Reviews • FOUNDATION REVIEW

Foundation review:
Virtual ligand screening: strategies, perspectives and limitations

Gerhard Klebe

Institute of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, D35032 Marburg, Germany

In contrast to high-throughput screening, in virtual ligand screening (VS), compounds are selected using computer programs to predict their binding to a target receptor. A key prerequisite is knowledge about the spatial and energetic criteria responsible for protein–ligand binding. The concepts and prerequisites to perform VS are summarized here, and explanations are sought for the enduring limitations of the technology. Target selection, analysis and preparation are discussed, as well as considerations about the compilation of candidate ligand libraries. The tools and strategies of a VS campaign, and the accuracy of scoring and ranking of the results, are also considered.

GERHARD KLEBE
Gerhard Klebe is a full professor of pharmaceutical chemistry at the University of Marburg, Germany, since 1996. The focus of his research is the understanding of protein–ligand interactions, including chemical synthesis, microcalorimetry, molecular biology, crystallography, bioinformatics and software development. He studied chemistry at the University of Frankfurt, Germany, and obtained his doctorate in physical chemistry. After postdoctoral positions in crystallography (Grenoble, CNRS and ILL, France, and the University of Berne, Switzerland) he was responsible for molecular modelling and crystallography at BASF-AG in Ludwigshafen, Germany. He recently turned down an offer from ETH Zürich, Switzerland, for a chair in pharmaceutical chemistry. He has published ~200 scientific articles, serves on the editorial board of several journals and is member of the board of governors of the Cambridge Crystallographic Data Centre.

Introduction

In the late 1980s and early 1990s, experimental high-throughput screening (HTS) and combinatorial chemistry were aggressively developed to overcome the lead discovery bottleneck in drug development. Using sophisticated large-scale automation, it was anticipated that this would generate an unprecedented number of novel leads, resulting in a substantial increase in novel drug entities launched to the market per year. However, in reality, the opposite was the case [1,2]. Frequently, the discovered hits could not be validated and further optimized into actual leads and preclinical candidates. Thus, the initial euphoria surrounding these approaches has subsided owing to the disappointingly low hit rates and significant costs involved [3,4].

Such situations fuel the consideration and development of alternative techniques. The expression 'virtual screening' (VS) was coined in the late 1990s; however, the techniques involved are much older. In an effort to show that searching for lead candidates using a computer is a serious alternative to HTS, the term 'VS' was adopted by the community. In contrast to HTS, which is largely phenomenological and technology driven, in VS, compounds are selected by predicting their binding to a macromolecular target using computer programs (in drug discovery, the term 'target' or 'receptor' is used frequently to describe the macromolecule to which a drug binds, which is usually a protein but can also be DNA or RNA). The compounds studied do not necessarily exist, and their 'testing' does not consume valuable substance material. Experimental deficiencies, such as limited solubility, aggregate formation or any sort of influence that could possibly interfere with experimentally applied assay conditions do not need to be considered in the initial computational screen. In contrast to HTS, VS requires as a key prerequisite knowledge about the spatial and energetic criteria responsible for the binding of a particular candidate ligand

Corresponding author: Klebe, G. (klebe@mailer.uni-marburg.de)

to the receptor under investigation. In consequence, either the three-dimensional (3D) structure of the macromolecular target – as given by crystal structure analyses, NMR or sophisticated homology modelling – or, at the very least, a rigid reference ligand with a known bioactive conformation mapping out the putative receptor binding site must be available [5]. This defines VS as a knowledge-driven approach. Even though it is likely that multiple screening campaigns have been performed in parallel in industry [6–9], it was only recently that McMaster University launched an open and unbiased competitive screening against *Escherichia coli* dihydrofolate reductase to detect inhibitors [10], to assess how well VS can enrich candidate ligands compared with HTS random screening [10,11]. The original sample set of compounds used for screening was split into two, and one fraction was tested by HTS. Obtained hits were reported and served as a training set to validate and tailor applied VS tools in different research groups, who entered into a competition to retrieve hits also detected by the experimental screen in the second portion of the screening sample [12–16]. Subsequently, in a totally unbiased fashion, the second part of the data sample was evaluated by HTS and VS in parallel. Unexpectedly, the second portion did not show any hits as competitive inhibitors, as would be expected for ligands docked into the substrate binding site of a target protein. Interestingly, Brenk *et al.* [16] reported on some promising hits found by docking that were not detected as inhibitors in the initial HTS. Retesting at higher concentration indicated weak inhibition. This observation could be seen as supporting the view that VS and HTS are complementary approaches and that they might find candidates missed by the other [7,9].

As mentioned above, the methods involved in VS are much older than the approach itself. Initial attempts to find ligands by docking or by mapping them onto ligand-based pharmacophore models were the generic prototypes of a VS approach. However, this term was not used at that time, probably because computers and algorithms were not fast enough to enable large scale applications. Focusing on approaches that actually make use of an available protein structure, one of the first systems to be studied by docking was HIV protease. Initial versions of the program DOCK, developed over many years by Kuntz's group [17,18], tried to dock rigid entries from the Cambridge Crystallographic Database into the protein receptor, focusing primarily on shape complementarity and later considering chemical complementarity. In 1990, the Kuntz group retrieved the neuroleptic drug haloperidol as a potential 'lead' from a docking screen using a database of known drug molecules as input. However, this compound would have had to be administered at a very high dose – far beyond a toxicologically tolerable concentration – to be an effective inhibitor of the protease [19]. Nevertheless, haloperidol as a 'lead' gave rise to some new ideas for developing a derivative possessing 15 μM inhibition of HIV protease [20]. Later, at Dupont-Merck, a 3D database search retrieved a substituted terphenyl derivative as a putative lead for inhibiting this protease. Further optimization via six- and seven-membered rings resulted in the class of cyclic ureas that are able to replace the crucial structural water molecule in the protease, at the same time targeting the carboxy groups of the two catalytic aspartates via two appropriately placed hydroxy functionalities [21]. Since these early VS attempts, a plethora of case studies has been performed and the list of success stories is steadily growing. Despite the fact that VS is still a young discipline, it has been reviewed frequently [22–30], most recently in a comprehensive overview by Kubinyi [31]. If one excludes purely retrospective studies, in which the potential of a method is demonstrated by its ability to enrich putatively active molecules from a sample of anticipated nonactive ones, ∼50 targets have been studied to date, and reports on the discovery of mostly micromolar binding ligands in a truly predictive fashion are available (Table 1). Still under development, and far from mature, the number of strategies followed in VS is nearly as large as the number of reported screen-

TABLE 1

Targets addressed by virtual screening [31]

