Topography-biased compound library design: the shape of things to come?

Irini Akritopoulou-Zanze, James T. Metz and Stevan W. Djuric

Medicinal Chemistry Technologies and Structural Biology, Advanced Technologies, Abbott Laboratories, Abbott Park, IL 60064, United States

The design and synthesis of quality compound libraries is of critical importance to any pharmaceutical company that relies on high throughput screening efforts for the identification of lead compounds. In this perspective, we use a moment of inertia derived shape analysis to interrogate potential libraries for chemical synthesis. An analysis of known 'Rule of Five' compliant drug shapes using this methodology clearly highlights compound libraries that may be reasonably expected, shape wise, to interact with biologically relevant protein active site topography and those that, although being structurally diverse in shape, have little chance of being pharmacologically productive. The use of multicomponent reactions as a means of producing structurally novel, bioactive compounds in a synthetically expeditious manner is also highlighted.

Introduction

Recently, most major pharmaceutical companies have invested considerable effort and dollars in improving the quality, size, and diversity of their corporate compound screening collections. These companies have reportedly spent a combined figure of well over a billion dollars to individually amass collections of between one and three million compounds with Pfizer being a well-publicized example [1]. Jacoby *et al.* have described analyses and processes to enhance the corporate compound collection at Novartis [2]. The Novartis effort includes both internal combinatorial synthesis and compound acquisition.

A key issue that companies face in this area is the selection of compounds to synthesize or acquire. For example, in establishing a relatively optimized collection one needs to consider not only near term but also future protein family targets. Not unexpectedly, the number of potential future targets has been the subject of substantial debate [3]. The current number of targets that oral small-molecule drugs 'hit' is 186 with 25–30% of these being GPCR based targets [4]. Pfizer scientists have suggested that there are approximately 1000 possible drug targets [5]. Interestingly, the drug binding domains of 399 targets (including 120 successful targets) are represented by 130 protein families, nearly a half of which are represented by six families [6]. Scientists at Lexicon have suggested, on the basis of murine KO data that there may be around 150 quality new targets left, which are to be interrogated [7]. Importantly, it has also been suggested recently that not more than 24% of the undiscovered targets will be from protein families not represented by known targets [8]. It is possible, that one may be left with the awkward task of trying to design compounds to interact with certain proteins of, as yet, unknown structure or function, in certain instances. The challenge, therefore, is how to design a compound collection that will meet the needs of current and future targets. Our ideas on the subject have been influenced by the work of Sauer and Schwarz at Serono who published two papers concerning library design using a moment of inertia (MoI) derived shape model [9,10]. Our working hypothesis is that the shapes of ligands, which will bind to future targets, are likely to be similar to the shapes of known bioactive ligands, which surprisingly, tend to adopt a somewhat limited range of MoI shapes.

Methods and results

The calculation of MoI shapes began with generation of 3D conformation of a small molecule or the utilization of existing 3D conformation using Pipeline Pilot [11].

The diagonalized components—Ixx, Iyy and Izz, of the moment of inertia tensor were calculated using routines written in Pipeline Pilot Script. Normalized ratios of the Ixx and Iyy

Corresponding author: Djuric, S.W. (stevan.w.djuric@abbott.com)

1359-6446/06/$ - see front matter © 2007 Elsevier Ltd. All rights reserved. doi:10.1016/j.drudis.2007.08.017

