

DiverseSolutions[®]

DESIGN, COMPARE, OR SELECT COMPOUND LIBRARIES



As key technologies in the drug discovery process, high throughput synthesis and screening of compound libraries have focused attention on molecular diversity in order to enhance the efficiency and productivity of research efforts.

DiverseSolutions is a suite of chemically intuitive tools for addressing a variety of diversity-related tasks, including: compound acquisition, iterative lead follow-up, focused and diverse library design, *de novo* design, and visualization of compounds and entire libraries in chemistry-space.

Benefits

- **Reduce redundancy** in synthetic or screening efforts by selecting a diverse, representative subset of molecules from a much larger population.
- **Focus discovery efforts** by selecting a subset of molecules within a limited population, such as neighbors of lead compounds within a large database or combinatorially-generated analogs of lead compounds.
- **Identify regions of chemistry-space** that are not occupied by compounds in a library and therefore unexplored.
- **Optimize compound acquisition** by assessing the diversity of two or more databases, compound collections, or combinatorial libraries to determine if their chemistry-space is complementary or overlapping.
- **Maximize diversity** or focus on activity, while minimizing the number of reagents and robotic operations required in full-array combinatorial library design.

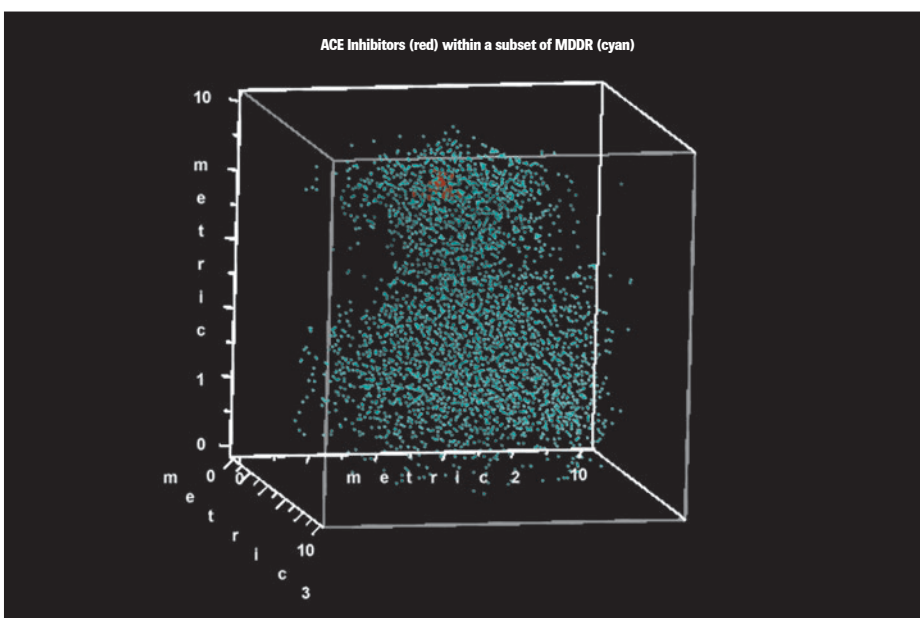
For a given population of molecules, DiverseSolutions generates chemistry-space metrics from atomic properties relevant to protein-ligand interactions (charge, polarizability, H-bonding ability, etc.) and connectivity information (bond types, interatomic distances, etc.)¹

DiverseSolutions then identifies those metrics which best distinguish structural differences between compounds and thereby best represent the diversity of the given population.

The optimal chemistry-space is partitioned into cells and the occupancy of each cell is easily determined based on the coordinates of compounds. Subset selection from millions of compounds can be accomplished

in just a few seconds simply by sampling from occupied cells. "Voids", or unoccupied areas of chemistry-space, are easily recognized. Filling in missing diversity and comparing two or more populations is similarly straightforward.

Cell-based methods for assessing diversity enable referencing to both inter-compound distance and absolute position in chemistry-space and are ideally suited for all diversity tasks. Distance-based methods are also provided in DiverseSolutions to facilitate analysis of high dimensional representations of chemistry-space (e.g., fingerprints), for rapid nearest neighbor analyses, and for subset selection.



