ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 439–446

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Methods for compound selection focused on hits and application in drug discovery

Florence L. Stahura, Ling Xue, Jeffrey W. Godden, Jürgen Bajorath*

Computer-Aided Drug Discovery, Albany Molecular Bothell Research Center Inc., 18804 North Creek Pkwy, Bothell, Washington, 98011, USA

Abstract

In the context of virtual screening calculations, a multiple fingerprint-based metric is applied to generate focused compound libraries by database searching. Different fingerprints are used to facilitate a similarity step for database mining, followed by a diversity step to assemble the final library. The method is applied, for example, to build libraries of limited size for hit-to-lead development efforts. In studies designed to inhibit a therapeutically relevant protein–protein interaction, small molecular hits were initially obtained by combined fingerprint- and structure-based virtual screening and used for the design of focused libraries. We review the applied virtual screening approach and report the statistics and results of screening as well as focused library design. While the structures of lead compounds cannot be disclosed, the analysis is thought to provide an example of the interplay of different methods applied in practical lead identification.

Keywords: Virtual screening; Database mining; Docking; Focused libraries; Molecular fingerprints; Protein–protein interaction; Drug discovery

1. Introduction

In computer-aided drug discovery, a variety of approaches are used for lead identification and optimization, dependent on the information available for computational design [1]. If the 3D structures of targets are known, various docking algorithms can be applied to virtually screen compound databases [1,2]. If small molecular hits have been identified, other computational approaches can be used to discover new leads [3]. Although computational design methods are intensely employed in drug discovery programs, the scientific literature is currently dominated by descriptions of methods, rather than examples, which are rarely reported due to the proprietary nature of the compounds identified. Thus, the emphasis is on "how can we do things" and not "what has actually been done". In order to address this situation, the symposium "Designing focused libraries for drug discovery: hit-to-lead to drug", held at the spring 2001 ACS National Meeting in San Diego, organized by C. Reynolds and A. Tropsha, provided a forum for the discussion of practical drug discovery examples and case studies, even if only part of the results could be disclosed. Our contribution presents a case study where different virtual screening approaches have been used to successfully identify new lead compounds.

We first describe some methodological aspects for database mining and focused library design [3,4]. Then we discuss a practical example, computational studies aiming to inhibit an intracellular protein–protein interaction. With the structure of the target known and one small molecular inhibitor identified by wet screening, virtual screening at both the macromolecular level (docking) and small molecular level (using fingerprints) was carried out. The selection of compounds for testing, further aided by binary QSAR calculations [5], is described in Section 2. Of 20 tested compounds, a total of five new synthetic inhibitors were identified using these approaches. Finally, based on one of the inhibitors, a small focused library was generated. Testing of 139 compounds in this library yielded three other inhibitors and four active analogs of the template structure.

Although the structures of these new inhibitors and their binding assay data cannot be disclosed (as they belong to a collaboration partner), the different strategies used and the statistics of virtual screening calculations and library analysis are described in detail to illustrate the computational discovery process.

2. Materials and methods

Fig. 1 summarizes the methods and calculations, as described in the following, and also shows the results

* Corresponding author. Present address: Department of Biological Structure, Albany Molecular Bothell Research Center Inc., University of Washington, Seattle, WA 98195, USA. Tel.: +1-425-424-7297; fax: +1-425-424-7299.
E-mail address: jbajorath@nce-mail.com (J. Bajorath).

