Pacific Symposium on Biocomputing 4:456-467 (1999)

LIGAND-RECEPTOR 3-D SIMILARITY STUDIES USING MULTIPLE 4-POINT PHARMACOPHORES

Jonathan S. MASON, Daniel L. CHENEY
Computer-Assisted Drug Design, Department of Macromolecular Structure,
Bristol-Myers Squibb, PO Box 4000, Princeton NJ 08543, USA,
jonathan_mason@ccmail.bms.com

A new method for 3-D similarity is presented based on the multiple potential 4-point 3-D pharmacophores expressed by ligands and complementary to receptors. These are calculated for ligands taking conformational flexibility into account, and for receptors through the use of complementary site-points. Through this common frame of reference both ligand-ligand and ligand-receptor similarity studies are possible. The application of the method to selectivity between different serine proteases (thrombin, factor Xa and trypsin) is discussed, and the need to use 4-point pharmacophores rather than 3-point pharmacophores is illustrated. A novel refinement to the potential pharmacophore method that uses a “special” feature to give a relative measure of similarity/diversity is also discussed.

1 Introduction

Methods for molecular similarity and diversity that use properties that can be calculated for both ligands and receptors and are needed for many computer-aided drug design (CADD) applications. These methods need to be able to handle rapidly large numbers of structures, often of a relatively high conformational flexibility, with applications for analysis and design such as virtual screening and combinatorial chemistry library design.[1,2,3]

3-D potential pharmacophores, consisting of triplets of distances between features considered to be important for ligand-receptor interactions, can be calculated systematically for structures using software such as the ChemDiverse module of Chem-X.[4] A recent extension to this software is the important move to 4-point pharmacophores, giving a significant increase in the amount of shape information and resolution, including the ability to distinguish chirality, a fundamental requirement for many ligand-receptor interactions. For a protein site, complementary site-points to groups in the site are generated and potential pharmacophores generated from all combinations of four site-points. For a ligand, the pharmacophoric features (hydrogen bond donors, hydrogen bond acceptors, acidic centers, basic centers, hydrophobic regions and aromatic ring centroids) are automatically identified and explicit "on-the-fly" conformational sampling is performed to enable the calculation of the potential 4-point pharmacophores (all geometries accessible for all combinations of four features).

These potential 4-point pharmacophores (~13 million possible with 10 ranges per feature-feature distance for combinations from 6 features) give a common frame

of reference for comparing different ligands and for comparing ligands to protein structures using the complementary potential pharmacophores. Studies that were used as part of the validation of this new method are presented for ligand-ligand comparisons using endothelin receptor antagonists, and for ligand-receptor interactions using serine protease inhibitors, with an emphasis on selectivity.

A further refinement of the method is to limit the definition of the potential pharmacophores to be only those that contain a particular feature, either one of the six normally used or an additional "special" feature that can be added to any substructure or site-point. This enables many powerful new design and analysis methods, such as the design of combinatorial libraries focused around "privileged" substructures found in many ligands for seven trans-membrane G-protein coupled receptors.[5,6] The use of this refinement in ligand-receptor studies is illustrated with the serine proteases, with an enhanced inter-enzyme selectivity predicted from the potential 4-point pharmacophore overlaps.

2 Methods

2.1 Chem-X software

The Chem-X[4] software is a general molecular modeling package, with specialist optional modules such as ChemDiverse and 4-centre pharmacophores that were used for this work. 3D structures were generated for ligands using the CONCORD[7] program, and were read into a Chem-X/ChemDBS-3D database from an SD file using a customized parameterization file and fragment database to assign atom types, as discussed below in section 2.2.. A single conformer is stored in the database, and conformational sampling is done "on the fly"; see section 2.2 for further details of the settings used. The studies were performed on a Silicon Graphics R10000 workstation; an identical version of the software is available for PC Windows95/NT.