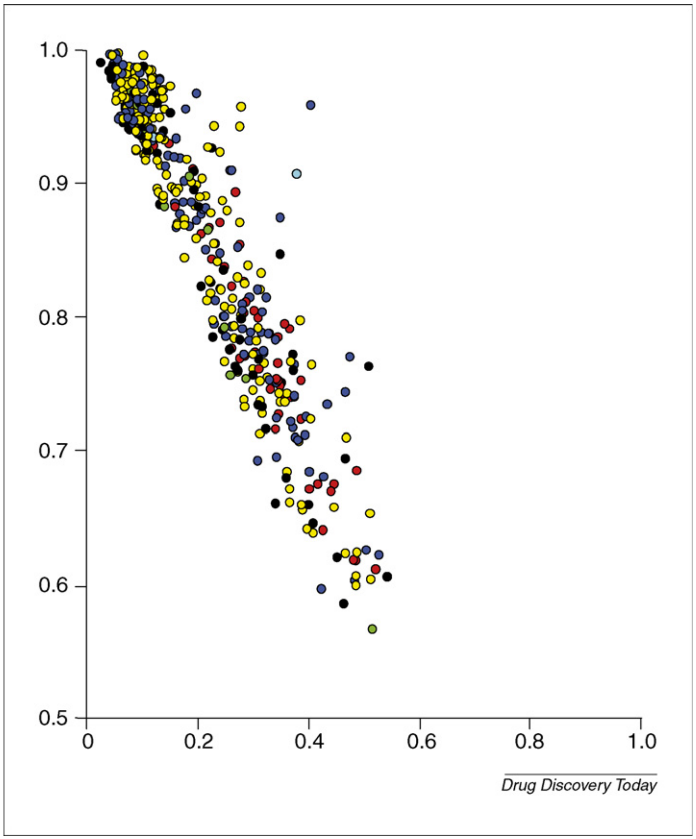

FIGURE 1

Distribution of MoI-derived shapes of 502 GPCR antagonist ligands derived from 13 GPCR homology protein structures.

components were obtained by dividing by the Izz component. The resulting normalized principal moments of inertia were plotted in two dimensions resulting in a triangular scatter plot, which facilitates the comparison of different compound sets of varying shape. This rather rudimentary approach views molecules as combinations of three types of shapes—rods (upper left corner), spheres (upper right corner) and 'pancakes' or discs (bottom). It should be noted that the absolute spatial extent of the shape is not considered in the 2D analysis, that is, spherically shaped molecules such as methane, adamantane and C_{60} all occupy the same upper right corner point. It is straightforward to construct a 3D space using molecular weight as an extra third dimension. This effectively separates molecules with the same shape, but different spatial extents. In our analyses (not reported) we have used the molecular weight dimension in an approximate way by using molecular weight range filters to reduce ambiguities involving both shape and spatial extent. Sauer and Schwarz reported that shape space coverage was found to originate mainly from the nature and the 3D geometry of the central scaffold, while the number and nature of the peripheral substituents and conformational aspects were shown to be of minor importance. Essentially, the diversity of a library is derived from the scaffolds used and their substituent patterns rather than the nature of the substituents/monomers used to construct the library. Building on the work of Sauer and Schwarz, we chose to evaluate whether focused libraries of compounds for a specific protein family exhibited particular shape preferences, for example, GPCR antagonists and kinase inhibitors (Figures 1 and 2). Furthermore, we also analyzed the MoI distribution of marketed, orally active, Rule of Five (Ro5) compliant drugs with respect to the Abbott corporate compound collection (organics only with molecular weights between 50 and 900) (Figure 3).

Perhaps not surprisingly, given the recent publication of Hajduk *et al.* at Abbott regarding the drugability of protein targets [12], most drugs choose to congregate in a region between rod and 'pancake' shapes with a higher density closer to the rod shape. It appears that most known protein active sites can accommodate molecules of this shape. Sauer originally used the Corina program to calculate MoI derived shapes. However, we chose to double-check our results in case that the computationally derived 'shapes' were an 'artifact' of the Pipeline Pilot program. To this end, we undertook a shape analysis of a subset of Ro5 compliant molecules that partnered proteins in the Protein Data Bank (PDB). It was found that PDB ligands generated from bound X-ray conformations occupied a region in MoI space similar to the MDDR drugs confirming our prior conjectures (Figure 4).