G-protein-coupled receptors	α1A adrenoceptor, dopamine D3 receptor, endothelin A (ETA) receptor, melanin-concentrating hormone type 1 receptor, muscarinic M3 receptor, neurokinin-1 receptor, neuropeptide Y receptor type 5, purinergic A2A receptor, urotensin II receptor (GPR14)
Nuclear receptors	Retinoic acid receptor, thyroid hormone receptor
Kinases	Akt 1 (also known as protein kinase Bα), Bcr-abl tyrosine kinase, checkpoint kinase 1, cyclin-dependent kinase 2, cyclin-dependent kinase 4, glycogen synthetase kinase, p56 lymphoid T cell tyrosine kinase, protein kinase CK2 (also known as casein kinase II), transforming growth factor β receptor kinase
Proteases	Cathepsin D, falcipain-2, HIV protease, plasmepsin II, severe acute respiratory syndrome CoV 3C-like proteinase, thrombin
Other hydrolases	Acetylcholinesterase, adenylate cyclase (oedema factor and CyaA, a toxin of the pathogenic bacteria *Bacillus anthracis* and *Bacillus pertussis*), AmpC β-lactamase, phosphodiesterase 4, protein tyrosine phosphatase 1B
Oxidases and reductases	Aldose reductase, dihydrofolate reductase, inosine 5′-monophosphate dehydrogenase inhibitors
Miscellaneous enzymes	5-Aminoimidazole-4-carboxamide ribonucleotide transformylase, carbonic anhydrase II, DNA gyrase, dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase, farnesyl transferase, guanine phosphoribosyl transferase, HIV-1 integrase, tRNA-guanine transglycosylase
Ion channels	T-type selective Ca²⁺ channel, Kv1.5 potassium channel, shaker potassium channel
Protein–protein Interfaces, protein complexes	Bcl-2 protein–protein interaction, cyclophilin A, FK506-binding protein, mesangial cell proliferation, Rac1 protein–protein interaction, VLA-4 (also known as α4β1 antigen)
Protein–RNA interactions	HIV-1 RNA transactivation response element

ing campaigns. Owing to space constraints, this review is unable to give a comprehensive overview of success stories or investigated targets, or to review all details of currently applied VS protocols; accordingly, the reader is recommended to consult the above-mentioned surveys [22–31]. Instead, this review attempts to summarize and comment on the agreed concepts and prerequisites for performing successful VS runs. Furthermore, it attempts to analyse the limitations of the approach in a frank and plain manner and seeks to explain why such limitations still exist. Accordingly, in the following sections, first the target selection, analysis and preparation are discussed, followed by considerations about the compilation of a candidate ligand screening sample. This is followed by a discussion of the actual requirements, tools and strategies of a VS campaign, and, finally, remarks concerning the accuracy of the scoring and ranking of the screening results.

Target selection, validation and analysis
Being a knowledge-driven approach, the scope of VS strongly depends on the amount and quality of information available about the system under investigation. Clearly, the availability of the target receptor is of great benefit compared with situations where only a rigid reference ligand is known. The target receptor could be any macromolecular biomolecule, either a protein, RNA or DNA. In terms of the rigid reference ligand, either a substrate, inhibitor molecule, agonist or antagonist (e.g. a steroid hormone) could serve. Because the structures of target receptors are becoming increasingly available, this review concentrates solely on such situations. However, it does not mean that VS cannot be applied to situations in which no structural information about the target receptor is available [5]. With respect to DNA and RNA as target receptors, the development of appropriate docking and scoring tools has been initiated only recently, and, accordingly, VS applications on such targets are still rare [32,33]. By contrast, the scope of applications addressing proteins ranges from enzymes to G-protein-coupled receptors and ion channels (Table 1).

Assessing the druggability of the target receptor
The first issue to be considered is the druggability of the selected target. Does the selected protein exhibit a binding pocket that can be successfully targeted by small molecule ligands? Clearly, characteristics such as pocket size and geometry, surface complexity and roughness, exposure of recognition properties and their complementarity in shape and polarity with respect to a putative druglike ligand, are of importance. Because such criteria are not immediately available, a pragmatic, but hardly generally applicable, approach would be to correlate gene families. If one member of such a family is able to bind a drug, other members might also be able to bind druglike ligands with related physicochemical properties [34]. This assumption is based simply on the fact that members of a gene family usually operate on related substrates or recognize similar endogenous ligands. However, the setup for a VS study requires more conclusive information about the actual binding-site architecture. Recently, Hajduk et al. [35] suggested several very decisive indices that help to discriminate druggable from non-druggable binding pockets. For druggability, the total surface area and a portion of a polar contact area below 75 Å^2 appear to be beneficial, along with an appropriate pocket compactness, surface roughness and complexity. This suggests that there is an optimal

size and composition of a protein-binding pocket that is best suited to recognize and accommodate small organic ligands. However, the analysis by Hajduk et al. also showed that no single index consistently dominates the correlation. This fact points to the complexity of the interrelationships among the various discriminatory indices and explains why, at present, our concepts for predicting druggability are still rudimentary.

Selection of the most relevant geometry of the target receptor
The selection of an appropriate 3D geometry for the target is another important issue when setting up a VS run. The most powerful method for learning about the spatial structure of proteins is crystal structure analysis. The accuracy and reliability of this method strongly depends on the resolution of the diffraction data. An alternative experimental technique to determine protein structures is NMR; there the accuracy of the structural determination depends strongly on the local distribution of the Nuclear Overhauser Effect distance information along the protein chain (this effect involves transfer of magnetization through space and gives information about distances between nuclei in a 3D structure). Furthermore, in the past, significant methodological enhancements in homology modelling have improved the quality of protein models [36–38]; accordingly, several successful VS screening campaigns have been reported based on model-built protein structures [39–41]. Homology modelling strongly depends on the availability of related proteins for which a crystal structure has been determined. The model will be particularly precise in those areas where the homology with the experimentally determined references is high, which is usually the case in the conserved core regions. However, functional binding sites, to be addressed by VS, are generally located in loop regions where, even among homologues of a gene family, significant differences are experienced (apart from catalytic residues in enzymes that are spatially highly conserved). To enhance the accuracy of model-built structures in regions where the binding sites are found, new algorithms have been suggested that consider the putative binding orientation of known ligands during the homology modelling process. These algorithms feed the information about the binding properties of possibly bound ligands back into the homology building process. Improved binding site geometries can be expected from such approaches [42,43].