1093-3263/02/$ – see front matter
PII: S1093-3263(01)00145-0

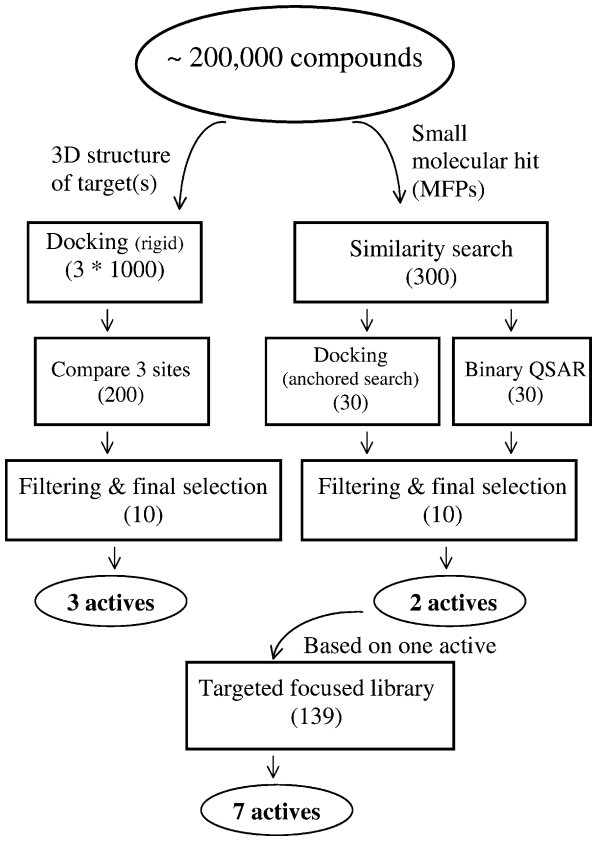

Fig. 1. Computational strategies: the diagram summarizes the methods applied in this study and how they were combined. The number of compounds selected at each stage is given in parentheses.

of the virtual screening analysis that are discussed later on.

2.1. Virtual screening at the macromolecular level

All docking calculations were carried out with DOCK 4.0 [6] and using the publicly available NMR structure of the Bcl-x_L/Bak peptide complex [7] (1BXL.pdb). Partial charges on protein atoms were assigned according to the AMBER force field. Seventy-two residues in the Bcl-x_L structure were selected to define the binding site for docking [8]. A negative image of the active site was constructed by 48 intersecting spheres. The centers of these spheres were used in conjunction with a matching algorithm to identify compounds that "fit" the docking site. Bcl-2 and Bcl-w are related to Bcl-x_L [9]; all are members of the Bcl-2 family and bind similar ligands. Therefore, these targets were also used for docking to further aid in the compound selection process. "Pseudo-Bcl-2" and "pseudo-Bcl-w" binding sites were generated by comparative modeling based on the structure of Bcl-x_L. Of the 72 residues forming the binding site, 33 are rigorously conserved among the three proteins, and three of these residues, Val126, Arg139, and Phe146, are intimately involved in the interaction with Bak [7]. To generate the binding site models based on Bcl-x_L, residues non-conserved in Bcl-2 and Bcl-w were replaced in rotamer conformations and intramolecular contacts re-optimized by minor energy minimization. Negative images of "pseudo-Bcl-2" and "pseudo-Bcl-w" were generated by 51 and 48 intersecting spheres, respectively. Although the accuracy of these binding site models is certainly limited, with more than half of the residues forming Bcl-x_L site not conserved, docking to these alternative receptor sites was thought to increase the chance of identifying novel inhibitors.

For compound screening, a database containing 195,906 ACD compounds were used [10]. Hydrogen atoms and Gasteiger and Marsili partial charges [11] were added to all compounds. For each molecule, a low energy 3D conformation was generated using MOE [12], and the resulting structure was docked using the rigid-body docking algorithm of DOCK [6]. Each ligand was ranked based on shape complementarity (contact scoring only), and the best 1000 molecules were pre-selected from each docking calculation on the Bcl-x_L, Bcl-2, and Bcl-w sites. Approximately 200 compounds common to the three lists were further analyzed within the binding sites. In addition, the compounds were filtered using criteria listed below, in part akin to the Lipinski et al. rules [13]. Thus, compounds chosen should:

- have a molecular weight between 150 and 600;
- have a log P value between -2 and 6;
- not be too flexible (visual inspection only, subjective selection);
- have a polar surface area (PSA[14]) $< 140\,\text{Å}^2$
- show potential for chemical diversification (subjective selection).

2.2. Virtual screening at small molecular level refined by structure-based evaluation

Our algorithms were implemented and used in combination with built-in functions of MOE [12]. A single hit, identified by wet screening of in-house libraries, was used as the template for virtual screening of ACD compounds using two of our mini-fingerprints (MFPs) [15,16] and the Tanimoto coefficient [17] (T_c) as similarity metric. T_c is defined as $T_c = b_c/(b_1 + b_2 - b_c)$, where b_1 is the number of bits set on in molecule 1, b_2 the number of bits set on in molecule 2, and b_c the number of bits common to both molecules. A T_c cut-off value of 0.7 was applied to quantify fingerprint overlap in our calculations. The MFPs used here, termed MFP1 and MFP2, are shown in Fig. 2. They consist of 54 and 62 bit positions, respectively, and monitor ranges of three numerical descriptors (accounting for the number of hydrogen bond acceptors and aromatic bonds as well as the fraction of rotatable bonds in a molecule) and the presence or absence of either 32 or 40 structural fragments or keys. Approximately 300 compounds were found to be "similar" to the hit template using our MFPs and a T_c cut-off value of 0.7. 3D