2.2 Generation of the 4-point pharmacophores

Six key features that are likely to be important for drug-receptor interactions are automatically identified for each molecule through the use of atom types, with the addition of dummy atoms for hydrophobic regions and aromatic rings. The six features that can be assigned to an atom type are hydrogen bond donor, hydrogen bond acceptor, acidic center (negatively charged at physiological pH 7), basic center (positively charged at pH 7), hydrophobic region and aromatic ring centroid. The atom types are automatically assigned when reading a molecule into Chem-X, through a customizable parameterization file and fragment database.[3,8] The dummy

atoms that are used to represent the hydrophobic regions are added by an automatic method within Chem-X that uses bond polarities (hydrophobic regions defined for groups of three or more atoms that are not bonded to atoms with a large electronegativity difference).

For an enzyme active site or a receptor site complementary site-points are added (as atoms, dummy atoms or functional groups), and relevant features assigned to these points. The site-points can be generated by many different methods, such as geometric ones (e.g. in Chem-X/ChemProtein where complementary site-points for all residue types are defined in template fragments and positioned relative to the residue atoms, or based on crystallographically determined positions of high or maximum probability), or via energetic surveys of the site, using a variety of probe atoms (as implemented in the GRID program[9]). For example, a dummy atom site-point assigned the hydrogen bond acceptor feature may be placed at a hydrogen bonding distance from sterically accessible N-H groups. The combined set of all site-points represents a theoretical molecule that binds to all available positions, and potential pharmacophores are generated for this “molecule” in the same way as for a normal compound. The positions of potential interacting groups can be refined using a force-field molecular dynamics simulation and/or minimization. The studies presented here used site-points from GRID probe analyses, where from the resultant energetically contoured maps, complementary site-points representing hydrogen bond acceptor, donor, acceptor and donor, acidic, basic, hydrophobic or aromatic interactions were located and atoms added that were assigned the relevant feature(s).

All combinations of four features (from the assignments to atoms in ligands or site-points) are considered, together with 7 or 10 distance ranges for each of the six distances (for 7 ranges: <2.5, 2.5-4, 4-6, 6-9, 9-13, 13-18, >18 Å). A pharmacophore “key” is thus generated that indicates the presence or absence of all the theoretically possible combinations of features and distances (potential pharmacophores); an additional chirality indicator can be added to applicable potential pharmacophores, and this was done for the work reported in this paper. About 3 million (7 distance ranges) or 13 million (10 distance ranges) 4-point potential pharmacophores are thus considered for each ligand or set of site-points, as illustrated in figure 1. For ligands, which normally have conformational flexibility, an effective conformational sampling is needed and is used for the potential pharmacophore analysis. The method used here is a customization of the Chem-X/ChemDiverse method, based on an "on-the-fly" generation of conformers done at search time, with a quick evaluation of the conformation performed based on a steric contact check to reject poor or invalid conformations. Using 3 rotamers for single bonds, 4 for alpha bonds (sp2-sp3, e.g. Ph-CH2) and 2 or 4 for conjugated bonds, the total analysis time is only up to 5–15 seconds on a Silicon graphics R10000 processor, using a systematic analysis where possible and a random analysis for very flexible molecules (method automatically selected according to whether a systematic analysis is predicted to finish in the maximum allowed time). Even with very flexible molecules, these sampling conditions were found to identify most of the

exhibited potential 4-point pharmacophores with a resolution of 7 distance ranges (giving 3 million possible), and was suitable for general virtual screening purposes. For more precise analyses, 6 rotamers were generated for the single and alpha bonds, increasing the maximum time by 5 or 10.

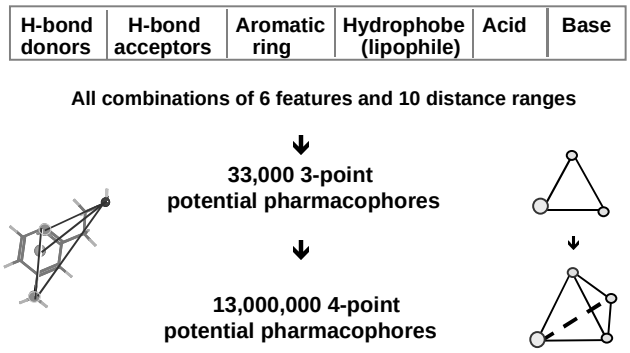


Figure 1: The multiple potential 3-D pharmacophore method

2.3 Comparisons of the 4-point pharmacophores

Logical operations on the pharmacophore "keys" were performed in Chem-X using no "tolerances" (i.e. an exact match in terms of features and distance ranges required). To identify potential pharmacophores that were common between a ligand and an enzyme active site, the potential 4-point pharmacophore keys were generated (1) for the ligand (with full conformational sampling) and (2) complementary to the active site (using the complementary site-points as a pseudo-molecule), and (3) a logical AND operation applied between the two keys. The resultant key of overlapping potential pharmacophores could be further analyzed in Chem-X, or written out to a file.