Dealing with protein flexibility and adaptability
Another obstacle that complicates VS attempts is molecular flexibility. Ligands and proteins possess internal degrees of freedom and can adopt various conformational states. With respect to the target protein, several methods have been described to simulate flexibility [44]. For VS, it is important to obtain an estimate of the protein conformers competent at accommodating a ligand. Once these conformers have been determined, a VS run can either be performed by considering the flexibility of the protein instantly during the calculation, or by addressing an ensemble of several rigid receptor conformations [45–48]. Ligand-binding competent conformers can either be sampled by exhaustive conformational searches (e.g. using molecular dynamics simulations [49,50]) or by examining multiple conformational states, observed in crystal structures with different ligands bound, to obtain insight into
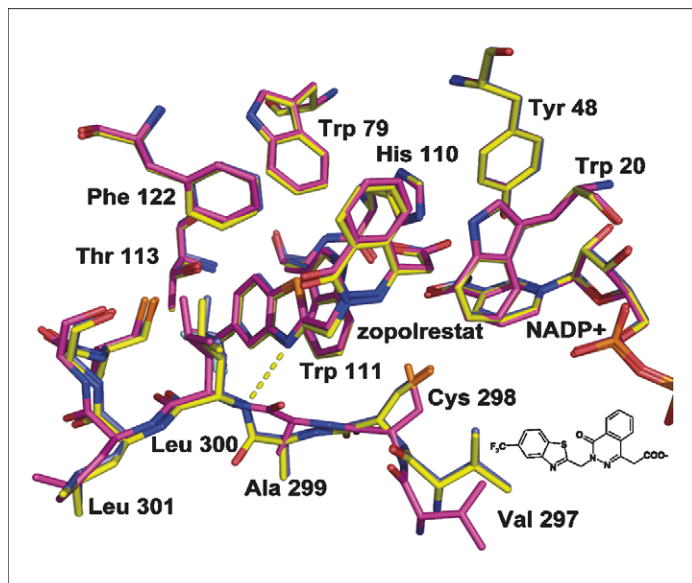
FIGURE 1

Binding mode differs with soaking and co-crystallization conditions.
Crystal structure of aldose reductase with the inhibitor zopolrestat is shown.
The crystal structure determined after one day of soaking (yellow) differs
from the structure obtained after six days of soaking (magenta) but is
identical to that obtained by cocrystallization (light blue). Whereas in the one
day-soaked structure (yellow) the amide bond between Ala299 and Leu300
orientates its NH group towards the inhibitor to form a hydrogen bond
(dotted yellow line) with the bound ligand, in the latter two structures
(magenta and light blue) this amide bond rotates away from the inhibitor and
no such hydrogen bond is observed [54].

the relevant conformations [51,52]. The latter approach appears
tempting because it can be assumed that a sample of experimen-
tally observed protein conformers corresponds to stable, low free-
energy states.

Crystal structures as the 'gold standard'

Even though the crystal structure is usually considered as the
'gold standard' for learning about the geometry of a protein, it is
highly questionable how representative a single structure deter-
mination really is. McGovern and Shoichet [53] reported on

increasing information decay, irrespective of whether the geo-
metry of a ligand-bound or ligand-free crystal structure was used
or a model-built structure was considered. Principally, crystal-
line protein–ligand complexes can be prepared by exposing
preformed crystals of a protein to the solution of a ligand, which
can then diffuse into the crystals and find its way to the binding
site. This process is called 'soaking'. Alternatively, in a cocrys-
tallization experiment, the protein–ligand complex is formed by
equilibrating them in solution, and then the assembled complex
is crystallized. We recently observed, depending on the soaking
protocol applied, different conformational states of the protein
aldose reductase complexed by the inhibitor zopolrestat [54]. In
one structure, the protein forms a hydrogen bond with the
ligand via one of its binding site-exposed amide bonds
(Figure 1). In a second structure, the amide bond is rotated
off from the binding site, and the same ligand can no longer
form this hydrogen bond. Such differences have a significant
impact on docking and scoring results in VS. Depending on the
crystallization conditions, ligands have been observed to accom-
modate a binding pocket with reversed orientation [55]; in other
cases, ligand-induced conformational adaptations of the protein
are observed [56] (Figure 2). Increasing the concentration of a
ligand in the soaking buffer will enhance the opportunity to
accommodate it in the binding pocket. However, it is likely that
soaking depicts some kind of kinetic trap for ligands; conse-
quently, surprising differences in cocrystallization have been
observed (several ligands bound at one time, multiple binding
modes [57]). Before selecting a particular crystal structure as a
reference for a VS run, detailed analysis of parameters such as
population of the bound ligand, B-factors next to the binding
site (B-factors indicate thermal motion in a crystal structure;
however, they are highly correlated with the population of a
ligand in the crystal) or consistency of the hydrogen bond
network is advisable.

Assignment of protonation states

In addition, most programs used for the actual computer screens
require properly defined protonation states of the active-site resi-
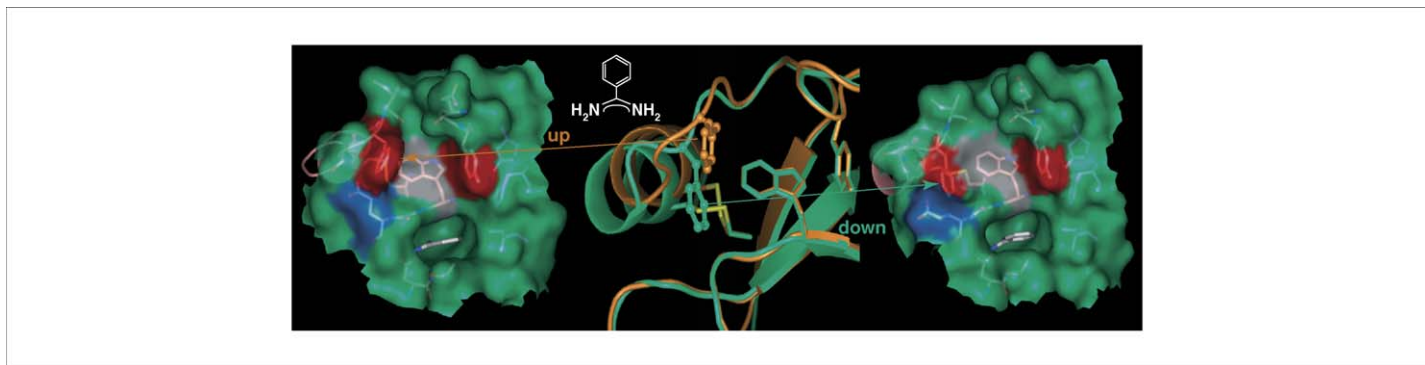dues. This is by no means easy to perform because local dielectric

