Knowledge-Driven Lead Discovery

Bernard Pirard*

Aventis Pharma Deutschland GmbH, A Company of the Sanofi-Aventis Group, Chemical Sciences, Drug Design, Building G878, D-65926 Frankfurt am Main, Germany

Abstract: Virtual screening encompasses several computational approaches which have proven valuable for identifying novel leads. These approaches rely on available information. Herein, we review recent successful applications of virtual screening. The extension of virtual screening methodologies to target families is also briefly discussed.

Keywords: Virtual screening, chemical biology, chemogenomics, lead discovery, similarity searching, pharmacophore, data mining, docking.

1. INTRODUCTION

The pharmaceutical industry is under pressure to increase its research productivity while reducing both the spiralling costs of research and the time to bring a compound to the market. In the early '90s, combinatorial chemistry and high throughput screening (HTS) were regarded as key technologies to enable the synthesis and the biological testing of large diverse chemical libraries and hopefully to increase the research productivity. Although the application of combinatorial chemistry and HTS led to the discovery of several lead compounds, among them Gleevec^{TM}[1], the number of new chemical entities brought to the market remained flat and even decreased over the last couple of years [2]. In the meantime, the pharmaceutical industry has become increasingly aware of the HTS limitations, in particular its low hit rate (typically below 0.1% for single dose HTS) and its large number of false positives [2-4]. Many experts have blamed the poor quality of the screening libraries for the low hit rate and the large number of false positives [5]. As a result, lead discovery has moved towards more knowledge-driven approaches like virtual screening (VS).

Moreover, the wealth of potential therapeutic targets provided by genomic initiatives [6] makes knowledge-driven lead discovery strategies almost essential, since both the costs of screening and the efforts required for data management grow with the number of assays [7]. Most of these proteins belong to large families like the kinases, the G protein-coupled receptors (GPCRs), the proteases, etc [8]. Obviously, one way to gain efficiency is to exploit knowledge and know-how available on well-characterised members of a given target family. For the pharmaceutical industry, this entails a move from a single target/disease to a cross-therapeutic areas target family-based paradigm, often referred to as chemical biology or chemogenomics [5,7,9].

Herein, we review recent applications of VS and extension of VS methodologies to target families. Virtual screening encompasses a variety of computational approaches which aim to reduce a large collection of compounds to a short list of screening candidates by applying a sequence of filters. This sequence depends on the amount of information available within a given discovery project. Biochemical screening of the molecules passing these filters should yield a higher hit rate, compared to screening of a random selection. While nearly all of the recent reviews on VS [10-13] have focused on methodological aspects, we would like to stress the significance of successful applications. Whenever possible, we discuss the chemical novelty of the compounds identified by VS, their suitability for future chemical optimisation, their potency and selectivity, compared with already known ligands, as well as their pharmacokinetic (PK) profile. In the final part of this review, we outline the extension of established VS methodologies to target families.

*Address correspondence to this author at the Aventis Pharma Deutschland GmbH, A Company of the Sanofi-Aventis Group, Chemical Sciences, Drug Design, Building G878, D-65926 Frankfurt am Main, Germany; Tel: + 49 (0) 69 305 83 768; Fax: + 49 (0) 69 33 13 99; E-mail: bernard.pirard@sanofi-aventis.com

2. VIRTUAL SCREENING

Typically, a VS cascade consists of fast and general filters followed by problem-specific and more time consuming ones such as docking of compounds in a protein binding site.

2.1. General Filters

These filters aim to discard or to flag molecules deemed unsuitable for biochemical screening [14], for instance covalent-acting drugs and 'frequent hitters' [15] – compounds that show up as positives in assays covering diverse targets.

The strategies for identifying such compounds involve substructure searches for toxic or reactive groups (such as sulfonyl chloride) and can also include limits on several features or properties like the molecular weight or the formal charge. Roche and colleagues [16] developed a model based on a neural network using atom types as descriptors and HTS data as well as known drugs as a training set, to identify potential 'frequent hitters'. In a validation study, this model was able to correctly classify 90% of the test set.

Other groups have also developed *in-silico* filtering or flagging tools that try to predict whether a compound possesses or does not possess the features of a drug (drug-likeness). These methods rely on the comparison of a set of

1389-5575/05 $50.00+.00

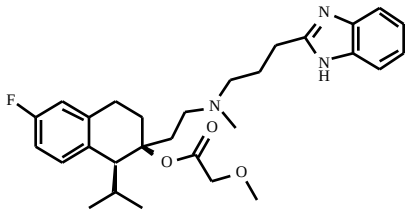
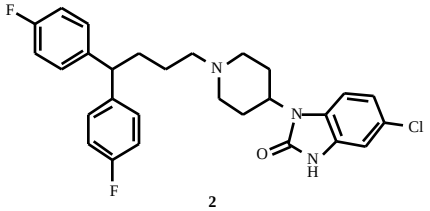
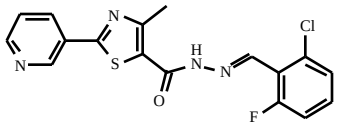
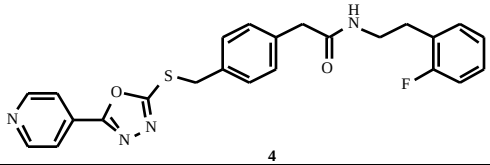
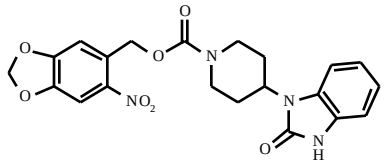
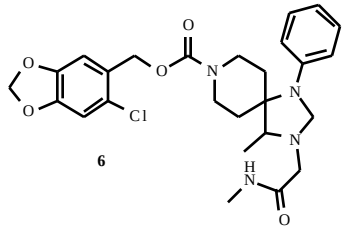
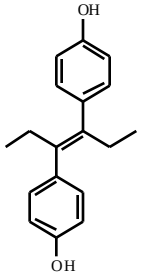
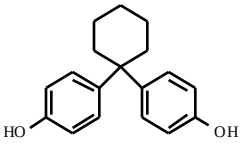
known drugs with compounds known or presumed to be non-drug in order to extract discriminating properties or characteristics between both sets. Since these approaches have been extensively reviewed over the last couple of years [17-19], we only mention that they can be grouped in three categories:

