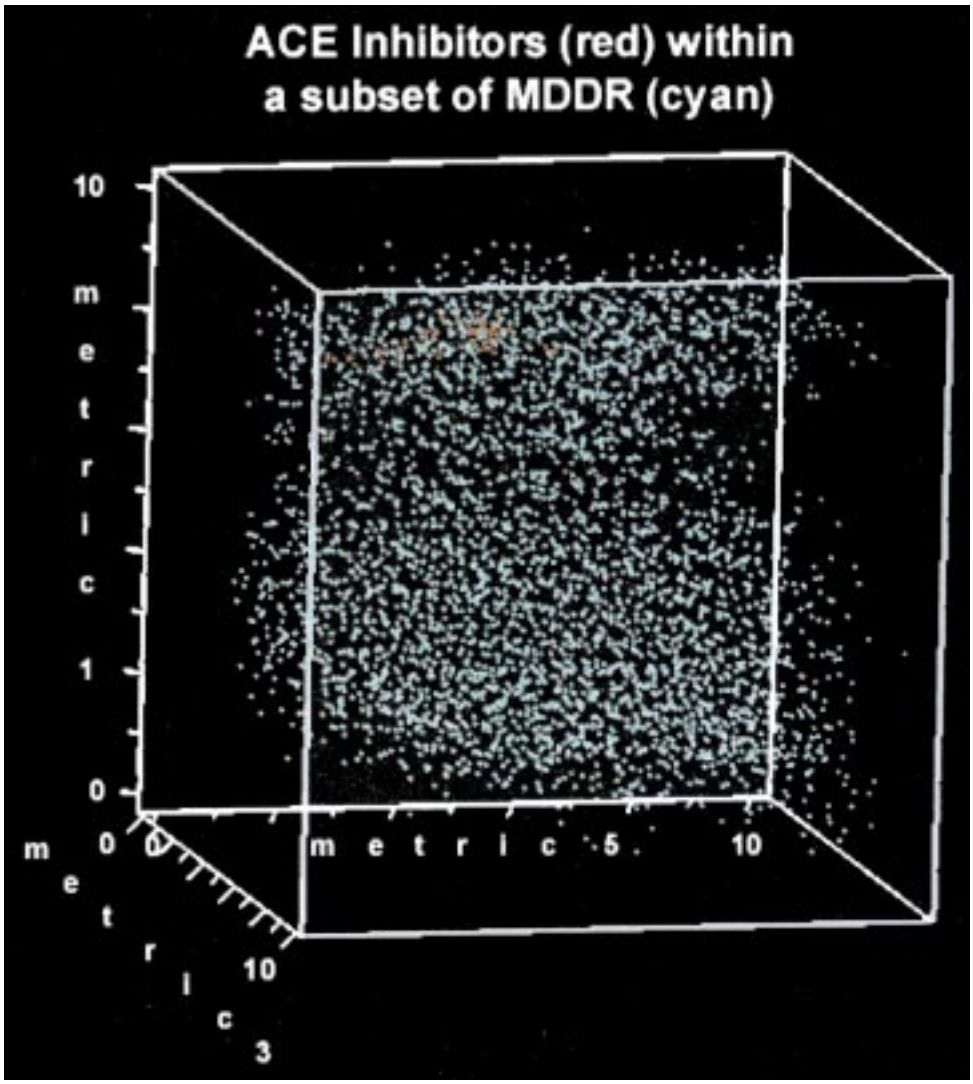

Figure 5. 74 ACE inhibitors in a three-dimensional space based on two relevant metrics (metric-1 and metric-2) and an irrelevant metric (metric-5). Note that the active compounds are tightly clustered within the respect to axes 1 and 2 but not clustered along the axis corresponding to the irrelevant metric.

testing might prevent or at least delay discovery of active compounds near the right-hand end of the cylinder. Based on distances computed in the receptor-relevant subspace, essentially all the active compounds are near neighbors of each other. Including the irrelevant metric-5 in the distance formula might also lead one to the erroneous conclusion that the BCUT metrics are not valid because active compounds appear to be quite distant from one another. However, based on distances computed in the receptor-relevant subspace, essentially all the active compounds are near neighbors of each other.

Figure 6 illustrates another important application of using receptor-relevant distances. (Once again, a two-dimensional projection is depicted merely because the three-dimensional view does not show up well when photographed.) Imagine that the (hypothetical) blue combinatorial library was screened for affinity for the ACE receptor and the resulting hits were highlighted in red. Should the (hypothetical) yellow combinatorial library be synthesized and screened? Given that there is no overlap between the yellow library and the blue library containing known actives, one might reasonably decide that screening the yellow library would be pointless. However, Figure 6 clearly indicates that the reason for which the libraries do not overlap is irrelevant to this particular receptor. Screening the yellow library could reveal additional actives (highlighted in red) which are very near the previously discovered actives within the receptor-relevant subspace but are distant from those compounds when an irrelevant metric is included in the distance formula. Please refer to Figure 5 in the article by Schnur[16] in this issue for a real-world example.