The tight clustering of 74 ACE inhibitors (in magenta) within the three-dimensional ACE-receptor-relevant subspace of the MDDR chemistry-space. Only a 5% diverse subset of the total MDDR population is shown (in cyan). The initial 6-dimensional chemistry-space of the MDDR compounds was reduced to three dimensions by considering only those metrics relevant to the ACE receptor.

Applications

- Use both distance-based and cell-based algorithms to perform various tasks based on the coordinates of compounds in chemistry-space.
- Generate BCUT descriptors for use as chemistry-space coordinates and QSAR/QSPR descriptors.
- Enhance rational compound acquisition; buying or synthesizing compounds to fill "under-populated" and/or "promising" regions of chemistry-space.
- Assess and compare diversities of one, two or more population(s) of compounds.
- Design focused, diverse, fill-in diverse, and focused-diverse parallel and combinatorial libraries.
- Perform iterative near-neighbor searches for highly efficient lead follow-up.

Cell-based Algorithms

Cell-based algorithms involve *binning* (dividing) each axis of a multi-dimensional chemistry-space, yielding a large number of hypercubic cells which increases exponentially with the dimensionality of the chemistry-space. Thus, for practical purposes, cell-based algorithms can only be applied in relatively low-dimensional chemistry-spaces. However, whereas distance-based algorithms only consider inter-compound distances, cell-based algorithms not only consider inter-compound distance but also consider the absolute positions of compounds in chemistry-space, making them far more useful for a wide range of diversity- and similarity-related tasks.

In order to benefit from cell-based algorithms, the axes of a low-dimensional chemistry-space must be associated with descriptors which are truly relevant to ligand-receptor interaction. Since traditional descriptors such as molecular weight, surface area, and even logP convey little or no information about substructural details important for understanding ligand-receptor interaction, they are of little or no use as axes of a meaningful, chemistry-space. In contrast, the novel and unique BCUT™ descriptors do convey substructural information and have proven, over the years, to be extremely useful for a wide range of CAMD applications.

BCUT Descriptors

BCUT descriptors are very powerful, structure-based molecular descriptors. Although some users have found them useful as QSAR and QSPR descriptors, BCUT descriptors are primarily useful as the coordinates of drug-sized compounds in a low dimensional chemistry-space. BCUT descriptors have, repeatedly, been shown to perform well in clustering compounds

known to be active at a particular target. Thus, BCUT descriptors are very useful as indicators of where, in chemistry-space, to look, or not to look, for novel actives. Hence, to some extent, BCUT descriptors convey the same sort of information as would be conveyed by the pharmacophore of the target of interest.

BCUT descriptors provide rough indications of how atomic properties relevant to ligand-receptor interaction are distributed across the potential ligand's molecular surface. More specifically, BCUT descriptors describe the surface distributions of positive charges, negative charges, H-bond donors, H-bond acceptors, regions of high polarizability ("greasiness"), and regions of low polarizability. BCUT descriptors for a particular compound are computed by first forming $N \times N$ matrices (where N is the number of atoms in the given molecule). The diagonal elements of the matrices are assigned the values of atomic descriptors related to the distributions listed above. The off-diagonal elements are assigned values related to either the topological or topographical distance between pairs of atoms. These two types of information are "combined" through a process called *matrix diagonalization* and the resulting eigenvalues are used to form the BCUT descriptors. However, since the matrices contain two different types of information (atomic property information and "connectivity" information), it is absolutely essential that carefully determined scaling-factors are applied to the diagonal elements prior to diagonalization.

Chemistry-space

A "chemistry-space" is a vector-space comprised of axes which are molecular descriptors related to molecular structure. The descriptor values of various compounds can be regarded as coordinates in this multi-

dimensional vector-space and the similarity or diversity of compounds can be expressed in terms of inter-compound distances. In order to be able to compute meaningful distances, the axes of the chemistry-space (i.e., the descriptors) must be normalized and must be mutually orthogonal.