- functional group filters, which discard molecules not possessing functional groups found in drugs [20];
- property filters limiting the selection to those molecules whose calculated physical properties fall within a specified range;
- classification tools like principal component analysis (PCA) [21], decision trees and neural networks.

Basically, VS is used in lead discovery programs which aim to identify molecules amenable to chemical optimisation for turning them into clinical candidates. Therefore, the consideration of drug-likeness criteria early in a VS cascade is questionable. Comparison of property distributions in databases containing compounds at different stages of their life cycle (lead identification, optimisation, clinical trials, marketed drugs) indicates that lead-like structures are on average less complex (lower molecular weight, less rings and rotatable bonds), less lipophilic and more water soluble, compared with their corresponding drugs, as well as with compounds in development [2,22]. Using a simple model of ligand receptor interactions, Hann and colleagues [23] also supported the inclusion of less complex molecules in libraries used for lead discovery purposes. Charifson and Walters warned against the increased risk of false negatives resulting from the use of too strict filters early in a VS campaign [24]. Instead, one could consider some of the drug-likeness filters (like binary classification schemes for cytochrome P450 isoform inhibition liabilities [25]) in the final stage of screening set selection, or even as a flagging routine, since issues like P450 inhibition can still be tackled in the lead optimisation phase.

Once these general filters have been applied, one can consider structure-based or ligand-based VS strategies, depending on whether the three-dimensional (3D) structure of the target protein is available or not. The aim of these approaches is to identify molecules which are desirable for *in vitro* screening.

Table 1. Successful Applications of Similarity Searching

Target	Query	Molecular representation	Representative hit
T-type Ca^{2+} channel	 **1**	topological pharmacophores	 **2**
GSK3	 **3**	topological pharmacophores	 **4**
μ receptor	 **5**	3D fingerprints	 **6**
ERα	 **7**	molecular fields	 **8**

2.2. Ligand-Based Virtual Screening

Despite the continuous growth in protein 3D structural data, ligand structures remain the predominant source of knowledge in many projects. Depending on the amount of information available, three different types of ligand-based VS approaches can be considered [26]. In the simplest type, one knows just one ligand (natural substrate, compound from a patent) which can be used as a query for substructure or similarity searches. In the second case, several actives are available, one can identify a pharmacophore and subsequently use it for 3D database searches. In the third case, statistical tools are used to extract information from vast amounts of noisy data (like HTS data).

2.2.1. Similarity Searches

Similarity searches aim to identify database molecules similar to one or several actives (query). The underlying assumption is that structurally similar molecules exhibit similar biological activity which has been validated by several researchers [27-29]. So wide is the variety of similarity searches approaches developed over the last decades that we refer the reader to more comprehensive reviews [26,30,31]. All these methods share the same conceptual framework consisting of two independent components: a molecular representation and a similarity index. Obviously, the effectiveness of any method depends on the project, the query and the content of the database [26,31]. Moreover, different methods produce different ranking of active compounds and hence tend to select different subsets of compounds. In particular, 3D descriptors like pharmacophore keys or molecular fields can detect similarities that are missed by two-dimensional (2D) descriptors such as Daylight fingerprints [26,31,32]. Therefore, several researchers have suggested that using several similarity approaches and combining their results might lead to superior performance [31].

We now move on to review some successful applications of similarity searching in drug discovery projects. Table **1** describes a synopsis of each case. Schneider and colleagues have developed a similarity searching technique referred to as Chemically Advanced Template Search (CATS) [33]. The CATS approach is based on the comparison of topological pharmacophoric features assigned to atom pairs. Using the T-type Ca^{2+} channel blocker mibefradil (**1**) as query for a CATS search, Schneider selected the 12 highest-ranking molecules for screening in a cellular assay. As a result, nine of them (75%) showed significant activity (IC_{50} < 10 μM), the best of them was clopimozid (**2**) with a novel scaffold and activity in the submicromolar range [33]. Recently, another group reported a successful application of a VS strategy based on a method derived from the CATS to identify a novel class of micromolar inhibitors of glycogen synthase kinase-3 (GSK3) [34]. Starting from HTS hits not amenable to further optimization (**3**), the CATS method was able to identify a structural class (**4**) suitable for single array solution phase synthesis whereas similarity searches on 2D fingerprints did not result in the identification of any structure similar to the queries. A HTS hit with micromolar activity on the μ opiate receptor (**5**) was also considered as the starting point for similarity searches on 3D pharmacophoric keys, which led to the discovery of two novel chemical classes of μ opiate receptor ligands [35]. The most active compound (**6**) exhibits submicromolar affinity on the μ opiate receptor, a different selectivity profile and improved water solubility, compared to the query. Mestres performed flexible superpositions of commercially available compounds, based on electrostatic and steric fields, on the bioactive conformation of diethylstilbestrol (**7**) bound to the estrogen receptor α (Erα), which resulted in the discovery of a novel agonist with micromolar activity (**8**) [36]. A method based on steric field comparison of single rule-generated conformers has also been successfully applied in hit follow-up exercises, but few structural and biological details have been disclosed so far [37].