FIGURE 2

Induced-fit adaptations in two different crystal forms. Two crystal structures of benzamidine (in both cases, accommodated deeply buried in the S1 pocket,
which is indicated as a deep depression in the left- and right-hand images) bound to a trypsin mutant exhibiting a binding pocket related to factor Xa. Both crystal
structures differ in terms of the conformational state of Phe174 (in the left- and right-hand images at the left rim of the binding pocket with the red surface), which
is exposed ('up') in one structure (orange and on the left) and buried ('down') in the other (green and on the right). This adaptation involves partial unwinding of a
short three-turn helix (in the centre, helix at the left-hand side of the binding pocket) and rearrangement of a disulfide bond (in the centre, yellow bond to the
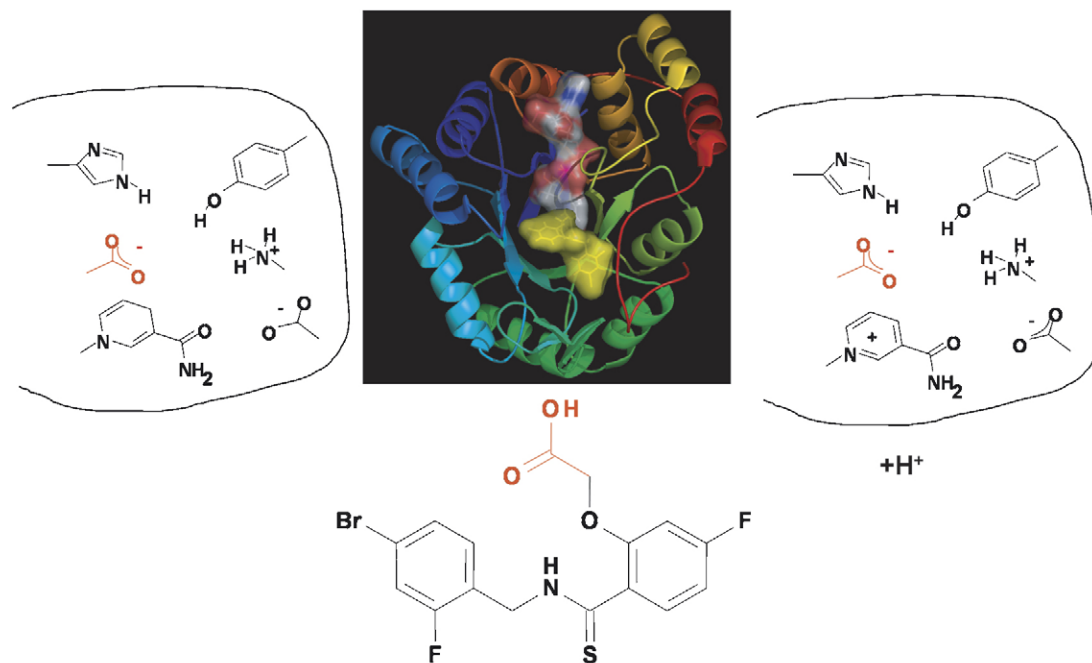back) [56,60].

Reviews • FOUNDATION REVIEW

FIGURE 3

Protonation states can change upon ligand binding. Depending on the oxidation state of the bound cofactor NADPH (left) and NADP$^+$ (right), aldose reductase binds to a carboxylate-type inhibitor (e.g. IDD594) by placing its acid functional group (red) into the catalytic centre; the complex thereby formed remains either in an unchanged protonation state (left) or picks up protons upon binding (right). In the centre, the crystal structure of the complex is shown. A short contact distance between the carboxylate function of the inhibitor (shown with a yellow surface) and the nicotinamide portion of the cofactor (shown with an atom-type coloured surface, where oxygen is red, nitrogen is blue, carbon is white, sulfur is yellow and phosphorous is orange) is formed. High resolution X-ray crystal structure analysis and neutron diffraction has provided evidence that the inhibitor binds with its carboxylate function in the deprotonated state, and the active site His is present in the neutral state. It remains unresolved where the proton goes in the case of the NADP$^+$-bound complex, or whether the residues of the catalytic centre exhibit deviating protonation states in the uncomplexed and inhibited situation. With respect to VS, a unique and precise assignment of the protonation states of the ligand and protein functional groups in a pK_a range between 3 and 11 is essential because, for example, for docking it is important whether such a group is considered as donor or acceptor of a hydrogen bond (Krämer and Klebe, unpublished).

conditions can modulate pK_a values of functional groups by several orders of magnitude. Even more complex and, at present, difficult to calculate are pK_a shifts of residues during the course of ligand binding. Isothermal titration calorimetry [58] measurements can make such changes apparent; these changes probably occur more frequently than we presently admit [59], and can easily turn an acceptor functional group into a donor or a charge-assisted hydrogen bond into a neutral one. These changes are of significance during various validation steps of a VS run. We have recently observed that even the protonation states of the residues in the binding pocket of inhibitor-bound aldose reductase can change depending on whether the cofactor NADPH/NADP$^+$ is present in the oxidized or reduced state (Figure 3) (O. Krämer and G. Klebe, unpublished results).