Reassuringly, data derived using energy minimized arbitrary conformations of these same ligands produced virtually identical results. The PDB-derived molecules also, by and large, adopted shapes between rods and pancakes. These results suggest that drug-like molecules should be designed (for current known protein family targets) to have these types of topography. An analysis of the Abbott corporate compound collection indicated a high degree

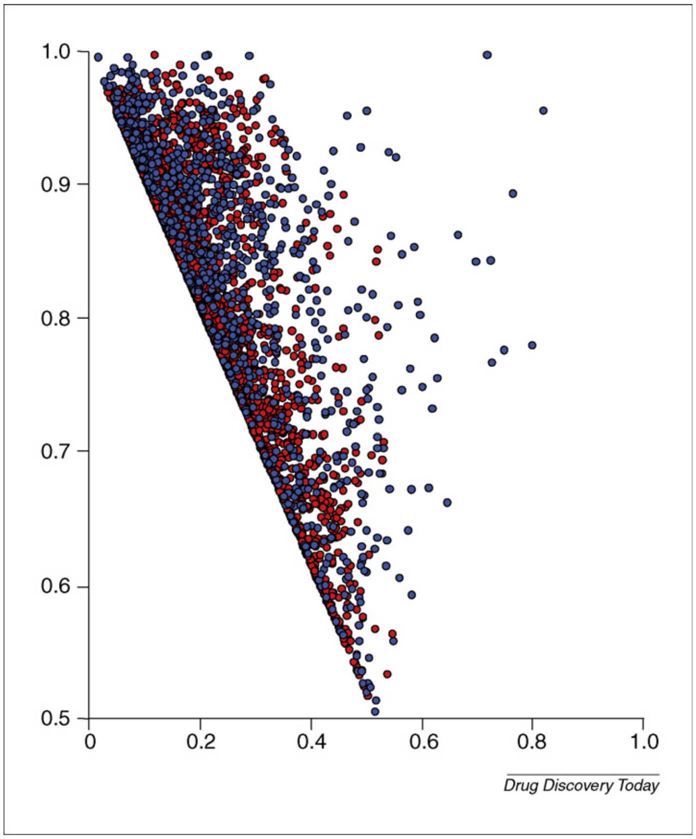

FIGURE 2

Distribution of MoI derived shapes of our kinase initiative compound collection (red) vs. the MDDR database (blue). As expected most kinase inhibitors are located primarily along a pancake/rod axis.

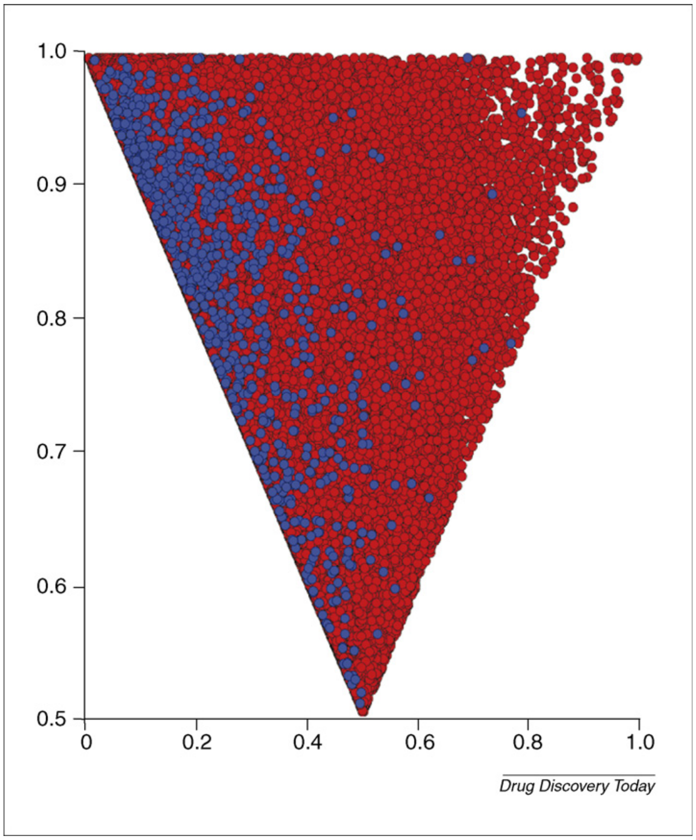

FIGURE 3

Distributions of MoI derived shapes for Ro5 compliant drugs found in the MDDR database (blue) and the Abbott corporate collection (red).