2.2.2. Pharmacophore Modeling and 3D Searching

A pharmacophore model, defined as the spatial arrangement of structural and physicochemical features relevant for biological activity, can be used in conjunction with 3D database searching for identifying novel leads, optimizing them and focusing combinatorial libraries [38,39]. In this subsection, we focus on pharmacophore models identified from a set of ligands which are supposed to bind to the same site of the target. Pharmacophores derived from protein binding sites are discussed in the section 2.3 on structure-based VS.

Over the past 15 years or so, software products for pharmacophore modeling and 3D searching have been developed [38,39]. Some of these software products have been evaluated retrospectively on data sets from the medicinal chemistry literature in order to investigate their efficiency in the area of VS [40,41]. Other scientists have performed a validation study on the ability of three commercially available pharmacophore modeling programs to reproduce pharmacophores, identified from crystal structures of protein ligand complexes [42]. As a result, key elements for the successful use of pharmacophore models in VS have been identified as: the training set selection (i.e. compounds used to identify a pharmacophore), the choice of feasible pharmacophoric features and the handling of conformational flexibility.

Many succesful applications of pharmacophore-based VS have been reported [39]. Herein, we review just a few of the most recent success stories. Several researchers have constructed pharmacophoric queries based on X-ray or nuclear magnetic resonance (NMR) structures of peptides as well as on structure-activity relationships (SAR) obtained for peptidic compounds [43,44]. These queries are used to search databases for nonpeptidic replacements. The identification of nonpeptidic lead compounds is essential in drug discovery in order to avoid the pharmacokinetic problems inherent to peptides, such as their metabolic instability and their low bioavailability. Recently, scientists at Aventis employed the solution structure of the vasoactive cyclic peptide urotensin II (U-II) and SAR of its analogues to set up a pharmacophore query for screening of the corporate database [43]. Biological screening of 500 compounds matching the pharmacophore revealed 10 antagonists (hit rate 2%) from six different chemical classes, with IC_{50} values ranging from 400 nM to 7 μM. This hit rate is about 20 times higher than seen in HTS screens for GPCR antagonists. A similar improvement in hit rate has also been reported by researchers from another pharmaceutical company who compared the performances of

pharmacophore-based VS and HTS for the discovery of farnesyl transferase inhibitors [45]. Pharmacophore-based database searches also served to identify a stable inhibitor of the serine protease chymase [46]. Starting from a training set of 26 chymase inhibitors unstable in human plasma, Koide and colleagues derived a four-point pharmacophore model. After validation on an external test set of 60 inhibitors, this model was used to search a database of 216599 compounds; 45 of the retrieved molecules were selected for biological screening. This resulted in the identification of three actives from three different chemical classes, and one of them was stable. In another recent work, pharmacophore-based VS led to the discovery of a novel chemical class of non-competitive glutaminergic antagonists with equivalent or increased potency compared to the training set members [47]. Successful applications of this VS strategy also include the discovery of novel antagonists or inhibitors for several therapeutically relevant proteins, like the muscarinic M3 receptor [48], the cytochrome 17 [49], and a monoamine transporter [50]. One advantage of pharmacophore-based VS is that it provides a molecular alignment of the hits with the training set members, which can be used to design novel compounds [47] or to derive 3D quantitative structure-activity relationship (QSAR) models.

2.2.3. Classification Methods

The awareness of HTS limitations has led to the emergence of the sequential screening paradigm recently reviewed by Bajorath [13,51]. Sequential screening is an iterative process, which combines VS for subset selection and HTS. The development of novel SAR and statistical methods which are able to handle large and heterogeneous data sets, has contributed to the emergence of this new screening paradigm.

Recursive partioning (RP) is one of these novel statistical techniques [52]. Briefly, RP identifies the most statistically relevant chemical feature that can split a data set of initially screened compounds into smaller and more homogeneous subsets, correlating chemical features with biological activities. A RP analysis produces a dendrogram, or SAR tree, which provides insight into complex SAR as well as a predictive model for selecting compounds for a second round of biological screening. Applying both RP and the sequential screening paradigm to GPCRs [53] and to kinases [54] yielded significant improvements in hit rate, compared to other compound selection methods. However, the authors of these studies didn't provide any structural information on the actives identified.

Binary QSAR is another recent statistical technique, which might prove valuable in a sequential screening setting [55]. In binary QSAR, a training set of compounds is classified as "active" or "inactive", a binary model is generated, which correlates a set of descriptors with the activity classes. This model can be used subsequently to predict the activity class of an external set of compounds. In a study designed to identify inhibitors of protein-protein interactions, binary QSAR was used to refine the hit list of a similarity search [56]. As a result, ten of 30 hits from the similarity search were selected for biological screening; two of them were active (hit rate of 20%).

2.3. Structure-Based Virtual Screening

Progress in molecular biology, protein chemistry and structural biology has led to an exponential growth in the number of 3D protein structures. These provide a sound basis for lead finding through VS.

A structure-based VS campaign usually starts with an analysis of all the relevant 3D protein structures. As a result, one identifies the key intermolecular interactions that need to be formed by a candidate ligand. Then, chemical databases are searched for candidate ligands by means of docking or pharmacophore matching. Before this more computer-intensive step, it may be worth applying some of the general and ligand-based filters described above. Once the database has been searched, the hit list has to be further analyzed in order to remove compounds with unreasonable binding modes or strained conformations. Finally, a short list of molecules is selected for biological testing. Actives identified in the primary assay can be further characterized in a secondary assay or crystallized with the target protein for confirmation of the binding mode predicted by docking. Herein, we review recent applications of both high throughput docking and protein-based pharmacophore searching.