Single or multiple binding-competent protein conformers to be considered

With respect to protein flexibility, the decision has to be taken as to whether a VS run should consider one single protein conformer or multiple binding-competent states of the receptor. Different strategies with respect to docking and scoring will be the consequence (see below). Experience shows that conformational adaptations observed in a protein-binding pocket are usually related to the structural modulations potentially needed by the protein to accomplish its functional role [51]. Such functional adaptations must correspond to low-energy transformations, otherwise dramatic shortcomings in the functional performance of the protein would be the consequence. Accordingly, enzymes that operate on a large palette of structurally diverse substrates (cf. aldose reductase and short chain dehydrogenases) or perform substantial conformational adaptations during catalysis (cf. kinases) will experience multiple binding-competent states that have to be considered in VS. Thus, a profound understanding of the functional properties of a protein might be the best option for predicting the conformational adaptability of a protein. Such information is not usually available at the beginning of a drug development project, when VS is used. Because proteins occur in gene families of closely related members, the analysis across different entries of the family might provide some insight into the flexibility properties inherently shared by the members of the family. However, it might also be that some family members display rigid solutions, as required by their function (e.g. trypsin and factor Xa among the serine proteases), whereas others possess pronounced flexible behaviour, as appropriate to their function (e.g. the serine protease factor VIIa) [60].
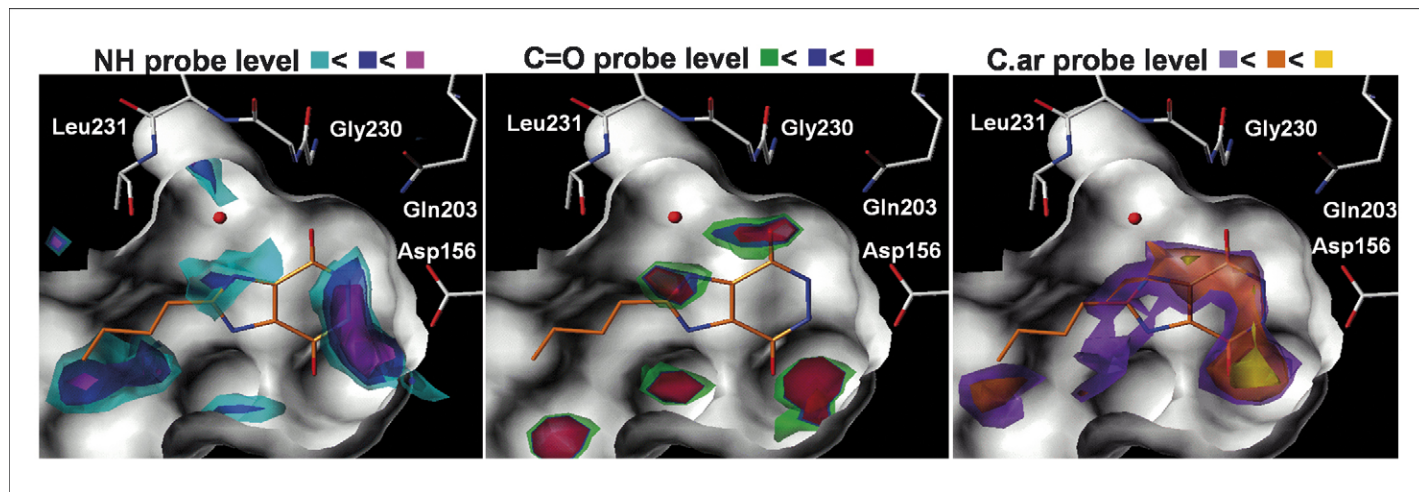
FIGURE 4

Mapping the hot spots of binding. Hotspot analysis using DrugScore for the binding pocket of t-RNA guanine transglycosylase (surface of the binding pocket indicated in white). Regions energetically favourable for the binding of a hydrogen bond donor group (represented by an NH group) or an acceptor group (represented by a C=O group), or a hydrophobic molecular portion (represented by aromatic carbon atoms, C.ar) have been analysed and contoured on three subsequent levels (indicated using three different colours) above the deepest energy minimum found for each atom type in the maps [88]. The crystallographically determined binding mode of an inhibitor is shown superimposed.

Mapping 'hot spots' of ligand binding in the target protein

After the reliability and relevance of the protein conformer(s) of the selected target protein have been assessed, its binding-site properties should be mapped before blindly starting a VS run (see below). This strategy helps the user to evaluate the properties of the target protein better and subsequently to examine the relevance of docking solutions suggested by VS. Several tools have been described to elucidate the 'hot spots' of binding in a particular binding pocket [61]. These methods are either based on thoroughly parameterized force fields [62] or on well-selected empirical information (SuperStar [63], DrugScore [64]; Figure 4). Most importantly, this analysis has to be performed on all multiple conformational states of a binding pocket because this will provide a composite picture of how the molecular recognition properties of a binding site might change upon protein adaptation.

Water, the nasty third binding partner in protein–ligand complexes

Finally, a very crucial decision with respect to the setup of the protein reference for a VS run concerns the consideration of water molecules in a binding site. The analysis of several thousand crystal structures of ligand–protein complexes using the waterbase module in Relibase [65,66] revealed that, in about two-thirds of all cases, a water molecule is involved in ligand binding, frequently mediating contacts between protein and ligand. Thus, any approach based on the prediction of binding modes of putative candidate ligands, as required in VS, must take water into consideration. One possible approach for learning about water molecules tightly bound to the protein is the analysis of crystallographic data with respect to the repeated occurrence of water molecules in structurally related binding sites (e.g. in a gene family) or multiple structural determinations of the same protein with many distinct ligands [66]. If water is present in all structures analysed at more or less the same location, this is a strong indica-

tion that this water molecule is tightly bound, and in a VS run could be considered as an integral part of the target structure.

Selection of candidate ligands

As with HTS, VS needs a thoughtfully designed and thoroughly compiled sample of small molecule candidate ligands for screening. Pharmaceutical companies will primarily screen their own proprietary compound collections, with the advantage that detected hits will be exclusive and will cover molecules for which the synthesis is well-established at their site. However, are such sample pools biased? How well is the available chemical space covered? These considerations have led large pharma companies to complement their inhouse collections with compounds offered through commercial suppliers.

Sample collection of candidate ligands for VS

Starting with the Available Chemical Directory as the initial prototype (see MDL, http://www.mdli.com), there are currently more than 10 million unique purchasable compounds on offer [67]. How well do they cover chemical space, and which portion of this space represents druglike molecules? Some dramatic figures have been proposed concerning the number of organic molecules that can be considered as druglike. Usually, molecules meet these criteria if they are composed only of the elements H, C, N, O, P, S, Cl and Br, and possess a molecular weight <500 Da [68]. How diverse should the entries of a compound database used for screening be? What physicochemical properties have to be met by the candidates to guarantee sufficient bioavailability? The concept of a well-balanced and homogeneously populated space of diverse druglike molecules appears very tempting; however, there is no proper definition of what descriptors to use as coordinates on the axes of such a compound space. What is 'diverse' in this context? As with the criterion 'similarity', the expression 'diversity' is a relative measure that relates to a reference point. In VS, the reference point is the target receptor and the difference in binding

affinity of two ligands with respect to this structure. This defines whether two molecules should be classified as 'similar' or 'diverse'. It can be imagined that, for one target, a correctly placed methyl group on a ligand might have a dramatic effect on binding affinity, whereas for another it will be tolerated with no change in binding affinity. In the first case, the ligands with and without a methyl group would have to be termed 'diverse', whereas in the second, they will be called 'similar'. However, consulting a 'diversity' scoring that concentrates solely on the topology of a ligand, the methylated and unsubstituted derivatives will probably be ranked as 'highly similar'.