of shape diversity as assessed by the Sauer and Schwartz model, however, a fair proportion of molecules in the collection displayed more 'spherical' shape character. Interestingly, subsequent analysis of the biological records of the compounds indicated that they had never 'hit' in a high throughput screen. These results suggested that the search for compounds of diverse shape should be highly regulated so as to avoid the synthesis of compounds with 'useless' diversity that can rarely, if ever, access biologically relevant space shapes. It is certainly clear that although molecules may meet Ro5 criteria they may not have shapes that allow for productive binding to protein targets. As a final check we asked whether the compounds with calculated 'spherical' shapes could attain desired rod/pancake shape by analyzing a set of conformations for each molecule. It is clear that because of protein–ligand 'induced fit', compounds do not necessarily bind in their lowest energy conformations [13]. Bostrom *et al.* [14] have recently suggested from a study of the conformational energies required for ligands to adopt their bioactive conformations that in 70% of the cases studied the conformational energies of the bioactive conformations were calculated to be $\leq$3 kcal/mol from the global minimum energy conformation. From this analysis, we judge that approximately 5–10% of the Abbott collection probably will not achieve energetically favorable binding to protein targets with a rod/'pancake' shape (>3 kcal/mol required).

If as Chen *et al.* surmise [8], no more than 24% of undiscovered targets will be from protein families not represented by known targets it would appear to make sense to bias library design and compound acquisition efforts towards compounds with rod/'pancake' shapes. This has been our strategy. Although rudimentary, as mentioned previously, we have found this analysis to be useful in the evaluation in the design of compound libraries and for compound acquisition from commercial sources. An analysis of one commercially available collection consisting of small sets of compounds built around several mutually diverse scaffolds is shown in Figure 5.

Figure 5 amply demonstrates the attractive and relevant shape/symmetry diversity exhibited by this collection. In this context, we have found that the degree by which a scaffold influences the shape of the final molecules is closely related to its contribution to the overall MW. For example: the shape of a scaffold that contributes substantially to the MW of the whole molecule also dictates, for the most part, its shape. However, small scaffolds decorated with large substituents provide libraries that are more 'spread out/diffuse' in diversity shape space.

Several libraries (~100 members each) derived from our novel scaffold synthesis initiative centered on multicomponent reaction chemistry have been designed using this shape analysis as a guideline [15–17]. Results are shown in Figure 6. Interestingly, the shape space occupied by the fourth member of this library (shown in black) is more 'diffuse' than the other examples. This, we believe, is because of the significant contribution of the R4 substituent to the overall shape of the library. Nilakantan (Wyeth-Ayerst) [18] and