2.3.1. High Throughput Docking

Docking of large compound collections requires an efficient search algorithm as well as a scoring function able to discriminate between binders and non-binders. We refer the reader to some recent reviews dealing with the methodological aspects of docking [57,58] and scoring [59].

Several researchers have also assessed the performance of docking and scoring tools on different proteins by compiling a test set including a few known actives and many drug-like molecules, which are assumed to be non-binders. This data set is subsequently docked in each of the test proteins and the performance of a given docking /scoring combination is based on its ability to enrich known ligands among the top-scoring docked molecules [60-64]. Some of these studies [60b,63,64] suggest that the performance of docking is strongly affected by the conformation of the receptor used for the calculations. However, the assumption that all the drug-like molecules present in the test set are inactive on the proteins under study is questionable. Furthermore, retrieval from a large data set of active compounds, already optimized for binding to the investigated protein, is significantly easier than finding novel and weaker ligands in a large collection of non-binders. Therefore, enrichment factors derived from these validation studies may not be very conclusive.

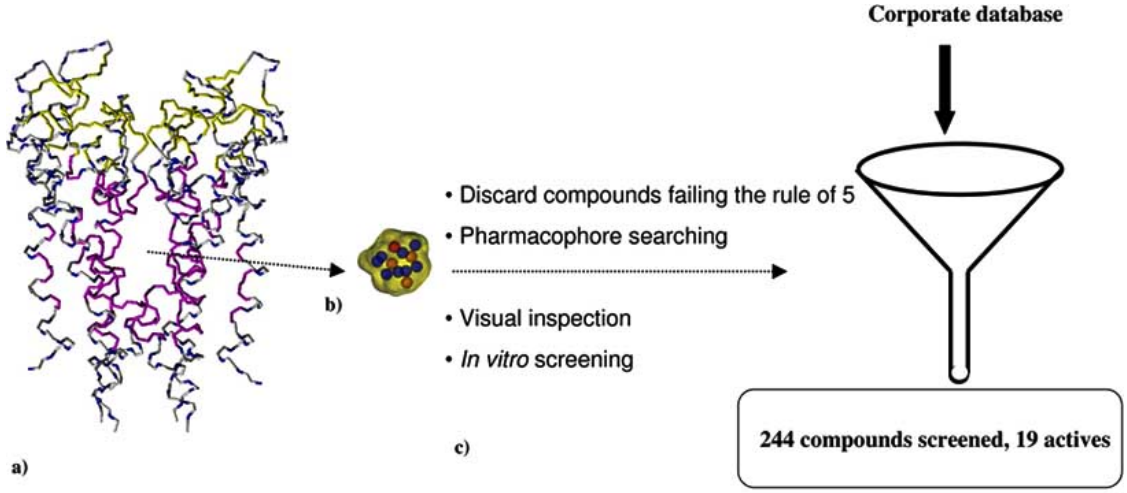
Over the last couple of years, several successful applications of high throughput docking for lead finding have been reported [65]. A protocol comprising general filters followed by more specific ligand-based filters was applied for reducing the size of two commercial databases before docking in carbonic anhydrase [66]. After visual inspection of the docking results for the top-scoring compounds, 13 compounds were selected for biological testing. Eleven of them turned out to be active, with four in the nanomolar range. These nanomolar activities are exceptional since most hits from VS exhibit activities in the micromolar range. However, one can argue that these carbonic anhydrase inhibitors are not truly novel, since they

show the same Zn binding groups as previously known inhibitors of the same enzyme. More importantly, the binding modes of two of these known inhibitors were confirmed by crystal structure determination. Crystal structure determination is an invaluable step in the follow-up of actives from VS, because protein plasticity can lead to a binding mode not predicted by docking [67,68].

An unexpected binding mode resulting from protein plasticity can provide new opportunities for structure-based VS, as illustrated by Brenk [68]. Brenk considered an unexpected binding mode observed for an inhibitor of tRNA-guanine transglycolase as a starting point for a VS cascade, similar to the one applied to carbonic anhydrase. As a result, all of the nine compounds selected for biological testing showed activities in the micromolar range, with two of them in the submicromolar range. Recently, another group described the successful application of a hierarchy of filters involving 3D pharmacophore searches followed by docking in an homology model of an important target for antimalarial chemotherapy [69]. This strategy resulted in the identification of 12 actives out of 24 selected for testing based on the docking results. Interestingly, these inhibitors, structurally unrelated to compounds used in the clinic, were found active at the micromolecular level on both the wild type enzyme and its mutants resistant to known chemotherapy.

Other scientists have skipped the general and ligand-based filters and only relied on docking and scoring. Wyss and colleagues [70] docked about 9000 compounds from a virtual library in the crystal structure of a bacterial dihydrofolate reductase (DHFR). They selected for synthesis 252 compounds among the top-scoring ones, and 54 of them were found active (21%). Interestingly, they also applied a diversity-based selection approach to the same pool of 9000 products and obtained a higher hit rate (21% compared to 3% for diversity-based selection) by means of structure-based VS. Other researchers reported similar observations on other targets [71,72]. However, as illustrated by Doman [72], both approaches should be considered as complementary techniques for lead discovery. Docking in a homology model followed by a hierarchical sequence of post-processing filters also served to identify ATP competitive inhibitors of casein kinase II, an enzyme with no previously known drug-like inhibitor [73]. As a result, 12 compounds were selected for screening and four were found active. The binding mode, proposed for the most active one, was supported by structure-activity relationships obtained in a subsequent hit follow-up program. Other successful applications of docking in a protein model include the discovery of thyroid hormone receptor antagonists [74], CDK4 [75] and BCL2 [76] inhibitors.