Size of the candidate ligands used for VS

Is there an optimal size for candidate ligands to be submitted to screening? In the past, combinatorial chemistry in particular has enabled pharmaceutical companies to develop screening libraries of several million candidate molecules; thus, in principle, the sheer number of test compounds is no longer an issue. Although large in total count, the individual members of such libraries are usually large in terms of size and molecular weight. They are mostly in the range of typical drug-size molecules, having been synthesized in other drug development projects. However, if they turn out to be a micromolar screening hit, they still require optimization to improve their affinity towards the target protein by two or three orders of magnitude. This has to be accomplished while maintaining reasonable molecular size and appropriate absorption, distribution, metabolism and excretion properties. In consequence, it involves stripping down to a scarcely decorated core skeleton while subsequently building up this core again with novel well-tailored side chains. This is a challenging task, and is often difficult to achieve. Experience from drug optimization programs has shown that small core fragments ('privileged templates' or 'fragments') known to bind with significant affinity are ideal starting points for further optimization [69,70]. Accordingly, the compounds selected for VS should leave some room for optimization, thus matching the range of so-called leadlike or fragment-like molecules [71].

Well-balanced selection of candidate ligands for screening

Verdonk *et al.* [72] have presented a very insightful discussion on approaches to be taken when constructing databases for VS, and have suggested that the methods involved should be validated (e.g. by assessing the enrichment rates of known binders). For this purpose, in a VS campaign, known binders are pooled with the sample set of screening candidates to assess how well the known binders 'enrich' during the course of the screening process. Such considerations are important for the development of new methods; however, in an actual drug development project, at the end of the day, the medicinal chemist will ask for novel leads that are worthwhile pursuing with synthesis and optimization. Impressive enrichment rates of known actives will not be convincing.

In light of these considerations, what is a suitable compound collection? Some general criteria have to be matched either by inhouse proprietary compounds or by substances offered by commercial suppliers. Compounds can be validated with respect to drug likeness considering Lipinski's rule of five, or lead likeness applying more stringent criteria (as defined by the rule of three) [73–75]. Recently, Martin tested Lipinski's frequently applied rule with

respect to experimentally determined bioavailability in the rat [76]. This study showed that the rule of five has predictive ability for neutral and positively charged molecules; however, anions obey different rules, and the size of their polar surface area confers some predictive power. Similarly to the considerations concerning the druggability of binding sites, the criteria for bioavailability are multifactorial. Irwin and Shoichet took the initiative to set up the freely available database ZINC with validated compounds for VS [67]. It is built from 2D compound information, generates 3D coordinates and curates, if possible, from stereo- and regioisomeric ambiguities. Multiple states with respect to protonation, charges and tautomers are enumerated. However, as described for proteins, the properties of compounds might be altered upon protein binding. Insoluble, reactive and aggregating compounds are not represented in the database. The rule of five is a straightforward filter to discard compounds with putatively undesired properties from the screening sample. Depending on the strategy pursued in the subsequent VS campaign, multiple conformers can be precalculated and stored as separate entries in the database.

Data sample of already synthesized or virtual candidate molecules

Limiting the search sample to purchasable compounds in VS is a pragmatic approach because screening hypotheses can be tested rapidly [67,77]. However, VS can also scan over virtual compound libraries, and synthesis can be postponed to a later stage, considering only the most promising hits. Reymond's group attempted to generate a database of all possible organic molecules up to 160 Da under the constraints of defined chemical stability and synthetic feasibility [78]. This database contains 13.9 million entries. It is possible that such a sample could be used for fragment screening. For larger compounds, exhaustive sampling of possible molecular skeletons will end up in a combinatorial explosion with too many possible solutions. However, proper design criteria, defined by the architecture of the binding site used in VS as the target, might guide the generation of target-tailored virtual libraries for VS. In particular, considering the criteria of combinatorial chemistry and parallel synthesis, such VS strategies can actually help to synthesize only the most promising entries of a large virtual combinatorial library.

Tools and strategies for a VS campaign

Principally, two strategies can be followed in a VS campaign: forward or backward filtering of hits obtained by docking. The most crucial step in VS is the docking of candidate molecules to the target protein [79–86]. In forward filtering, various criteria are used to reduce the initial data sample, which might comprise several millions of test compounds, to the several hundred or thousand most promising candidates to be docked [87]. In backward filtering, all entries from the data sample are docked to the target protein, and filter criteria are subsequently applied to rank the generated docking solutions. Nowadays, the speed of computers is no longer the limiting factor in selecting the strategy.

Forward filtering applying increasingly stringent selection criteria

The forward technique requires fewer computational resources. This is mainly because of the fact that at each hierarchical filtering

step, a significant amount of the original data sample is discarded. However, with a decreasing number of compounds, the filtering becomes computationally increasingly demanding and sophisticated. Flexible docking is the computationally most intensive step of all; thus, the fewer candidates to be considered here, the more effort can be spent in controlling, validating and assessing docking results. This is clearly an advantage of this strategy. Forward filtering eliminates compounds initially according to simple descriptors such as molecular weight, number of rotatable bonds, lipophilicity (usually expressed by the logarithm of the partition coefficient, log P) or crude shape descriptors, such as the ellipticity of the overall structure. Subsequently, information about the receptor's binding site is exploited. Once a hotspot analysis of the most likely anchoring positions in the binding pocket has been performed, a protein-based pharmacophore can be derived [88]. This pharmacophore sets the constraints for the minimal requirement of functional groups to be matched by putative ligands (e.g. number of hydrogen bond donors, acceptors or hydrophobic groups). Molecules satisfying such criteria can be retrieved by any database engine capable of a functional group substructure search. Once the topographical arrangement of the protein-based pharmacophore has been incorporated into the search and the remaining candidates are requested to match this pattern, the study can be further focused. UNITY (UNITY Chemical Information Software, version 4.1, Tripos) and CATALYST [89] are prototypes of such database engines supporting this screening step. An alternative tool is FeatureTrees, which can retrieve molecules of similar topology in feature space [90]. In this context, 'features' are considered as being similar types of functional groups or molecular building blocks. Similarity can be considered as a further filter criterion in VS. For example, if candidate molecules are requested to match with a predefined protein-based pharmacophore, this condition already involves a selection in terms of similarity criteria; however, they are regarded very generally. Furthermore, information about known binders can be used for filtering, although the danger then exists that, owing to biased filters and preconceived concepts, some unexpected and novel chemistry is discarded during the early filtering steps. Finally, docking is pursued, usually considering only 1–10% of the initial sample collection. An advantage of the forward filtering approach is that it enables more elaborate docking protocols to be performed – for example, taking multiple protein conformational states into account or reflecting a protein-based pharmacophore as a restraint in docking. Most importantly, this hierarchical filtering strategy enables the tracking of the performance at the various filter levels by human intervention, and, in particular, visual inspection of docking solutions remains feasible.