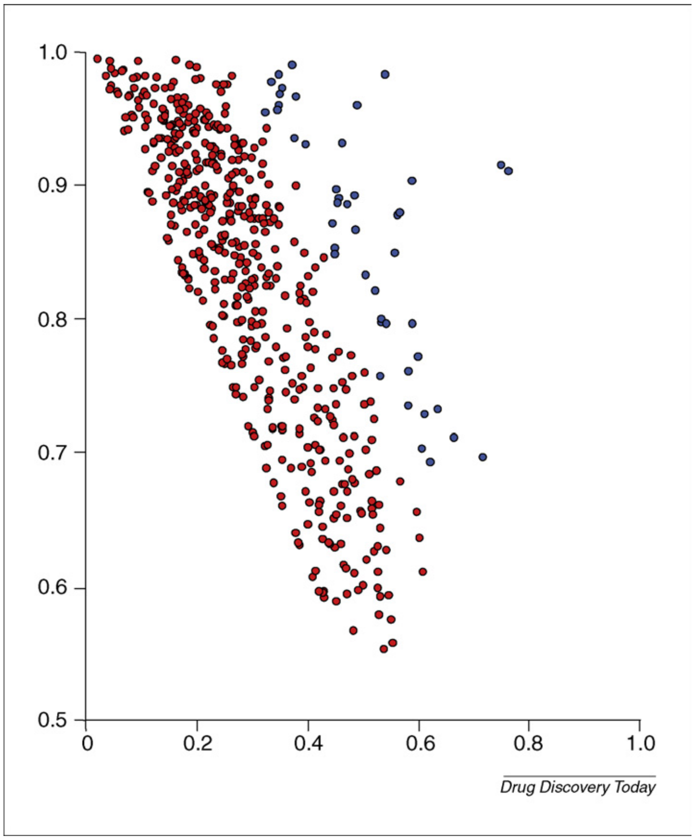

FIGURE 4

Distributions of MoI derived shapes for Ro5 compliant PDB ligands, using their respective bound conformations. 492 'West Coast' (MDDR drug region) compounds in red, 47 'East Coast' (non MDDR drug region) compounds in blue.

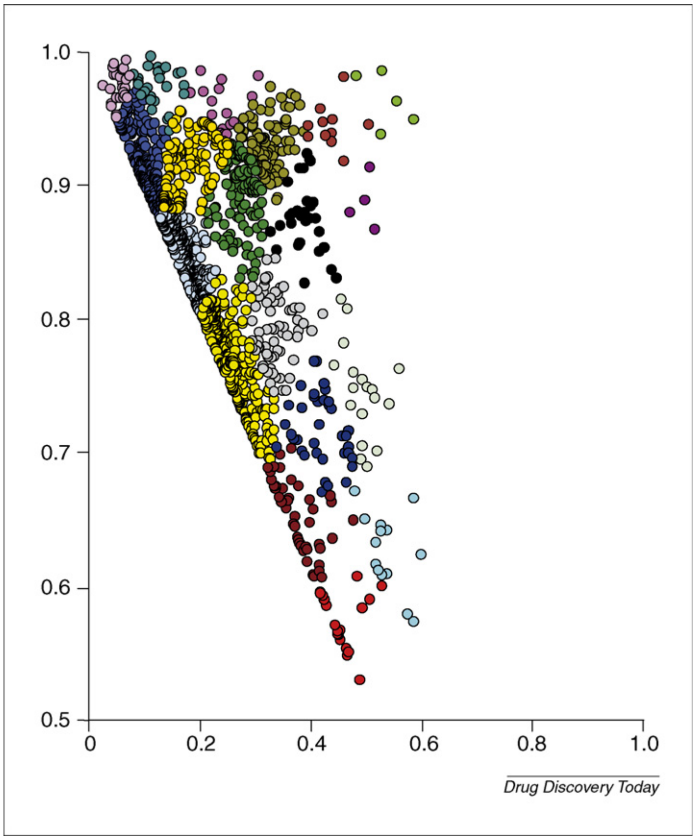

FIGURE 5

Distributions of MoI derived shapes for a Ro4 compliant vendor compound collection.

Harper (GlaxoSmithKline) [19] have constructed statistical models suggesting optimal library designs from clustering and expected HTS hit rates. In accord with ideas proposed by Nilakantan and Harper, we routinely create libraries of ~100 members around selected scaffolds to try to provide a balance of focus and diversity.

Conclusions

To this point, we have approached shape analysis in a forward type of fashion, that is, enumerated virtual libraries of potential molecules have been evaluated to ensure fit into 'west coast' space. A far more challenging task would be the development of tools that provide suggestions for biologically relevant shape space. A thorough analysis of ligands and protein folds [20,21] might provide new insights into the way small molecules fit into proteins and possibly identify new 'privileged shape' chemical structures for library synthesis.

Although these MoI derived shape analyses have proved to be useful for library design they are, of course, an oversimplification in terms of trying to design compounds with correct shape and electronic characteristics to fit known protein active sites. It goes without saying that if electrostatic complementarity is not achieved between ligand and protein productive binding will not occur even if the molecule has the 'right' shape. In a recent publication, McGaughey and coworkers found that the usage of shape alone (ROCS) resulted in a considerably smaller mean enrichment factor for ligand-based virtual screening than did shape and atom-types (ROCS + color) [22]. No publications com-