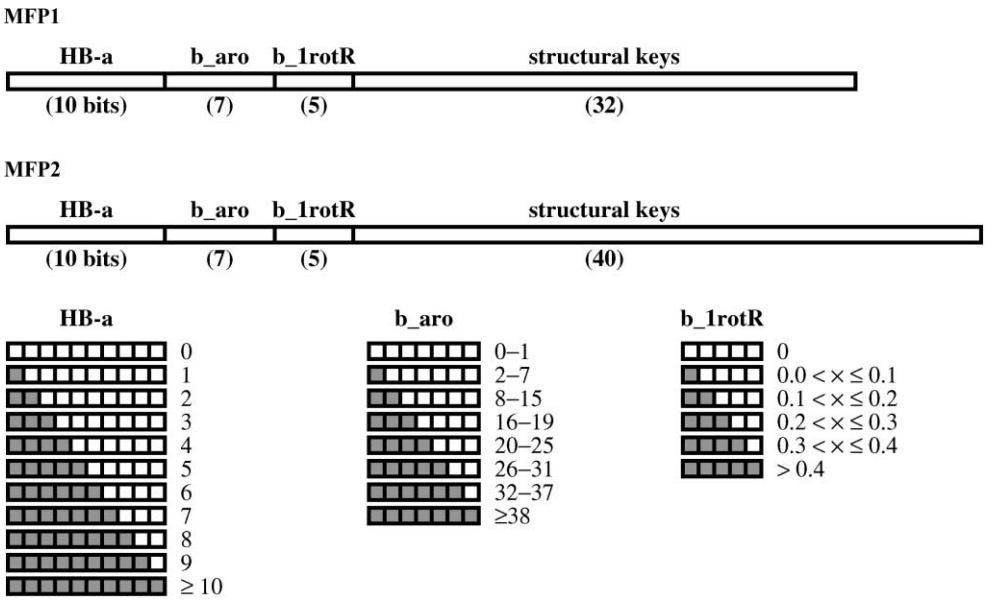

Fig. 2. Mini-fingerprints: two binary MFPs applied in this study are schematically illustrated. They share three numerically encoded descriptors, HB-a (number of hydrogen bonds acceptors in a molecule), b_aro (number of aromatic bonds), and b_1rotR (fraction of rotatable bonds). The bottom part of the figure illustrates how bit segments are used to encode these descriptors. Shaded bit positions are set on (i.e. 1).

structures were generated for these molecules and docking calculations were performed using the Bcl-x_L site and an anchored search protocol [18]. The top 100 molecules were ranked using contact scores and analyzed visually within the docking site. Pre-selected compounds were filtered using criteria listed above and ∼30 compounds were selected.

2.3. Binary QSAR

To further refine the selection of compounds, binary QSAR models were generated and applied. The binary QSAR method and some applications have been described previously [5,19] and are only briefly discussed herein. A learning set of compounds is classified as "active" (i.e. "active" = 1) or "inactive" (i.e. "inactive" = 0), molecular structures and properties are correlated with biological activity using a set of descriptors, and a binary model is generated. Using this model, binary QSAR predicts activity, assigning scores between 0 and 1 for each molecule, based on their calculated descriptor values. Descriptors were chosen by QuaSAR-Contingency [12] calculations. The first binary model, modelA0, was calculated using a training set of 12 molecules, with only five actives and seven inactives, and 83 descriptors (74 2D descriptors and 9 MACCS keys [20]). To further evaluate this model, five test models were generated where, in each case, one of the active compounds was omitted from model derivation. The prediction was considered successful if the calculated activity was >0.5. The three best models were chosen to calculate the predicted activity of the ∼300 compounds obtained by similarity searching. Compounds that had a calculated activity above 0.7 were selected, and the list was reduced to ∼30 compounds.