Backwards filtering applying selection criteria on generated docking solutions

Backwards filtering starts with high-throughput docking and analyses the generated docking modes as subsequent steps. This approach is especially challenging because docking returns multiple solutions for most candidate molecules. Strongly discriminative and reliable scoring functions must be available for analysing the computed results in an automated fashion because visual inspection for the large number of diverse compounds to be examined is hardly feasible. However, it is questionable whether the existing scoring functions are reliable enough to succeed with such heroic demands (see below). Nevertheless, to proceed with the multiple docking solutions from a high-throughput run, the generated docking modes can be filtered with respect to achieved matching of the protein-based pharmacophore, contact complementarity of protein and ligand surfaces or the remaining residual unoccupied voids along the protein–ligand interface. Because nature probably avoids gaps in molecular assemblies, the latter criterion could be a powerful indicator for irrelevant binding modes or the putative accommodation of interstitial water molecules.

Selection of the most appropriate docking tool

Docking is the crucial step in VS. The seminal program DOCK, originally described in 1982 by Kuntz et al. [17], has evolved as the first VS tool. Later, other programs were successfully applied in VS, such as GOLD [91], FlexX [92], Glide [93,94] or Autodock [95], to name the most popular prototypes. These have been recently reviewed [96–102]. All docking tools follow slightly different concepts. This might give individual programs a particular advantage in one task with respect to another – for example, incorporating aspects such as flexibility of ligand and receptor or restraining the docking search engine to particular regions in configuration space (e.g. mapping a protein-based pharmacophore) [103]. However, all docking tools are still far from perfect. An even more challenging, but carelessly disregarded, aspect of docking is the appropriate consideration of water molecules. As indicated above, water molecules are involved in binding in about two-thirds of the known protein–ligand complexes. However, most docking tools ignore them simply because conclusive concepts of how to consider them correctly are missing. If structural evidence is given, preplacement of water molecules in a docking run is a feasible strategy [66]. The docking tool Slide [104] treats preplaced water molecules in a way that still enables their replacement by ligand atoms in docking. The particle concept in FlexX [105] enables the placement of water molecules 'on the fly' during the course of the generation of individual docking solutions. In the docking tool GOLD, water molecules can be switched on or off, and can spin around their principal axes to achieve good contacts with a docked ligand [106]. Recently, the popular docking tools DOCK and FlexX have been equipped with features that enable docking on a preconceived pharmacophore or property distributions [107,108]. Consideration of such criteria will drive docking solutions especially into regions either frequently trapped by other bound ligands or featured by complementary analytical tools as being particularly relevant for binding. The program AutoDock [95] performs docking on a precalculated grid, storing potential values from any sort of interaction field [109]. In the original implementation, Lennard-Jones and Coulomb potential values are used. Sotriffer et al. [110] replaced these by knowledge-based potentials, originally implemented into DrugScore. The latter potentials have proven to be powerful for discriminating, and rank among multiple ligand poses. The potential grid approach in AutoDock also enables one to average across the fields produced by various protein frames, so that the conformational degrees of freedom of a protein can be considered. In addition, adapted fields optimized with respect to the binding properties of some known actives in the comparative molecular field-type approach 'adaptation of fields for molecular comparison' (AFMoC) [111] can be used as target potential values in AutoDock [112]. The

latter docking tool performs multiple stochastic searches on the potential hypersurface; accordingly, the frequency of occurrence of certain docking solutions can be used as an additional figure of merit for their relevance [113].

Docking of small ligand fragments

The issue of an optimal ligand size for screening has been addressed above. The complexity of the docking problem increases with the size of the ligand and its number of rotatable bonds. Thus, smaller molecules in the typical range of ligand fragments should be simpler to dock. Perplexingly, present experience indicates the opposite to be the case. Fragments are easily scattered over a binding site by docking; reliably successful docking can only occur if the binding site itself is restricted in size and shows dimensions similar to the fragments [48]. Interestingly, experimental approaches show the opposite: small molecular fragments (>200 Da) usually populate, in a well-defined manner, in a very limited number of sites in binding pockets [69,70].

Enrichment rates to control the achievements of virtual screening

VS runs are usually monitored and validated by comparing the performance of a set of known actives with a large number of 'randomly' picked compounds. The actives are pooled with the random entries. All compounds are submitted to the selected VS protocol, and the performance ranks of the known actives with respect to the remaining pool are converted into enrichment plots. Such a process is essential for keeping control over the performance and achievements of VS. However, the choice of the random compound library is crucial and can strongly affect the enrichments obtained [72]. Accordingly, it has to be examined whether the pooled, randomly picked decoy structures are actually nonbinders. In a real-life scenario, it should be noted that merging known actives with a set of candidate ligands should result in a gradually declining enrichment rate at the subsequent hierarchical filter steps because novel actives, retrieved by VS, will populate in prominent ranks and dilute the set of predefined known actives.

Scoring, ranking and validation of VS docking results

Independent of the actual VS strategy applied, docking and subsequent scoring of the suggested solutions is the key performance-determining factor in VS. The discriminatory power of the applied scoring function is of utmost importance for ranking, and hopefully enriching, potentially active binders at the top of the list of docking solutions. In consequence, a myriad of scoring functions has been developed over the past few years [114–116]. Approaches have been taken not simply to rely on one single function but to take on board the consensus picture of several scoring schemes [117,118].

Scoring of docking hits by the expected binding affinity

To characterize the binding affinity of putative lead candidates experimentally, the binding constant or its inverse, the dissociation constant, is determined (or approximated by values such as the IC_{50}). If we assume that the basic rules of equilibrium thermodynamics can be applied, we can define, according to the law of mass, an equilibrium constant that describes the formation of a protein–ligand complex. This equilibrium constant is logarithmically related to the Gibbs free energy, comprising both an enthal-

pic and entropic contribution. Whereas the former relates to energetic features, the latter concerns configurational and ordering phenomena. The entropic contribution estimates how the energy content of the system is distributed over internal and external molecular degrees of freedom.

First-principle methods to compute binding affinity

Up until now, three strategies have been followed to predict binding affinities on the basis of a given protein–ligand binding geometry. The most rigorous and theoretically most solid approaches are first-principle methods [59]. Using quantum mechanics or computationally less demanding (however approximate) force fields, the partition function of a system is computed and the free energy differences between the bound and unbound state are determined. With the increasing speed of computers, such methods are becoming more accurate and obtaining a growing relevance for scoring [119]. However, screening large samples of docked solutions to estimate binding affinities is still far beyond tractability. Nevertheless, first-principle methods do not need any calibration or training in experimentally determined affinity data; thus, they will not suffer from inherent experimental shortcomings or accuracy limits. This differs from the other two approaches described in the following section, the regression- and knowledge-based scoring functions, which are based on empirical concepts [59].