3 Results

3.1 Ligand-ligand 3D similarity studies

As an example of the power of the method to find similarity between compounds with a similar biological activities, two endothelin receptor antagonists with about 20 nM activity as antagonists of the ETA receptor were compared (see figure 2). The two compounds have very low 2D similarity, but have significant overlap of their 4-point potential pharmacophores. 2959 common 4-point potential

pharmacophores were found, from a total of 6904 for the smaller compound. The ability of the pharmacophore method to identify and focus on features important for drug-receptor interactions was important for this result; for example the assignment of the acidic feature to the acylsulfonamide group affects increases the overlap by about a third (acids were also considered as general hydrogen-bond acceptors for this analysis).

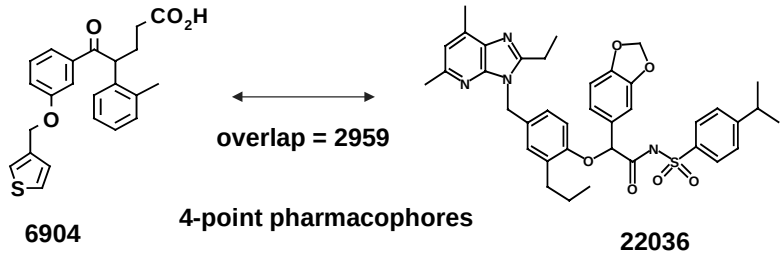


Figure 2: Ligand-ligand 3D similarity: Total and common (overlap) multiple 4-point potential 3-D pharmacophores for two potent endothelin antagonists

3.2 Ligand-receptor 3D similarity studies

The multiple potential pharmacophore key calculated from a ligand can also be compared to the multiple potential pharmacophore key of complementary site-points in its target binding site. The 4-point pharmacophore method thus provides a novel method to measure similarity when comparing ligands to their binding site targets, with applications such as virtual screening and structure-based combinatorial library design.

An example of the method comes from studies on three closely related serine proteases thrombin, factor Xa and trypsin. 4-point multiple potential pharmacophore keys were generated from site-points positioned in the active sites using the results of GRID analyses (see Figure 3). Keys were also generated using full conformational flexibility for some highly selective and potent thrombin and factor Xa inhibitors (see Figure 4). We thus investigated whether receptor-based similarity as a function of common potential 4-point pharmacophores for each ligand/receptor pair could resolve enzyme selectivity. The expected benefit of using 4-point pharmacophores with their improved shape information was probed by using identical studies with only 3-point pharmacophores. The goal of the studies were to see if common potential pharmacophores could give information pertaining to relative enyzme selectivity, not to predict binding affinities, the goal of the work of many others.[10,11,12]