2.4. Design of a targeted library

To establish a focused compound library, we implemented a metric that makes use of built-in functions of MOE. Instead of building molecules by selecting scaffolds and R-groups, as it is done using QuaSAR-CombiDesign [12], the algorithm generates a library from compound databases using fingerprints. This dual fingerprint-based metric has recently been introduced [4] and we will present here some new features that that have been added. The algorithm includes two steps for selecting molecules from databases. First a "similarity step" is carried out where compounds are selected if they are "similar", based on T_c calculations, to at least one template molecule (i.e. a hit or known inhibitor). The second step is a "diversity" step where molecules are accepted if they are "different" from each other. The T_c of the compound previously accepted in the similarity step is calculated against the rest of the growing library, and the compound is added to the library if its value is below a defined threshold value for all molecules previously selected. Compounds are added to the library until a desired number is reached. The lower this second T_c cut-off value, the more diverse the library will be. MFPs are usually used for the "similarity" step and larger or more complex fingerprints (like MACCS keys [20]) are used for the "diversity" step. Using this approach, a balance between focusing and chemical diversification is achieved by adjusting the parameters. It is also possible to use only one of the steps if no diversity or similarity criteria is desired. New filters were implemented in the algorithm including Lipinski rules and binary QSAR models. Here, three binary QSAR models were used as filters, modelA0, modelB1 and modelC1. Twenty-one

2D descriptors were chosen by QuaSAR-Contingency for modelB1, generated from seven active and 19 inactive compounds. The last model, modelC1, was based on a training set of 190 molecules, only six of which were considered active, and using 32 newly introduced implicit 3D descriptors [21]. For the focused library, 100 compounds were selected from a pool of ACD molecules using MFP2 for the similarity step and filtering by Lipinski rules and three binary models. No diversity step was applied here. As the similarity criterion, a $T_c > 0.8$ was used. During the filtering process, molecules were accepted if at least one of the activity values predicted by the three binary models was above 0.8. In addition, compounds were added to the library based on substructure search using two large fragments of the most potent previously identified inhibitor. Compound distribution in 3D "chemical space" was compared for some of the active compounds following principal component analysis [22] of fingerprint-encoded molecular descriptors. The first three principal components were used as a coordinate system for graphical representation of libraries.

3. Results and discussion

3.1. The target(s)

Our primary goal was to inhibit the Bcl-x_L/Bak protein–protein interaction, for which there were no synthetic leads reported when the project was started. Bcl-x_L and Bak are members of the Bcl-2 family of proteins and are intracellular regulators of apoptosis, downstream of caspases [9]. Their heterodimerization is a key event in program cell death, which when deregulated, contributes to many diseases including cancer and autoimmunity. Some proteins, like Bcl-2 or Bcl-x_L, inhibit apoptosis and others, like Bak or Bax, promote it. Identifying small molecule inhibitors of Bcl-2 family suppressor proteins is thought to selectively promote the death of cancer cells in which Bcl-2 is expressed in excess. Fig. 3A and B show representations of the 16 amino-acid peptide from Bak bound to Bcl-x_L [7]. The interaction is driven by binding of an amphipatic α-helix, formed by the Bak peptide, into the large hydrophobic groove-type of Bcl-x_L. We hypothesized that these features would render the binding site a promising target for finding small molecule inhibitors.

3.2. Compound selection on the basis of virtual screening at the macromolecular level

On the basis of the described docking calculations, only 10 molecules were chosen and acquired for testing, and three synthetic inhibitors were identified having different in vitro specificity for Bcl-2 and Bcl-w. It is interesting to note that these initial results were achieved although only simple parameters and scoring functions (and thus many approximations) were applied. Although DOCK 4.0 is designed to permit flexible docking options (anchored search docking, also used here, being one of them), we initially screened a database containing only one minimized conformation of each compound and used only rigid-body docking; the main