Comparisons of the properties of leads and drugs have led to the conclusion that a low molecular weight compound might be a suitable starting point for a lead discovery program [22,23]. This observation serves as rationale for fragment-based screening, also referred to as needle screening, using either biophysical or computational methods. Virtual needle screening has been successfully applied to identify low molecular weight inhibitors of two targets for anti-infectives, namely DNA gyrase [77] and inosine 5'-monophosphate dehydrogenase (IMPDH) [78]. In both projects, lead identification efforts relying on HTS or



	1	2	3	4	5
1	1.00				
2	0.28	1.00			
3	0.29	0.51	1.00		
4	0.25	0.25	0.26	1.00	
5	0.32	0.54	0.43	0.28	1.00

d)

Fig. (1). a) Homology model of the human Kv1.5 pore, with the amino acids forming the external binding site colored in dark blue and those forming the internal site in magenta. **b)** protein-derived pharmacophore for the internal site. The H bond acceptor, donor and hydrophobic sites are colored respectively in red, blue and orange. The volume of the protein cavbity is displayed in yellow. **c)** VS protocol. **d)** Similarity matrix on Unity fingerprints computed for the five compounds with $IC_{50} < 10$ μM.

traditional medicinal chemistry did not provide any suitable compound. The work on DNA gyrase inhibitors, which combined structure-based pharmacophore searches and docking, is noteworthy because the authors took care to validate screening hits using several biophysical methods. In addition, they showed how structure-based design could be used to turn a weak binder into a potent lead candidate. In the more recent paper on IMPDH, Pickett and colleagues applied a docking procedure for reducing the size of the list of screening candidates by a factor of 50; a hit rate of 10% was recorded.

2.3.2 Protein-Based Pharmacophore Searching

Pharmacophores derived from the analysis of protein binding sites can be used before docking or at the bottom of a VS cascade. Computational tools to characterize proteins binding sites have been recently reviewed [79]. Earlier successful applications of protein-based pharmacophore searching include the discovery of thyroid hormone receptor ligands [80] and the identification of HIV integrase inhibitors [81].

Recently, Wu and colleagues derived a pharmacophore from the analysis of crystal structures of complexes between cyclophilin and peptidic ligands [82]. They subsequently used this model as query to screen databases of commercial compounds and found a micromolar non-peptidic ligand. This ligand was considered a suitable starting point for a hit exploration program. Using the crystal structure of a bacterial potassium channel as template [83], we constructed a homology model of the human voltage-dependent potassium channel Kv1.5, a target for which no HTS assay was available (Fig. (**1**)). Based on geometrical criteria [84], we identified two putative ligand binding sites and selected the internal one for further characterization with the GRID force field [85]. A pharmacophore was derived from this analysis and subsequently used as query to screen our corporate databases. As a result, 244 compounds were selected for *in vitro* screening; 19 of them showed activity (hit rate 7.8%). Among these actives, five compounds belonging to four different chemical classes had IC_{50} < 10 μM (Fig. (**1**))♣.

3. CHEMICAL BIOLOGY

All the approaches presented so far focus on single targets. However, target family-based approaches are emerging, as we move into the post-genomic era [9].

Within this context, interest in developing new methods for protein classification has grown. Earlier approaches to group proteins into families are based on comparisons either of their overall sequence or of their tertiary structure [86]. With respect to ligand discovery, focusing the comparison on smaller protein regions that encompass putative ligand binding sites seems most relevant. The strategy to localize the sequence homology to the different ligand fragment binding sites has been recently applied to monoamine-related GPCRs [87]. Other researchers have considered descriptors of protein ligand interactions for detecting similarities among protein binding sites [86, 88-92].

One of the most promising aspects of protein binding sites comparisons is to suggest structural motifs or features for chemical libraries to be screened on orphan proteins, related to well-characterized targets. In particular, privileged substructures have been identified for GPCRs ligands [93]. Classification schemes to recognize molecules acting at members of specific target families like kinases, GPCRs and serine proteases have also been developed [94-96]. These models, which are based on 2D descriptors and neural networks, can achieve over 80% correct classifications. Other researchers have extended the scope of similarity searches for identifying ligands binding to targets that belong to the same family as the target of the reference ligand [97]. So far, only retrospective validation studies of these concepts have been carried out while successful applications to lead discovery have not been reported.

4. CONCLUSIONS

Over the last couple of years, VS has emerged as a complementary alternative to HTS for lead finding. Depending on the information available, different computational approaches can be applied for reducing a large database to a short list of screening candidates. Knowledge of only one active compound is sufficient to identify novel leads for a given target. When more information is available, more sophisticated and time-consuming approaches like pharmacophore searching or docking can be used. Results produced by these latter approaches can provide a starting point for subsequent hit exploration or even lead optimization. Despite several successful applications of VS, lead discovery would still benefit from further methodological developments. In particular, one still needs accurate and fast scoring functions for protein ligand interactions.

In the post-genomic era, the challenge is to select the most druggable targets and find the corresponding drug-like molecules. In order to focus the search space, researchers have suggested exploiting the available information and know-how, gained from well-characterized related targets. Hence, grouping proteins and designing libraries focused on target families are becoming key activities. Established VS methodologies have been extended to deal with target families. However, despite the interest in target family-based approaches, their real benefits have not yet been demonstrated.