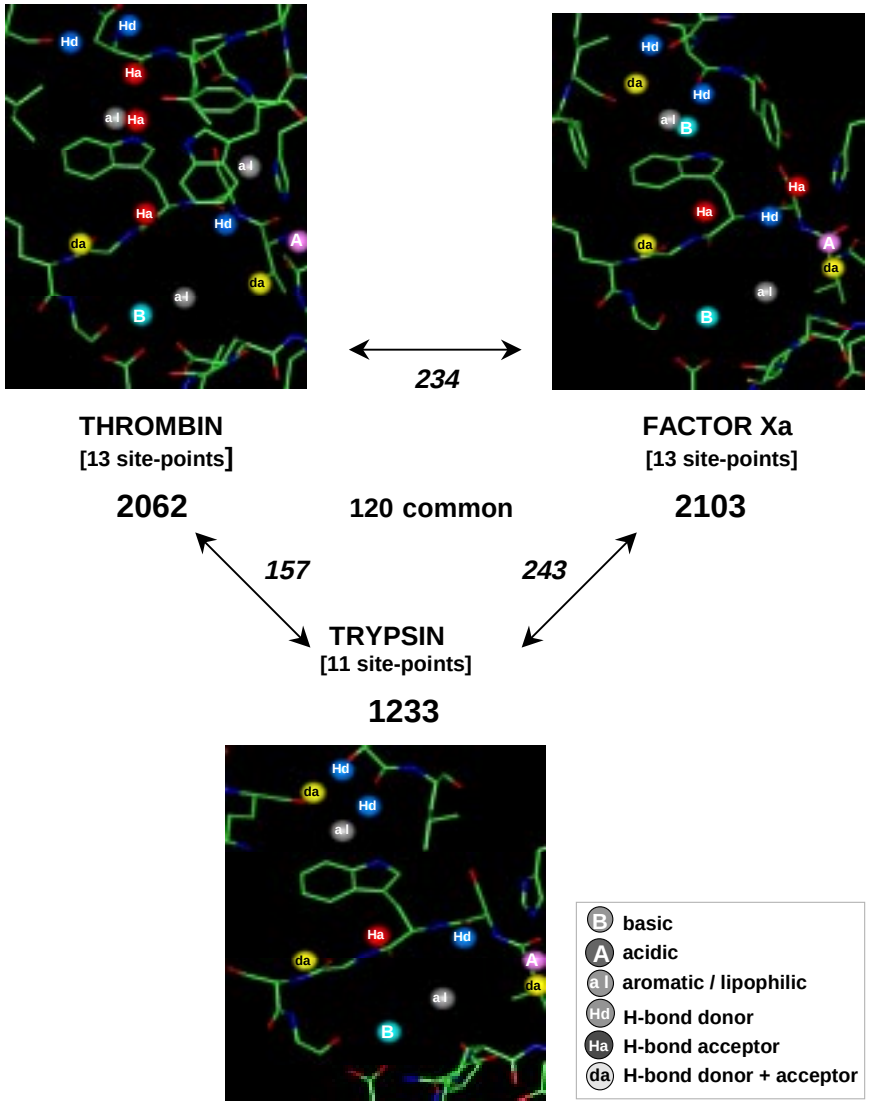
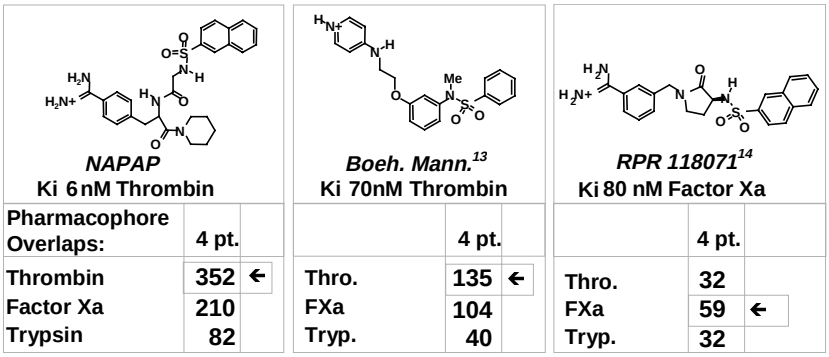


Figure 3: Number of potential 4-point pharmacophores calculated on the basis of complementary site-points inserted into the active sites of thrombin, factor Xa, and trypsin, and number of overlapping pharmacophores (pair-wise and for all 3 serine protease sites)

The results shown in figure 4 indicate that the use of just 4-point potential pharmacophores gives correct indications as to relative selectivity for this set of

related enzymes. The thrombin and factor Xa inhibitors exhibit greater similarity with the complementary 4-point potential pharmacophore keys of the thrombin and factor Xa active sites, respectively, than with the potential pharmacophore keys generated from the other enzymes. Clearly other factors such the strength of hydrogen bonds and hydrophobic interactions will affect actual binding energies, but for this set of ligands these relative overlap numbers alone are sufficient to indicate relative selectivity.