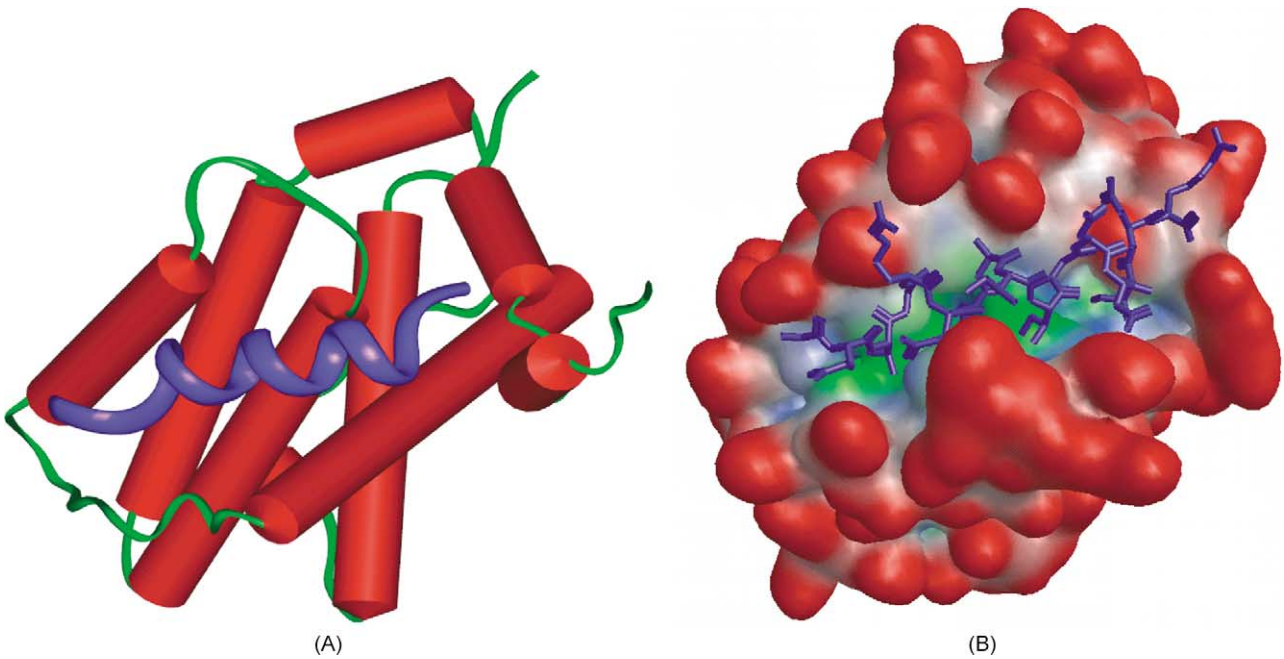

Fig. 3. Protein–protein interaction targeted in this study: the NMR structure of a Bak peptide tightly bound to Bcl-x_L is shown. (A) Ribbon representation of the complex. The Bak peptide forms an α-helix shown in purple. (B) Surface representation of the Bcl-x_L protein, colored by hydrophobic "pockets". The Bak peptide is shown within the binding site in purple, using a stick representation.

reason being computational efficiency. Using this approach, one cannot expect to find active small molecules for which no "binding conformation" was by chance predicted when building the database, which is a major shortcoming. Furthermore, only shape complementarity was taken into consideration and no force field energy-based score was applied for ranking of compounds in initial screens. However, the results show that a high degree of shape or surface complementarity between plausible database conformations of compounds and protein binding sites can readily suffice to identify hits. In addition, to generate related binding sites for compound selection, Bcl-w and Bcl-2 binding were modeled by homology (based on simple replacements of the residues that differ from Bcl-x_L, followed by conformational adjustment). In fact, screening large numbers of databases compounds on multiple binding sites (for example, a protein family) increases the probability of identifying hits statistically, even if simple calculation parameters are applied. On the other hand, the more sophisticated the calculations, the fewer compounds can be evaluated, due to computational constraints. In this case, the increased "accuracy" of more elaborate docking and scoring functions has to compensate for lower statistical probability of identifying hits. Although scoring based on shape complementarity scales with the number of atoms (and has thus a tendency to score larger molecules better than smaller ones), active compounds appeared in the top scoring lists for all three binding sites. However, as discussed above, the use of only single conformation of all database compounds suggests that various other "well-fitting" molecules may not have been found. In general, an important step in the final selection has been the visual analysis of molecules docked into the binding sites (thus adding a subjective or intuitive "scoring" function). However, this required to reduce the number of candidate compounds to a feasible minimum.

3.3. MFP screening with subsequent docking and binary QSAR screens

Virtual screening of ∼200,000 ACD molecules using a small molecule as template and two mini-fingerprints pre-selected only approximately 300 molecules. These compounds were docked into the Bcl-x_L site by anchored search [18] and evaluated by contact scoring. Anchored search docking initially divides compounds into fragments, places a core structure in the binding site and progressively adds the remaining fragments, which is computationally much more costly than rigid-body docking. We thus limited this approach to a relatively small number of pre-selected compounds. In a previous study, we have found that this technique is often more accurate than rigid-body docking [23]. On the basis of these advanced docking calculations, the number of compounds selected by MFP screening was further reduced to ∼30, corresponding to 10% of the pre-selection.