Advantages

- The BCUT metrics calculated by DiverseSolutions reflect substructural differences and are based on properties relevant to intermolecular interactions. As a result, they better describe chemistry-space than traditional whole-molecule metrics such as logP, surface area, and dipole moment.
- The cell-based representation of chemical diversity facilitates biased sampling, in which a maximally diverse subset is selected based on factors in addition to structural diversity. Property-biased subset selection and biased compound acquisition are possible because you can easily compare different compounds representing the same region of chemistry-space.
- A novel reactant-biased, product-based library design algorithm enables design of full array combinatorial libraries that satisfy user-specified limits on the number of reactants used and the robotic operations performed on plates on the robot-table.
- DiverseSolutions performs receptor-relevant subspace² perception by identifying the subset of metrics which describe structural features important for affinity to a particular receptor. This provides a rational approach to reducing the dimensionality of chemistry-space. The tight clustering of active compounds can be visualized, which aids in focusing discovery efforts where active compounds are most likely to be.

Features

- Low-dimensional chemistry-space construction using metrics that best represent the diversity of a particular population of compounds.
- Active compound cluster visualization and relative positioning using meaningful coordinates in a receptor-relevant subspace.
- Diversity voids identification and fill-in for corporate screening collections.
- Full-array combinatorial library design that optimizes objectives and provides synthetic economy.
- Novel list-based near-neighbor search algorithm specifically developed for lead follow-up and iterative searching.
- Compound library comparison with sensible, intuitive indicators of chemistry-space overlap.
- Property-biased subset selection based on structural diversity.
- Imported descriptors, fingerprints, and properties such as logP, toxicity, sample availability, cost, etc.
- 166 MACCS SS-keys and corresponding vector generation.
- Powerful duplicate structure detection.
- UNIX-based tool kit, enabling power-users to build customized applications.

Validation

DiverseSolutions has been used to assess the diversity of sets of compounds with measured affinities for a variety of receptors. Compounds that are active against the same target cluster together, indicating that the metrics generated by DiverseSolutions are truly indicative of structural features relevant to bioreceptor interactions.³ Chemistry-spaces determined by DiverseSolutions have been shown to be transferable from one population of molecules to many others.⁴

Hardware and Software Requirements

DiverseSolutions requires a separate license, and runs on workstations operating under IRIX® (SGI®) or Linux® (x86).

Complementary Software

Integration of DiverseSolutions in the SYBYL expert molecular modeling environment allows users to access SYBYL's powerful molecular design and analysis tools.

- **Selector™** for characterizing and sampling compound libraries.
- **ClogP/CMR** for including molar refractivity and logP in QSAR and ADME models.
- **Concord®** for generating accurate 3D coordinates.
- **Molconn-Z™** for computing a wide range of topological indices based on molecular structure.
- **StereoPlex®** for expanding the stereochemical diversity of a database.
- **Legion™/CombiLibMaker™** for building virtual combinatorial libraries in cSLN format.
- **UNITY®** for locating compounds in databases that match a pharmacophore or fit a receptor site.

Acknowledgements

Scientific Partner:
Professor Robert S. Pearlman

References

1. Pearlman, R.S.; Smith, K.M. "Novel Software Tools For Chemical Diversity." *Perspectives in Drug Discovery and Design*. **1998**, 9/10/11, 339-353.
2. Pearlman, R.S.; Smith, K.M. "Metric Validation And The Receptor-Relevant Subspace Concept." *J. Chem. Inf. Comput. Sci.* **1999**, 39, 28-35.
3. Schnur, D. "Design And Diversity Analysis Of Large Combinatorial Libraries Using Cell-Based Methods." *J. Chem. Inf. Comp. Sci.* **1999**, 39, 36-45.
4. Menard, P.R.; Mason, J.S.; Morize, I.; S. Bauerschmidt. "Chemistry Space Metrics In Diversity Analysis, Library Design, And Compound Selection." *J. Chem. Inf. Comp. Sci.* **1998**, 38, 1204-1213.



WWW.TRIPOS.COM				CONTACT_US@TRIPPOS.COM		
AUSTRALIA +61 (7) 5439 9775	CANADA +1 450 4334500	FRANCE +33 1 69 59 29 49	GERMANY +49 89 45 10 300	JAPAN +81 3 5166 1721	UNITED KINGDOM +44 1 908 650000	UNITED STATES 800 323 2960 +1 314 647 1099
©Tripos, Inc. 2005 0075 Printed in USA All trademarks are the property of their respective owners.						