NAPAP
Ki 6 nM Thrombin

Pharmacophore Overlaps:	4 pt.	
Thrombin	352	←
Factor Xa	210	
Trypsin	82	

***Boeh. Mann.*[13]**
Ki 70nM Thrombin

	4 pt.	
Thro.	135	←
FXa	104	
Tryp.	40	

***RPR 118071*[14]**
Ki 80 nM Factor Xa

	4 pt.	
Thro.	32	
FXa	59	←
Tryp.	32	

Figure 4: Number of potential 4-point pharmacophores for ligands that overlap with those calculated for the active sites of thrombin, factor Xa, and trypsin on the basis of complementary site-points; the arrow points to the enzyme for which the ligand shows biological activity (Ki for binding given in top box). Using 4-point pharmacophores each ligand has more pharmacophores in common with the enzyme for which it shows binding than with the other enzymes. The figures thus show that for these ligands enzyme selectivity can be resolved in terms of relative numbers of common potential 4-point pharmacophores for each ligand/enzyme pair.

3-point pharmacophore-based similarity gave however poor resolution of enzyme selectivity (see Figures 5 and 6), the enhanced resolution of the 4-point method thus being needed for comparisons based just on pharmacophore overlap, without taking any further account of the shape of the site. It would clearly be advantageous to take into account the shape of the site for each ligand-receptor pharmacophore match, for example to reject matches where a steric clash with the site cannot be avoided, and through a further collaboration with Chemical Design[4] this concept has been put into a new DiR (Design in Receptor) module that is now becoming available;[15] the method is equivalent to doing multiple 3D database pharmacophoric searches, using each potential 3- or 4-point pharmacophore as a 3D query, but with only one conformational sampling step per molecule.

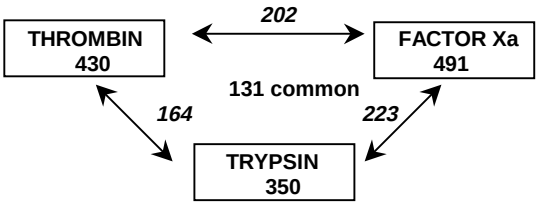
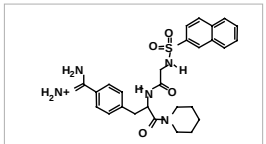
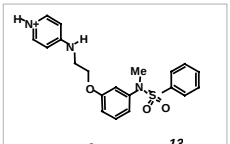


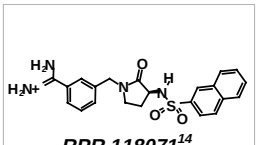
Figure 5: Potential 3-point pharmacophores calculated from the complementary site points inserted into the active sites of thrombin, factor Xa, and trypsin. The figures show the total number for each active site, pair-wise overlaps, and the common for all three active sites. Compared to the results for 4-point pharmacophores, there are similar numbers of overlapping (common) pharmacophores, but many less "unique" (non-overlapping) pharmacophores (e.g. 195 compared to 1811 for thrombin).



NAPAP
Ki 6 nM thrombin



***Boeh. Mann.*[13]**
Ki 70 nM thrombin



***RPR 118071*[14]**
Ki 80 nM Factor Xa

Pharmacophore Overlaps:	**3 pt.**	
Thrombin	**168**	←
Factor Xa	**144**	
Trypsin	**99**	

	3 pt.	
Thro.	**58**	←
FXa →	**64**	
Tryp.	**31**	

	3 pt.	
Thro. →	**79**	
FXa	**56**	←
Tryp.	**49**	

Figure 6: Number of potential 3-point pharmacophores for ligands that overlap with those calculated for the active sites of thrombin, factor Xa, and trypsin on the basis of complementary site-points; the right-side arrow and box points to the enzyme for which the ligand shows binding. Using 3-point pharmacophores two of the ligands have more pharmacophores in common with an enzyme different from the one with which it shows (selective) binding. The figures thus show that poor enzyme selectivity is resolved in terms of relative numbers of common potential 3-point pharmacophores; the left side arrows show where incorrect selectivity is indicated.