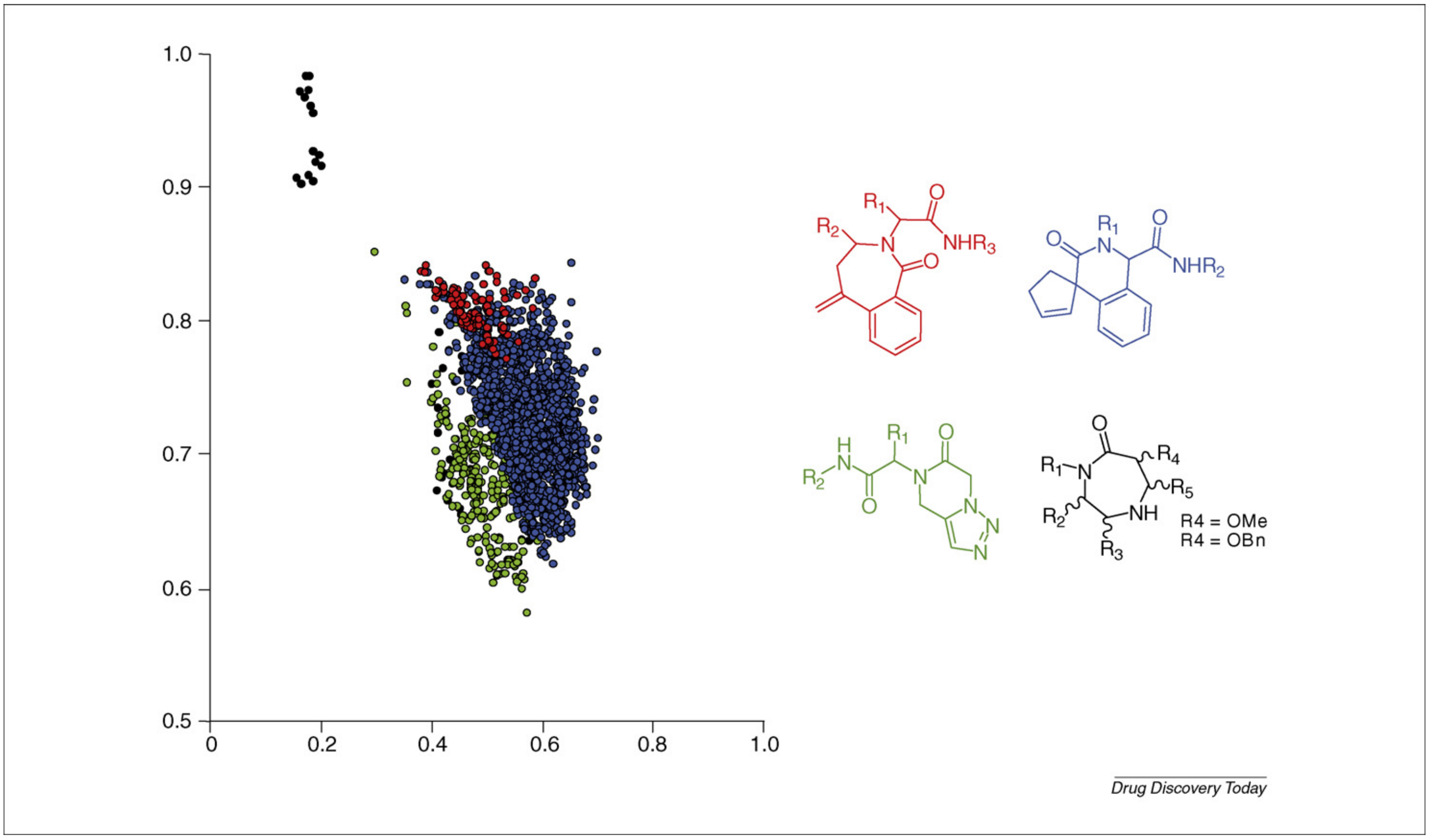

FIGURE 6

Distributions of MoI derived shapes for Ro5 compliant libraries deriving from the corresponding color-coded scaffolds.

paring ROCS, ROCS + color, and MoI with respect to enrichment factors have appeared to-date. Efforts to develop models and ligand classification schemes which incorporate shape—spherical harmonics [23], ROCS [24] and electrostatics, EON [25], PARASURF [26] are underway.