LIST OF ABBREVIATIONS

CATS = Chemically Advanced Template Search

DHFR = DiHydroFolate Reductase

Erα = Estrogen receptor α

GPCRs = G Protein-Coupled Receptors

GSK3 = Glycogen Synthase Kinase-3

HTS = High Throughput Screening

IMPDH = Inosine 5'-MonoPhosphate Dehydrogenase

NMR = Nuclear Magnetic Resonance

PCA = Principal Component Analysis

♣B. Pirard, Poster, Gordon Research Conference "Ligand Recognition and Molecular Gating", 2002

PK = Pharmacokinetic

QSAR = Quantitative Structure-Activity Relationships

RP = Recursive Partioning

SAR = Structure-Activity Relationships

3D = Three-Dimensional

2D = Two-Dimensional

U-II = Urotensin-II

VS = Virtual Screening

REFERENCES

[1] Auld, D. S.; Diller, D.; Ho, K.-K. *Drug Discov. Today*, **2002**, *7*, 1206-1213.

[2] Oprea, T. I. *J. Comput.-Aided Mol. Des.*, **2002**, *16*, 325-334.

[3] Lahana, R. *Drug Discov. Today*, **1999**, *4*, 447-448.

[4] Ramesha, C. *Drug Discov. Today*, **2000**, *5*, 43-44.

[5] Wess, G.; Urmann, M.; Sickenberger, B. *Angew. Chem. Int. Ed.*, **2001**, *40*, 3341-3349.

[6] Drews, J. *Science*, **2000**, *287*, 1960-1964.

[7] Bleicher, K. H. *Curr. Med. Chem.*, **2002**, *9*, 2077-2084.

[8] Hopkins, A. L; Groom, C. R. *Nat. Rev. Drug Discov.*, **2002**, *1*, 727-730.

[9] Caron, P. R.; Mullican, M. D.; Mashal, R. D.; Wilson, K. P.; Su, M. S; Murcko, M. A. *Curr. Opin. Chem. Biol*., **2001**, *5*, 464-470.

[10] Finn, P. W. *Drug Discov. Today*, **1996**, *1*, 363-370.

[11] Walters, W. P.; Stahl, M. T.; Murcko, M. A. *Drug Discov. Today*, **1998**, *3*, 160-178.

[12] Klebe, G. Virtual Screening. An Alternative or Complement to High Throughput Screening, Kluwer: Dordrecht, **2000**.

[13] Bajorath, J. *Nat. Rev. Drug Discov.*, **2002**, *1*, 882-894.

[14] Rishton, G. M. *Drug Discov. Today*, **2003**, *8*, 86-96.

[15] Mc Govern, S L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. *J. Med. Chem.*, **2002**, *45*, 1712-1722.

[16] Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E.-M.; Rogers-Evans, M.; Neidhart, W.; Stadler, H.; Dillon, M.; Sjögren, E.; Foutouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G. *J. Med. Chem.*, **2002**, *45*, 137-142.

[17] Clark, D. E.; Pickett, S. D. *Drug Discov. Today*, **2000**, *5*, 49-58.

[18] Matter, H.; Baringhaus, K.-H., Naumann, T.; Klabunde, T.; Pirard, B. *Comb. Chem. High Throughput Screen.*, **2001**, *4*, 453-475.

[19] Clark, D. E. *Comb. Chem. High Throughput Screen.*, **2001**, *4*, 477-496.

[20] Muegge, I.; Heald, S. L.; Brittelli, D. *J. Med. Chem.*, **2001**, *44*, 1841-1846.

[21] a) Oprea, T. I.; Gottfries, J. *J. Comb. Chem.*, **2001**, *3*, 157-166. b) Oprea, T. I.; Zamora, I.; Ungell, A.-L. *J. Comb. Chem.*, **2002**, *4*, 258-266.

[22] a) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. I. *Angew. Chem. Int. Ed.*, **1999**, *111*, 3962-3967. b) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1308-1315.

[23] Hann, M. M.; Leach, A. R.; Harper, G. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 856-864.

[24] Charifson, P. S.; Walters, W. P. *J. Comput.-Aided Mol. Des.*, **2002**, *16*, 311-323.

[25] Zuegge, J.; Fechner, U.; Roche, O.; Parrott, N. J.; Engkvist, O.; Schneider, G. *Quant. Struct.-Act. Relat.*, **2002**, *21*, 249-256.

[26] Bajorath, J. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 233-245.

[27] Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. *J. Med. Chem.*, **1996**, *39*, 3049-3059.

[28] a) Matter, H. *J. Med. Chem.*, **1997**, *40*, 1219-1229. b) Pötter, T.; Matter, H. *J. Med. Chem.*, **1998**, *41*, 478-488.

[29] Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. *J. Med. Chem.*, **2002**, *45*, 4350-4358.

[30] Willett, P.; Barnard, J. M.; Downs, G. M. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 983-996.

[31] Sheridan, R. P.; Kearsley, S. K. *Drug Discov. Today*, **2002**, *7*, 903-911.

[32] Schuffenhauer, A.; Gillet, V. J.; Willett, P. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 295- 307.

[33] Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem. Int. Ed.*, **1999**, *38*, 2894-2896.

[34] Naerum, L.; Nørskov-Lauritsen, L.; Oelsen, P. H. *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 1525-1528.

[35] Poulain, R.; Horvath, D.; Bonnet, B.; Eckhoff, C.; Chapelain, B.; Bodinier, M.-C.; Deprez, B. *J. Med. Chem*., **2001**, *44*, 3378-3390.

[36] Mestres, J. *Biochemical Society Transactions*, **2002**, *30*, 797-799.

[37] a) Cramer, R. D.; Poss, M. A.; Hermsmeier, M. A.; Caulfield, T. J.; Kowala, M. C.; Valentine, M. T. *J. Med. Chem.*, **1999**, *42*, 3919-3933. b) Andrew, K. M.; Cramer, R. D. *J. Med. Chem.*, **2000**, *43*, 1723-1740. c) Cramer, R. D.; Jilek, R. J.; Andrews, K. M. *J. Mol. Graph. Model.*, **2002**, *20*, 447-462.