3.4. Binary QSAR predictions

Following initial experimental feedback, another binary QSAR selection was applied to the ∼300 molecules found by MFP screening. The first model used five active compounds (three synthetic compounds found by our docking calculations, the original hit, and one Bcl-2 inhibitor published in the literature while our studies were underway [24]) and seven molecules experimentally confirmed to be inactive. Then, additional models were generated as described in Section 2. In total, six models were tested on the training set. Utilizing an activity threshold value of 0.5, modelA0, modelA4, and modelA5, which were able to identify all five active compounds, were selected for further calculations. ModelA4 and modelA5 predicted an activity of one for the active compound that was omitted to generate the model. In general, binary QSAR models predicted the activity of compounds well, although only a few molecules could be used to generate these models. The fact that active molecules are structurally diverse may explain why accurate predictive models could be generated with such a small training set. ModelA0, modelA4, and modelA5 were used to calculate the predicted activity of the ∼300 compounds from the similarity search and, again, approximately 30 molecules were selected.

3.5. Compound selection by refined similarity search

The two lists from the docking and binary QSAR calculations were compared and 10 compounds present in both lists were selected for testing. Doing so, two new synthetic inhibitors (SS1 and SS2), again distinct from the three ones found previously, were identified.

Compounds	Calculated activity with modelA0	Calculated activity with modelA4	Calculated activity with modelA5	Anchor search docking: ranking by shape complementarity
SS1	0.96	1	1	5
SS2	1	1	1	81

All binary QSAR models predicted high activity for the selected molecules (above 0.95). Interestingly, although these inhibitors score very well by shape complementarity in anchored search docking, they would not have been found by rigid-body docking, since they did not appear in the top 1000 list in any of the calculations. It is noteworthy that MFPs recognized these compounds on the basis of a single

template molecule. An interesting experiment would have been, for example, to test the top 10 scoring compounds identified by similarity search only, without the additional selection criteria. However, this test was not carried out in the course of the collaboration.

3.6. Targeted library design and evaluation

The most potent inhibitor was one of the molecules found by MFP screening (SS2) and was chosen as a single template to design a targeted library. Approximately 2,500 molecules were selected by similarity searching in ACD on the basis of this synthetic inhibitor, using the two MFPs and applying a T_c threshold value of 0.6 (almost 10 times more than when the small molecule hit was used as template with a T_c value of 0.7). One hundred compounds were selected from this compound pool using four filters (Lipinski-like rules and three binary QSAR models). Another 39 molecules were added to this library based on substructure search using two core structure fragments of the inhibitor template. These compounds are best considered as analogs of this inhibitor. In total, 139 compounds were tested and seven additional synthetic inhibitors were identified. Three of these were selected by similarity search only and share a core distinct from the other inhibitors. The four other active compounds were analogs from substructure searching.

3.7. Virtual screening strategies

Why have we selected so few compounds from various screening calculations, only 10 in each case? One goal was to reduce the number of experiments as much as possible, the other to put the selection to a stringent test and see "how well we can do", at least in this case. We also felt encouraged to do so since, from our point of view, the Bcl-x_L/Bak interaction should be a promising target for molecule inhibitor design. The computational approaches applied in the course of this project are summarized in Fig. 1, as mentioned earlier. We were indeed successful in finding a variety of new synthetically accessible inhibitors of this protein–protein interaction, and different strategies provided diverse compounds. These results suggest that no virtual screening approach was a priori superior, and we indeed believe that a meaningful combination of the evaluated approaches was critical