Acknowledgements

The authors would like to thank Drs Bill Greenlee (Schering-Plough), Wolfgang Sauer (Serono) and Jan Schulz (Pyxis) for valuable comments and suggestions during the preparation of the manuscript.

References

1 Koppal, T. (2003) Finding New drugs. *Drug Disc. Dev.* 9, 22–28
2 Jacoby, E. *et al.* (2005) Key aspects of the Novartis Compound Collection Enhancement Project for the compilation of a comprehensive chemogenomics drug discovery screening collection. *Curr. Topics Med. Chem.* 5, 397–411
3 Imming, P. *et al.* (2006) Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Disc.* 5, 821–834
4 Overington, J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Disc.* 5, 993–998
5 Hopkins, A.L. and Groom, C.J. (2002) The druggable genome. *Nat. Rev. Drug Disc.* 1, 727–730
6 Chantry, D. (2003) G protein coupled receptors: from ligand identification to drug targets. *Exp. Opin. Emerg. Drugs* 8, 273–276
7 Zambrowicz, B.P. and Sands, A.T. (2003) Knockouts model the 100 best-selling drugs- will they model the next 100? *Nat. Rev. Drug Disc.* 2, 38–51
8 Zheng, C. *et al.* (2006) Therapeutic targets: Progress of their exploration and investigation of their characteristics. *Pharmacol. Rev.* 58, 259–279
9 Sauer, W.H.B. and Schwarz, M.K. (2003) Molecular shape diversity of combinatorial libraries: A prerequisite for broad bioactivity. *J. Chem. Inf. Comput. Sci.* 43, 987–1003
10 Sauer, W.H.B. and Schwarz, M.K. (2003) Size doesn't matter: Scaffold diversity, shape diversity and biological activity of combinatorial libraries. *Chimia* 57, 276–283
11 Pipeline Pilot, SciTegic, Inc. 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA, http://www.scitegic.com/products_services/pipeline_pilot.htm
12 Hajduk, P.J. *et al.* (2005) Predicting protein druggability. *Drug Discov. Today* 10, 1675–1681
13 Teague, S.J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Disc.* 2, 527–541
14 Bostrom, J. *et al.* (1998) Conformational energy penalties of protein-bound ligands. *J. Comput. Aided Mol. Des.* 12, 383–396
15 Gracias, V. *et al.* (2004) Sequential Ugi/Heck cyclization strategies for the facile construction of highly functionalized N-heterocyclic scaffolds. *Tetrahedron Lett.* 45, 417–420
16 Akritopoulou-Zanze, I. *et al.* (2004) Synthesis of novel fused isoxazoles and isoxazolines by sequential Ugi/INOC reactions. *Tetrahedron Lett.* 45, 3421–3423
17 Vasudevan, A. *et al.* (2004) Synthesis of diazepinones via intramolecular transamidation. *Org. Lett.* 6, 3361–3364
18 Nilakantan, R. *et al.* (2002) A novel approach to combinatorial library design. *Comb. Chem. High Throughput Screen.* 5, 105–110
19 Harper, G. *et al.* (2004) Design of a compound screening collection for use in high throughput screening. *Comb. Chem. High Throughput Screen.* 7, 63–71
20 Breinbauer, R. *et al.* (2002) From protein domains to drug candidates. *Angew. Chem., Int. Ed.* 41, 2878–2890
21 Balamurugan, R. *et al.* (2005) Design of compound libraries based on natural product scaffolds and protein structure similarity clustering (PSSC). *Molecular BioSystems* 1, 36–45
22 McGaughey, G.B. *et al.* (2007) Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* 47, 1504–1519
23 Morris, R.J. *et al.* (2005) Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics* 21, 2347–2355
24 ROCS, version 2.2, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, 2007 http://www.eyesopen.com
25 EON version 1.1, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, 2007 http://www.eyesopen.com
26 PARASURF, Clark, T. *et al.* ParaSurf'06, Universitaet Erlangen-Nuernberg and Cepos InSilico Ltd., Ryde, UK, 2006, http://www.ceposinsilico.de/