[38] Güner, O. F. *Pharmacophore Perception, Development, and Use in Drug Design*, IUL Biotechnology Series, La Jolla, **2000**.

[39] Kurogi, Y.; Güner, O. F. *Curr. Med. Chem.*, **2001**, *8*, 1035-1055.

[40] Hecker, E. A.; Duraiswami, C.; Andrea, T. A.; Diller, D. J. *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 1204-1211.

[41] Krovat, E. M.; Langer, T. *J. Med. Chem*., **2003**, *46*, 716-726.

[42] Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. *J. Comput.-Aided Mol. Design*, **2002**, *16*, 653-681.

[43] Flohr, S.; Kurz, M.; Kostenis, E.; Brkovich, A.; Fournier, A.; Klabunde, T. *J. Med. Chem*., **2002**, *45*, 1799-1805.

[44] Singh, J.; van Vlijmen, H.; Liao, Y.; Lee, W.-C.; Cornebise, M.; Harris, M.; Shu, I; Gill, A.; Cuervo, J. H.; Abraham, W. M.; Adams, S. P. *J. Med. Chem*., **2002**, *45*, 2988-2993.

[45] Kaminski, J. J.; Rane, D. F.; Rothofsky, M. L. In *Pharmacophore Perception, Development, and Use in Drug Design*, IUL Biotechnology Series, La Jolla, **2000**; pp. 251-268.

[46] Koide, Y.; Tatsui, A.; Hasegawa, T.; Murakami, A.; Satoh, S.; Yamada, H.; Kazayama, S.-I.; Takahashi, A. *Bioorg. Med. Chem. Lett*., **2003**, *13*, 25-29.

[47] Barreca, M. L.; Gitto, R.; Quartarone, S.; De Luca, L.; De Sarro, G.; Chimirri, A. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 651-655.

[48] Marriott, D. P.; Dougall, I. G.; Meghani, P.; Liu, Y.-J.; Flower, D. R. *J. Med. Chem*., **1999**, *42*, 3210-3216.

[49] Clement, O. O.; Freeman, C. M.; Hartmann, R. W.; Handratta, V. D.; Vasaitis, T. S.; Brodie, A. M. H.; Njar, V. C. O. *J. Med. Chem.*, **2003**, *46*, 2345-2351.

[50] Enyedy, I. J.; Wang, J.; Zaman, W. A.; Johnson, K. M.; Wang, S. *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 1775-1778.

[51] Stahura, F. L.; Bajorath, J. *Curr. Med. Chem.*, **2003**, *10*, 707-15.

[52] a) Young, S. S.; Hawkins, D. M. *J. Med. Chem.*, **1995**, *38*, 2784-2788. b) Hawkins, D. M.; Young, S. S.; Rusinko, A. *Quant. Struct.-Act. Relat.*, **1997**, *16*, 296-302. c) Young, S. S.; Hawkins, D. M. *SAR QSAR in Environ. Res.*, **1998**, *8*, 183-193. d) Rusinko, A. III; Farmen, M. W.; Lambert, C. G.; Brown, P. L.; Young, S. S. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1017-1026.

[53] Jones-Hertzog, D. K.; Mukhodpadhyay, P.; Keefer, C. E.; Young, S. S. *J. Pharmacol. Toxicol.*, **1999**, *42*, 207-215.

[54] Barker, E. J.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Morris, J. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 346-356.

[55] Labute, P. *Pac. Symp. Biocomput.*, **1999**, *7*, 444-455.

[56] Stahura, F. L.; Xue, L.; Godden, J. W.; Bajorath, J. *J. Mol. Graph. Model.*, **2002**, *20*, 439-446.

[57] Muegge, I.; Rarey, M. *In Rev. Comput. Chem.*; K. B. Lipkowitz, D. B. Boyd, Eds.; VCH: New York, **2001**; Vol. *17*, pp.1-60.

[58] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput.-Aided Mol. Des.*, **2002**, *16*, 151-166.

[59] Gohlke, H.; Klebe, G. *Angew. Chem. Int. Ed.*, **2002**, *41*, 2644-2676.

[60] a) Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem*., **2000**, *43*, 4759-4767. b) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. *Proteins: Struct., Funct. Genet.*, **2003**, *50*, 5-25

[61] Stahl, M.; Rarey, M. *J. Med. Chem*., **2001**, *44*, 1035-1042.

[62] Schulz-Gasch, T.; Stahl, M. *J. Mol. Model*, **2003**, *9*, 47-57.

[63] Mc Govern, S. L.; Shoichet, B. K. *J. Med. Chem.*, **2003**, *46*, 2895-2907.

[64] Schapira, M.; Abagyan, R.; Totrov, M. *J. Med. Chem*., **2003**, *46*, 3045-3059.

[65] Shoichet, B. K.; Mc Govern, S. L.; Wei, B.; Irwin, J. J. *Curr. Opin. Chem. Biol*., **2002**, *6*, 439-446.