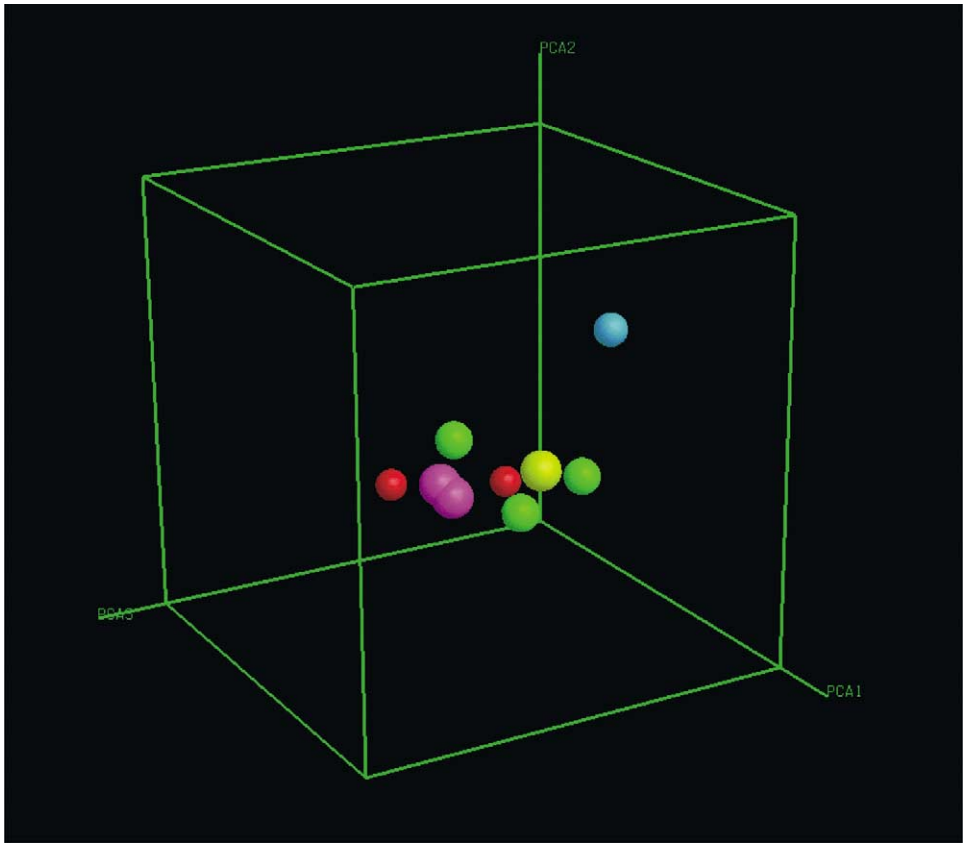

Fig. 4. Analysis of inhibitors in 3D "chemical space". The coordinate system is the result of principal component analysis of 35 MFP-encoded descriptors, calculated for 17 identified inhibitors. Ten of these inhibitors are shown: the original hit used as template for MFP searching (yellow), docked compounds (green), compounds from MFP similarity search (red), compounds from the small focused library (purple), and a literature compound [24] (blue). The other seven are Bcl-x_L inhibitors that have recently been reported in the literature [25], and are not shown for clarity. PC-1, PC-2 and PC-3 are the principal component axes of molecular descriptor space.

for the success with this project, as more information could be exploited for compound selection. Although "rational" criteria were used throughout the selection process, the role of visualization and "chemical intuition" should also not be underestimated. Approximately 200,000 molecules are, according to today's standards, a relatively small pool for compound selection, as virtual libraries in the millions are routinely evaluated. Nevertheless, this comparably small (yet statistically relevant) number of database compounds was sufficient to produce a total of 12 hits. These hits are compared in the "chemical space" representation shown in Fig. 4.

4. Conclusions

Practical drug discovery applications are rarely described in the literature. Because most of these studies are carried out in commercial pharmaceutical settings, successful test cases are seldom disclosed. On the other hand, an increased discussion of successful case studies and pitfalls would certainly advance the field of computational drug design and discovery. In this context, the recent ACS symposium has clearly been a step in the right direction. Although we were not able to disclose specific compounds and experimental data, we could at least explain in some detail the computational approaches and recipes used to produce results. In this study, we have combined several virtual screening strategies to select small sets of compounds for testing. We have implemented conceptually simple 2D tools for biological similarity searching and developed an algorithm to assemble focused libraries by database mining. In addition, we made extensive use of the DOCK approach and binary QSAR analysis. In the example discussed, starting from a single hit and the 3D target structure, a combined 2D and 3D virtual screening approach has yielded diverse synthetic hits.

Additional active compounds were identified from a small focused library, which was assembled from public domain compound sources. The selection of relevant protein targets is a major determinant for the success of small molecule design and discovery projects. Our results make it possible to draw some general conclusions regarding the proteins evaluated in this study. Protein–protein interactions driven by recognition of amphipatic α-helices in cavity-type binding sites represent attractive targets for the discovery of diverse small molecular inhibitors. These types of protein binding sites usually provide a variety of hydrophobic subsites suitable to bind organic molecules and shield them from the solvent environment. Therefore, in these cases, identified inhibitors may often have different modes of action. In contrast, "difficult" targets, for example, protein–protein interactions driven by specific binding of larger and relatively flat surfaces (such as seen, for example, in numerous protein–antibody complexes) may not permit to identify such molecules, even if large numbers of compounds are screened.