How would one ascertain that metric-5 is an irrelevant metric based only on the results of screening the blue library? After all—the actives found within the blue library appear to be tightly clustered along the metric-5 axis. The answer is simple once one realizes that the entire blue library (population) is tightly clustered along the metric-5 axis. By determining the chemistry-space which best represents the blue library, the axis with the same BCUT-type as metric-5 would be expanded to better distinguish the limited diversity of that common-core population with respect to that type of BCUT value. Close inspection of Figure 6 reveals that the actives in the blue library span the full range of metric-5 covered by the blue library. The receptor-relevant subspace identified for the blue library would also reveal this fact and would reveal which BCUT-types are relevant to this receptor.

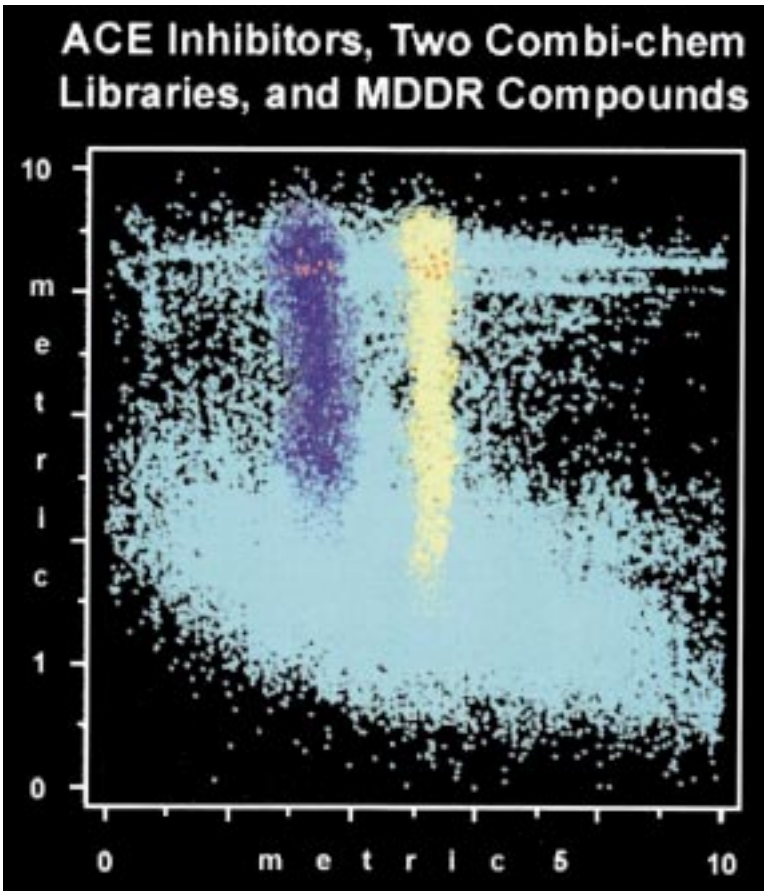

Figure 6. Two-dimensional projection of the three-dimensional subspace depicted in Figure 5. Note that the two combinatorial libraries and the active compounds found in each appear to be nonoverlapping with respect to the axis corresponding to the irrelevant metric. This might have lead to erroneous conclusions as discussed in the text.

Selecting the corresponding metric-types from the full six-dimensional space containing MDDR (and both libraries) would yield the receptor-relevant subspace of the MDDR space.

4. SUMMARY

Low-dimensional chemistry-space metrics enable the use of cell-based algorithms which are much better suited for most diversity-related tasks than the distance-based algorithms required when using high-dimensional (fingerprint) representations of chemistry-space. Traditional molecular descriptors are often cross-correlated, provide little or no substructural information, and are only weakly related to modes of drug-receptor interaction. Thus, traditional molecular descriptors are often poor choices as low-dimensional chemistry-space metrics. BCUT-values constitute a novel class of molecular descriptors which encode both substructural topological (or topographical) information as well as atomic information relevant to the strength of ligand−receptor interaction and have been shown to be quite useful as chemistry-space metrics for pharmaceutical and agrochemical diversity purposes.

Confidence in diversity software is related to confidence in the metrics on which the software is based. BCUT metrics have been validated repeatedly by various methods in various laboratories. A fundamental question to be addressed by metric validation is the extent to which metrics cluster compounds which show affinity for a given receptor. The activity-seeded, structure-based clustering algorithm described in this report directly addresses this question and is applicable with any low- or high-dimensional chemistry-space metrics when computer graphic visualization of active compounds is impossible.

Reducing the dimensionality of chemistry-space would enable visual validation of metrics and other important tasks such as library comparisons. However, dimensional reduction based upon mathematical considerations alone inevitably lead to the loss of potentially important information conveyed by the discarded metrics. The receptor-relevant subspace algorithm described in this report provides a simple, novel approach for reducing dimensionality by discarding metrics (axes) which are irrelevant to a given receptor while retaining metrics which convey receptor-relevant information. Since the receptor-relevant subspace of a full BCUT chemistry-space is often just two- or three-dimensional, computer graphic visualization is enabled. However, regardless of the dimensionality, working within the receptor-relevant subspace ensures the use of receptor-relevant intercompound distances for tasks such as identifying near neighbors of lead compounds and comparing libraries. The receptor-relevant subspace concept has been implemented in the DiverseSolutions software package and will be generally available in versions 4.1 and higher.