To evaluate this method for ligand-receptor similarity in the context of compound design and virtual screening the above analysis was repeated using two fibrinogen receptor antagonists (see figure 7). These compounds have 2D structural features (e.g. benzamidine) that resemble trypsin-like serine protease inhibitors, but had no reported activity for this class of enzymes. With 4-point pharmacophore profiling the degree of similarity is very small, whereas using 3-point pharmacophores the molecules exhibited pharmacophoric similarity against all three

enzymes, one being significant, with figures equivalent to that observed for the Boeh. Mann. thrombin inhibitor.

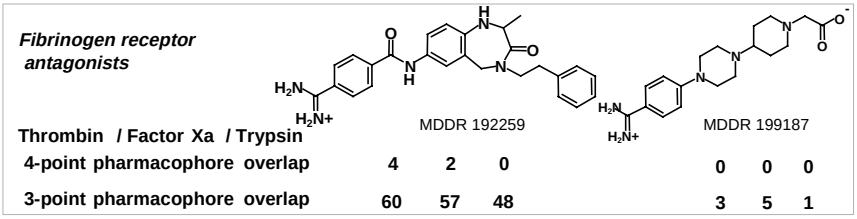


Figure 7: Number of common potential 4- and 3-point pharmacophores for inactive ligands with similar 2D structures to active compounds with those calculated from the complementary site-points for the active sites of thrombin, factor Xa, and trypsin; the ability of the 4-point pharmacophore method to distinguish these compounds is shown by both compounds.

3.3 Ligand-receptor studies using internally referenced "relative" similarity

The multiple potential pharmacophore keys calculated from ligands and from site-points can be further refined by using the concept of "relative" or "internally referenced" similarity/diversity.[5] The 4-point pharmacophore definition is modified to force one of the points to be the "special" feature, which can be assigned to a to an existing feature of interest, a dummy atom on a substructure or to a site-point of interest, as illustrated in figure 8. It is thus possible to readily measure and compare pharmacophoric shapes containing this special feature, again with a common frame of reference between different ligands and between ligands and receptor sites. For example in structure-based library design the method can be used to identify all complementary potential pharmacophores relative to an attachment point, and use this information to identify reagents that would exhibit these pharmacophores (relative to their attachment point).

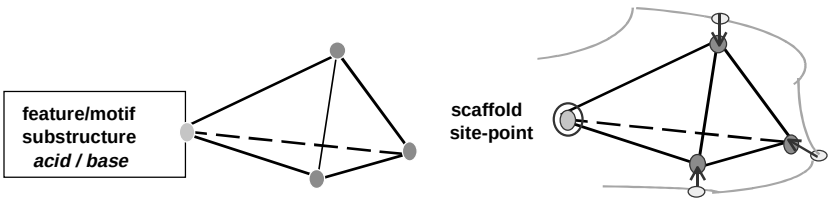
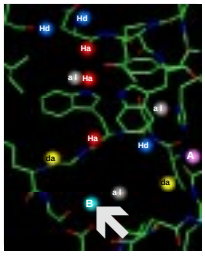
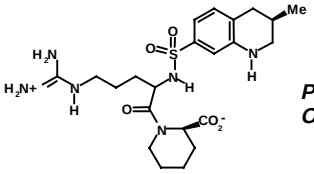


Figure 8: Modified definition of relative 4-point pharmacophores, for a ligand and for site-points; all potential pharmacophores considered will contain the "special" feature (shown on left side).

To illustrate this method the 4-point pharmacophores for the serine proteases were internally referenced to the "P1" basic group position, i.e. only potential pharmacophores containing this site-point were considered. This was found to give increased resolution of selectivity, as shown for the thrombin inhibitor MQPA in figure 9.



Thrombin
475 4-point relative pharmacophores



MQPA
Ki 19nM thrombin

Pharmacophore Overlaps:		**3 pt.**	**4 pt.**	**4pt. relative**	
Thrombin:		**183**	**192**	**76**	←
Factor Xa:	→	**199**	**191**	**50**	
Trypsin:		**128**	**94**	**34**	

Figure 9: Number of common potential 3-point, 4-point and 4-point relative (to P1 basic group) pharmacophores for the MQPA thrombin ligand with those calculated from the complementary site-points for the active sites of thrombin, factor Xa, and trypsin; the left side arrow indicates the incorrect indication of factor Xa selectivity from the 3-point figures, and the right side arrow the observed activity and the increased resolution of selectivity using the 4-point relative pharmacophores.