References

[1] H. Kubinyi, Combinatorial and computational approaches in structure-based drug design, Curr. Opin. Drug Discov. Develop. 1 (1998) 16–27.
[2] P.J. Gane, P.M. Dean, Recent advances in structure-based rational drug design, Curr. Opin. Struct. Biol. 10 (2000) 401–404.
[3] J. Bajorath, Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening, J. Chem. Inf. Comput. Sci. 41 (2001) 233–245.
[4] L. Xue, J. Godden, F.L. Stahura, J. Bajorath, A dual fingerprint based metric for the design of combinatorial libraries and analogs, J. Mol. Mod. 7 (2001) 125–131.
[5] P. Labute, Binary QSAR: a new method for the determination of quantitative structure activity relationships, Pac. Symp. Biocomput. 7 (1999) 444–455.
[6] E.C. Meng, D.A. Gschwend, J.M. Blaney, I.D. Kuntz, Orientation sampling and rigid body minimization in molecular docking, Proteins 17 (1993) 266–278.
[7] M. Sattler, H. Liang, D. Nettesheim, R.P. Meadows, J.E. Harlan, M. Eberstadt, H.S. Yoon, S.B. Shuker, B.S. Chang, A.J. Minn, C.B. Thompson, S.W. Fesik, Structure of Bcl-x_L-Bak peptide complex: recognition between regulators of apoptosis, Science 275 (1997) 983–986.
[8] S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, An all atom force field for simulation of proteins and nucleic acids, J. Comp. Chem. 7 (1986) 230–252.
[9] J.M. Adams, S. Cory, The Bcl-2 protein family: arbiters of cell survival, Science 281 (1998) 1322–1326.
[10] Available Chemicals Database (ACD), MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577.
[11] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity: rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3288.
[12] Molecular Operating Environment (MOE), version 1999.05, Chemical Computing Group Inc., 1255 University Street, Montreal, Que., Canada, H3B 3X3.
[13] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Freeney, Experimental and computational approaches to estimate solubility and permeability is drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25.
[14] D.E. Clark, Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. Part 1. Prediction of intestinal absorption, J. Pharm. Sci. 88 (1999) 807–814.
[15] L. Xue, J. Godden, J. Bajorath, Database searching for compounds with similar biological activity using short binary bit string representations of molecules, J. Chem. Inf. Comput. Sci. 39 (1999) 881–886.
[16] L. Xue, J. Godden, J. Bajorath, Evaluation of descriptors and mini-fingerprints for the identification of molecules with similar activity, J. Chem. Inf. Comput. Sci. 40 (2000) 1227–1234.
[17] P. Willett, Chemical similarity searching, J. Chem. Inf. Comput. Sci. 38 (1998) 983–996.
[18] A.R. Leach, I.D. Kuntz, Conformational analysis of flexible ligands in macromolecular receptor sites, J. Comp. Chem. 13 (1992) 730–748.
[19] F.L. Stahura, J. Godden, L. Xue, J. Bajorath, Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations, J. Chem. Inf. Comput. Sci. 40 (2000) 1245–1252.
[20] MACCS structural keys, MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577.
[21] P. Labute, A widely applicable set of descriptors, J. Mol. Graph. Model. 18 (2000) 464–477.
[22] W.G. Glen, W.J. Dunn, Principal component analysis and partial least squares regression, Tetrahedron Comp. Methodol. 2 (1989) 349–376.

[23] J. Godden, F. Stahura, J. Bajorath, Evaluation of docking strategies for virtual screening of compound databases: cAMP-dependent serine/threonine kinase as an example, J. Mol. Graph. Model. 16 (1998) 139–143.
[24] J.-L. Wang, D. Liu, Z.-J. Zhang, S. Shan, X. Han, S.M. Srinivasula, C.M. Crose, E.S. Alnemri, Z. Huang, Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 7124–7129.
[25] A. Degterev, A. Lugovskoy, M. Cardone, B. Mulley, G. Wagner, T. Mitchison, J. Yuan, Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-x_L, Nature Cell Biol. 3 (2001) 173–182.