REFERENCES AND NOTES

(1) Jilek, R. J.; Cline, M. A.; Spencer, R. W.; Blake, J. F. Comparison of Representative and Diverse Subset Selection Algorithms Suitable for

Use in High Throughput Screening. *J. Chem. Info. Comput. Sci*, submitted for publication.
(2) Pearlman, R. S. *DiverseSolutions User's Manual*; University of Texas, Austin, TX, 1995.
(3) Pearlman, R. S. Novel Software Tools for Addressing Chemical Diversity. *Network Science* 1996, http://www.awod.com/netsci/Science/combichem/feature08.html.
(4) Pearlman, R. S. and Smith, K. M. Novel Software Tools for Chemical Diversity. *Perspectives Drug Discovery Design* **1998**, *9*, 339−353.
(5) *DiverseSolutions* was developed by R. S. Pearlman and K. M. Smith at the University of Texas, Austin TX and is distributed by Tripos, Inc., St. Louis, MO.
(6) Pearlman, R. S.; Deanda, F., manuscript in preparation.
(7) Modern Drug Data Report database is distributed by MDL Information Systems, San Leandro, CA.
(8) Sweet, C. S.; Ulm E. H.; Gross, D. M.; Vassil T. C.; Stone C. A. A New Class of Angiotensin-Converting Enzyme Inhibitors. *Nature* **1980**, *288*, 280−283.
(9) Suh, J. T.; Skiles, J. W.; Williams, B. E.; Youssefyeh, R. D.; Jones, H.; Loev, B.; Neiss, E. S.; Schwab, A.; Mann, W. S.; Khandwala, A.; Wolf, P. S.; Weinryb, I. Angiotensin-Converting Enzyme Inhibitors. New orally active antihypertensive (Mercaptoalkanoyl)- and [(Acylthio)alkanoyl]glycine derivatives. *J. Med. Chem.* **1985**, *28*, 57−66.
(10) Menard, P. R.; Suh, J. T.; Jones, H.; Loev, B.; Neiss, E. S.; Wilde, J.; Schwab, A.; Mann, W. S. Angiotensin-Converting Enzyme Inhibitors. (Mercaptoaroyl)amino acids. *J. Med. Chem.* **1985**, *28*, 328−332.
(11) Wyvratt, M. J.; Patchett, A. A. Recent Developments in the Design of Angiotensin-Converting Enzyme Inhibitors. *Med. Res. Rev.* **1985**, *5*, 483−531.
(12) Karanewsky, D. S.; Badia, M. C.; Cushman, D. W.; DeForrest, J. M.; Dejneka, T.; Loots, M. J.; Perri, M. G.; Petrillo, E. W.; Powell, J. R. (Phosphinyloxy)acyl Amino Acid Inhibitors of Angiotensin-Converting Enzyme (ACE). 1. Discovery of (S)-1-[6-amino-2-[[hydroxy(4-phenylbutyl)phosphinyl]oxy]-1-oxohexyl-L-proline novel orally active inhibitor of ACE. *J. Med. Chem.* **1988**, *31*, 204−212.
(13) Yanagisawa, H.; Ishihara, S.; Ando, A.; Kanazaki, T.; Miyamoto, S.; Koike, H.; Iijima, Y.; Oizumi, K.; Matsushita, Y.; Hata, T. Angiotensin-Converting Enzyme Inhibitors. 2. Perhydroazepin-2-one derivatives. *J. Med. Chem.* **1988**, *31*, 422−428.
(14) Krapcho, J.; Turk C.; Cushman, D. W.; Powell, J. R.; DeForrest, J. M.; Spitzmiller, E. R.; Karanewsky D. S.; Duggan, M.; Rovnyak, G.; Schwartz, J.; Natarajan, S.; Godfrey, J. D.; Ryono, D. E.; Neubeck, R.; Atwal, K. S.; Petrillo, E. W. Angiotensin-Converting Enzyme Inhibitors. Mercaptan, carboxyalkyl dipeptide, and phosphinic acid inhibitors incorporating 4-substituted prolines. *J. Med. Chem.* **1988**, *31*, 1148−1160.
(15) Karanewsky, D. S.;, Badia, M. C.; Cushman, D. W.; DeForrest, J. M.; Dejneka, T.; Lee, V. G.; Loots, M. J.; Petrillo, E. W. (Phosphinyloxy)acyl amino acid inhibitors of angiotensin-converting enzyme. 2. Terminal amino acid analogues of (S)-1-[6-amino-2-[[hydroxy(4-phenylbutyl) phosphinyl]oxy]-1-oxohexyl]-L- proline. *J. Med. Chem.* **1990**, *33*, 1459−1469.
(16) Schnur, D. Computer Aided Design and Diversity Analysis of Large Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **1999**, in press.

CI980137X