4 Conclusion

The 4-point multiple potential 3-D pharmacophore method provides new approaches to measure the similarity (and diversity) for compounds in terms of both ligand-ligand comparisons and ligand-receptor interactions. This can be further enhanced using a new "relative" measure of similarity and diversity wherein only a subset of pharmacophoric shapes that contain a special feature is considered.

Acknowledgements

The authors wish to thank Isabelle Morize, Paul Menard, Stephen Pickett and Richard Labaudiniere at Rhone-Poulenc Rorer for useful discussions on the multiple potential pharmacophore method.

References

1. J.S. Mason and S.D. Pickett, “Partition-based Selection”. *Perspectives in Drug Discovery and Design*, **7/8**, 85-114 (1998)
2. M.J. Ashton, M.C. Jaye and J.S. Mason,” New Perspectives in Lead Generation II: Evaluating Molecular Diversity” *Drug Discovery Today*, **1,** 71-78 (1996).
3. S.D. Pickett,, J.S. Mason and I.M. McLay, “Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries (PDQ)”. *J. Chem. Inf. Comput. Sci***. 36**, 1214-1223 (1996).
4. Chemical Design Limited: Roundway House, Cromwell Park, Chipping Norton, Oxfordshire OX7 5SR, U.K.
5. J. S. Mason, I. Morize, P.R. Menard and D.L.. Cheney, “A new 4-point pharmacophore method for molecular similarity and diversity applications, Part I: Overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures”, manuscript submitted 1998.
6. B.E. Evans. et al., “Methods for Drug Discovery: Development of Potent, Selective, Orally Effective Cholecystokinin Antagonists”, *J. Med. Chem.*, **31**, 2235-2246 (1988).
7. R.S. Pearlman,*CDA News*, **2**, 1-7 (1987); , R. Balducci, C.M. McGarity, A. Rusinko III, J.M. Skell, K. Smith and R.S. Pearlman, University of Texas at Austin, CONCORD is available from Tripos Inc, St. Louis, Missouri.
8. J.S. Mason, “Experiences in searching for molecular similarity in conformationally flexible 3D databases”, In *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Blackie Academic and Professional: Glasgow, pp. 138-162 (1995).
9. Molecular Discovery Limited: West Way House, Elms Parade, Oxford OX2 9LL, U.K.
10. P. H. Lee, G.M. Maggiora, W. J. Howe, K.W. Watenpaugh, P.K. Tomich, B. Honig, “Binding Energy Calculations of HIV-2 Protease Inhibitions”**,** 213th ACS National Meeting, San Francisco, April 13-17 (1997).
11. R.D. Head, M.L. Smythe, T.I. Oprea, C.L.Waller, S.M. Green, G.R. Marshall, “VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands.” *J. Am. Chem. Soc*. **118**, 3959-69 (1996).
12. T. Hansson, J. Marelius, J. Aqvist, “Ligand Binding Affinity Prediction by Linear Interaction Energy Methods*”, J. Comput.-Aided Mol. Des.*,**12**, 27-35 (1998).

13. W. von der Saal, R. Kuczierz, H. Leinert, and R. Engh, "Derivatives of 4-Amino-Pyridine as Selective Thrombin Inhibitors", *Bioorg. Med. Chem. Lett*. **6**, 2387 (1996).

14. W.R. Ewing, Y-M. Choi, M. Becker, V. Manetta, D. Green, R. Davis, H. Mason, C. Ly, D. Cha, D. McGarry, W. Dankulich, A. Spada, D. Cheney, J. Mason, K. Brown, D. Colussi, V. Chu, R. Bentley, S. Morgan, C. Heran, R. Leadley, C. Dunwiddie, M. Perrone, H. Pauls, "Sulfonamidopyrrolidinones: Design, SAR and parenteral activity of a novel class of factor Xa inhibitors". Book of Abstracts, 215th ACS National Meeting, Dallas, March 29-April 2 (1998).

15. http://www.chemdesign.com/prod_dir.htm