[66] a) Grüneberg, S.; Wendt, B.; Klebe, G. *Angew. Chem. Int. Ed.*, **2001**, *40*, 389-393. b) Grüneberg, S.; Stubbs, M. T.; Klebe, G. *J. Med. Chem.*, **2002**, *45*, 3588-3602.
[67] Fritz, T. A.; Tondi, D.; Finer-Moore, J. S.; Costi, M. P.; Stroud, R. M. *Chem. Biol.*, **2001**, *8*, 981-995.
[68] Brenk, R.; Naerum, L.; Grädler, U.; Gerber, H.-D.; Garcia, G. A.; Reuter, K.; Stubbs, M. T.; Klebe, G. *J. Med. Chem.*, **2003**, *46*, 1133-1143.
[69] Rastelli, G.; Pacchioni, S.; Sirawaraporn, W.; Sirawaraporn, R.; Parenti, M. D.; Ferrari, A. M. *J. Med. Chem.*, **2003**, *46*, 2834-2845.
[70] Wyss, P. C.; Gerber, P.; Hartman, P. G.; Hubschwerlen, C.; Locher, H.; Marty, H.-P.; Stahl, M. *J. Med. Chem.*, **2003**, *46*, 2304, 2312.
[71] Paiva, A. M.; Vanderwall, D. E.; Blanchard, J. S.; Kozarich, J. W.; Williamson, J. M.; Kelly, T. M. *Biochim. Biophys. Acta*, **2001**, *1545*, 67-77.
[72] Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. *J. Med. Chem.*, **2002**, *45*, 2213-2221.
[73] Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabro, D.; Furet, P. *J. Med. Chem.*, **2003**, *46*, 2656-2662.
[74] Schapira, M.; Raaka, B. M.; Das, S.; Fan, L.; Totrov, M.; Zhou, Z.; Wilson, S. R.; Abagyan, R.; Samuels, H. H. *Proc. Natl. Acad. Sci. U. S. A.*, **2003**, *100*, 7354-7359
[75] Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T.; Fukasawa, K.; Iwama, T.; Ikeura, C.; Ikuta, M.; Suzuki-Takahashi, I.; Iwasawa, Y.; Hayama, T.; Nishimura, S.; Morishima, H. *J. Med. Chem.*, **2001**, *44*, 4615-4627.
[76] Enyedy, I. J.; Ling, Y.; Nacro, K.; Tomita, Y.; Wu, X.; Cao, Y.; Guo, R.; Li, B.; Zhu, X.; Huang, Y.; Long, Y.-Q.; Roller, P. P.; Yang, D.; Wang, S. *J. Med. Chem.*, **2001**, *44*, 4313-4324.
[77] Boehm, H.-J.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbers, T.; Meunier-Keller, N.; Mueller, F. *J. Med. Chem.*, **2000**, *43*, 2664-2674.
[78] Pickett, S. D.; Sherborne, B. S.; Wilkinson, T.; Bennett, J.; Borkakoti, N.; Broadhurst, M.; Hurst, D.; Kilford, I.; McKinnell, M.; Jones, P. S. *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 1691-1694.
[79] Sotriffer, C.; Klebe, G. *Farmaco.*, **2002**, *57*, 243-251.
[80] Greenidge, P. A.; Carlsson, B.; Bladh, L.- G.; Gillner, M. *J. Med. Chem.*, **1998**, *41*, 2503-2512.
[81] Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. *J. Med. Chem.*, **2000**, *43*, 2100-2114.
[82] Wu, Y.-Q.; Belyakov, S.; Choi, C.; Limburg, D.; Thomas IV, B. E.; Vaal, M.; Wei, L.; Wilkinson, D. E.; Holmes, A.; Fuller, M.; McCormick, J.; Connolly, M.; Moeller, T.; Steiner, J.; Hamilton, G. S. *J. Med. Chem.*, **2003**, *46*, 1112-1115.
[83] Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science*, **1998**, *280*, 69-77.
[84] Brady, G. P. Jr.; Stouten, P. F. W. *J. Comput.-Aided Mol. Des.*, **2000**, *14*, 383-401.
[85] Goodford, P. J. *J. Med. Chem.*, **1985**, *28*, 849-857.
[86] Schmitt, S.; Kuhn, D.; Klebe, G. *J. Mol. Biol.*, **2002**, *323*, 387-406.
[87] Jacoby, E. *Quant. Struct.-Act. Relat.*, **2001**, *20*, 115-123.
[88] Stahl, M.; Bur, D.; Schneider, G. *J. Comput. Chem.*, **1999**, *20*, 336-347.
[89] Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E. J.; Fox, T. *J. Med. Chem.*, **2000**, *43*, 3033-3044.
[90] Sheridan, R. P.; Holloway, M. K.; McGaughey, G.; Mosley, R. T.; Singh, S. B. *J. Mol. Graph. Model.*, **2002**, *20*, 217-225.
[91] Capello, V.; Tramontano, A.; Koch, U. *Proteins: Struct., Funct. Genet.*, **2002**, *47*, 106-115.
[92] Greenbaum, D. G.; Arnold, W. D.; Lu, F.; Hayrapetian, L.; Baruch, A.; Krumrine, J.; Toba, S.; Chehade, K.; Brömme, D.; Kuntz, I. D.; Bogyo, M. *Chem. Biol.*, **2002**, *9*, 1085-1094.
[93] Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. *J. Med. Chem.*, **1999**, *42*, 3251-3264.
[94] Manallack, D. T.; Pitt, W. R.; Gancia, E.; Montana, J. G.; Livingstone, D. J.; Ford, M. G.; Whitley, D. C. *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 1252-1262.
[95] Balakin, K. V.; Tkachenko, S. E.; Lang, S. A.; Okum, I.; Ivashchenko, A. A.; Savchuk, N. P. *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 1332-1342.
[96] Lang, S. A.; Kozyukov, A. V.; Balakin, K. V.; Skorenko, A. V.; Ivashchenko, A. A.; Savchuk, N. P. *J. Comput.-Aided Mol. Des.*, **2002**, *16*, 803-807.
[97] a) Schuffenhauer, A.; Zimmermann, J.; Stoop, R.; van der Vyver, J.-J.; Lecchini, S.; Jacoby, E. *J. Chem. Inf. Comput. Sci.*, **2002**, 42, 947-955. b) Schuffenhauer, A.; Floersheim, P.; Acklin, P.; Jacoby, E. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 391-405.