Empirical scoring functions derived by regression analyses

Regression-based approaches assume additivity of individual terms considered in a master equation to describe the total Gibbs free energy of binding. In this context, a 'term' can reflect any physicochemical property of relevance for the protein–ligand binding process – for example, the number of charged or uncharged hydrogen bonds, the size of the polar or nonpolar surface portion, the number of rotatable bonds or the enthalpy required to desolvate either the ligand or the protein, among others. They are assumed to be independent from each other, and their individual contribution in reproducing the known affinities of some training set ligands is extracted by regression analysis, partial least-squares analysis or neural networks [59]. However, independence of the 'terms' is unlikely, and fair to strong correlations between terms are probable. In the correlation analysis, this fact might point out that another, at first glance surprising, property pops up as the best explanation. However, it might be the case that the latter property is not on its own essential in a physical sense, but is highly correlated with another property that is actually crucial. We recently parameterized a new regression-based scoring function on the basis of a large sample of crystal structures with known affinities for the bound ligands. Depending on the composition of the training set, different terms are found to be relevant in the analysis. This clearly shows that such empirical scoring functions reflect a best fit with respect to the training set used but they rarely achieve generality.

Empirical scoring functions derived from a knowledge base

An alternative to regression-based approaches are knowledge-based scoring functions [59]. These evaluate the occurrence fre-

quencies of some properties of interest – for example, the mutual distance between particular atom types found across the protein–ligand interface. The sample distributions describe occurrence probabilities and can be compared with a statistical mean reference situation. Any deviations from this average can be translated using some type of mathematical relationship into statistical preferences that can be used to determine the geometry of protein–ligand complexes. Conceptionally, the knowledge-based functions appear to be more general because no master equation with preconceived 'terms' is required. However, the data selected to derive the function, and the definition of atom types, together with many adjustable parameters needed to actually establish the method, also attenuates the generality of this approach.

Accuracy of the estimate of binding affinity

To estimate binding constants, both the regression- and knowledge-based scoring functions require experimental affinity data for internal calibration. Thus, their prediction accuracy can never be better than the precision by which binding data can be measured. It is interesting to note that the estimated standard deviations of such empirical scoring functions are reduced if data for a selected number of targets, determined in one laboratory based on assay data recorded under strictly conserved conditions, are used or if broad-range data covering many targets are considered based on assay data determined in various laboratories. The relative differences in binding data within a series of compounds can be determined precisely – usually much better than across data from various systems for which the comparison has to be performed on an absolute scale. In consequence, the standard deviations of predicted binding affinities of presently available functions range between 0.7 and 1.5 logarithmic units in binding affinity. This range matches the experimental accuracy achieved for binding data across training sets of growing heterogeneity covering a broad spectrum of targets. In our experience, knowledge-based scoring functions are better able to extract binding geometries generated by a docking program that closely approximate the experimentally confirmed binding mode from a sample set of decoy placements. By contrast, regression-based scoring functions are better in the actual affinity prediction, provided that a fairly accurate binding geometry is given.

Unlike the professed opinion that in docking the geometry problem has been resolved to a sufficient extent, and that the scoring problem remains an open question [120,121], it appears that both are intimately related. We recently developed a knowledge-based scoring function based on accurate contact data from small molecule crystal data [122]. This function reliably recognizes the experimentally determined binding mode found in a crystal structure among a set of decoy poses. It appears self-evident that the scoring problem can only be alleviated if more relevant, near-native binding poses are produced by docking programs and reliably recognized by a scoring function. Accordingly, it appears advisable to drive docking solutions as close as possible to the native geometries – that is, as they would show up in a corresponding crystal structure. This can also be achieved by optimizing them towards the near-native situation. At best, this involves minimization with respect to the function used for scoring. As a disadvantage, this process would be computationally demanding.

Binding affinity: a sufficiently well understood property?

Any discussion of scoring and ranking must ask the question as to how well we understand the target value Gibbs free energy. Is it advisable to focus scoring on free energy or would it be better to treat enthalpy and entropy as separate terms in the scoring [123]? There is a mutual compensation between enthalpy and entropy owing to the fact that both entities scatter over much larger ranges than does the free energy itself. Considering the binding of a ligand to a protein, enthalpy (ΔH) and entropy (-$T\Delta S$) can easily scatter over a 3–4-fold larger range than the spread of Gibbs free energy (ΔG). The mutual compensation of enthalpy and entropy can even be found across closely related molecules [124] (Figure 5). Furthermore, many physicochemical phenomena of relevance for the binding process are not yet fully understood and, therefore, have not yet been correctly incorporated into scoring functions (e.g. the role of water, change in protonation states or an appropriate consideration of entropy). Interestingly, microcalorimetry indicates that with increasing temperature, the protein–ligand binding process becomes more exothermic and entropically less favourable [58,59]. Because this observation applies to all targets, it suggests that general phenomena are involved which are not yet understood at a molecular level. Despite our present deficiencies in understanding the physics of the binding process, scoring still works satisfactorily – most likely because we consider the binding of ligands to a protein on a relative scale to each other. Accordingly, any unappreciated phenomena, similar across all complexes in the analysis, will simply be cancelled out. For example, as mentioned above, one approach for reflecting protein flexibility applies parallel docking into several rigid binding-competent conformers of the protein. The disadvantage of this strategy is the fact that additional degrees of scoring are created: what discriminates the scoring against different protein conformers? Cancellation of unreflected internal protein energy contributions is no longer certain. Studies have shown that a special scoring scheme or protocol is required [48,52]. Because dramatic energy differences between low-energy conformers are unlikely, a modulated pocket size for the different protein conformers might require individual scoring of the altered desolvation properties of the binding site.

Required accuracy of scoring: enrichment or hit identification?

VS is used to enrich putative actives from a large sample set of test ligands. Accordingly, the desired accuracy of scoring depends on whether, at first glance, prioritization of the sample set for testing is anticipated or whether putative actives are expected among the top ten or 100 of a hit list. The latter requires very powerful discrimination of actives over inactive decoy binders. Present scoring functions have been optimized to discriminate for a particular ligand decoy binding mode from near-native ones. The discrimination of binders from decoy nonbinders still remains as major challenge for VS protocols [125]. At worst, in these cases the above-mentioned poorly understood contributions to binding become overwhelmingly important. Perhaps the consideration of similarity criteria in the search [87,107,108,110–112] enables the problem to be alleviated to some degree because this drives the search towards more closely related ligands for which some of the disregarded effects in scoring cancel themselves out.
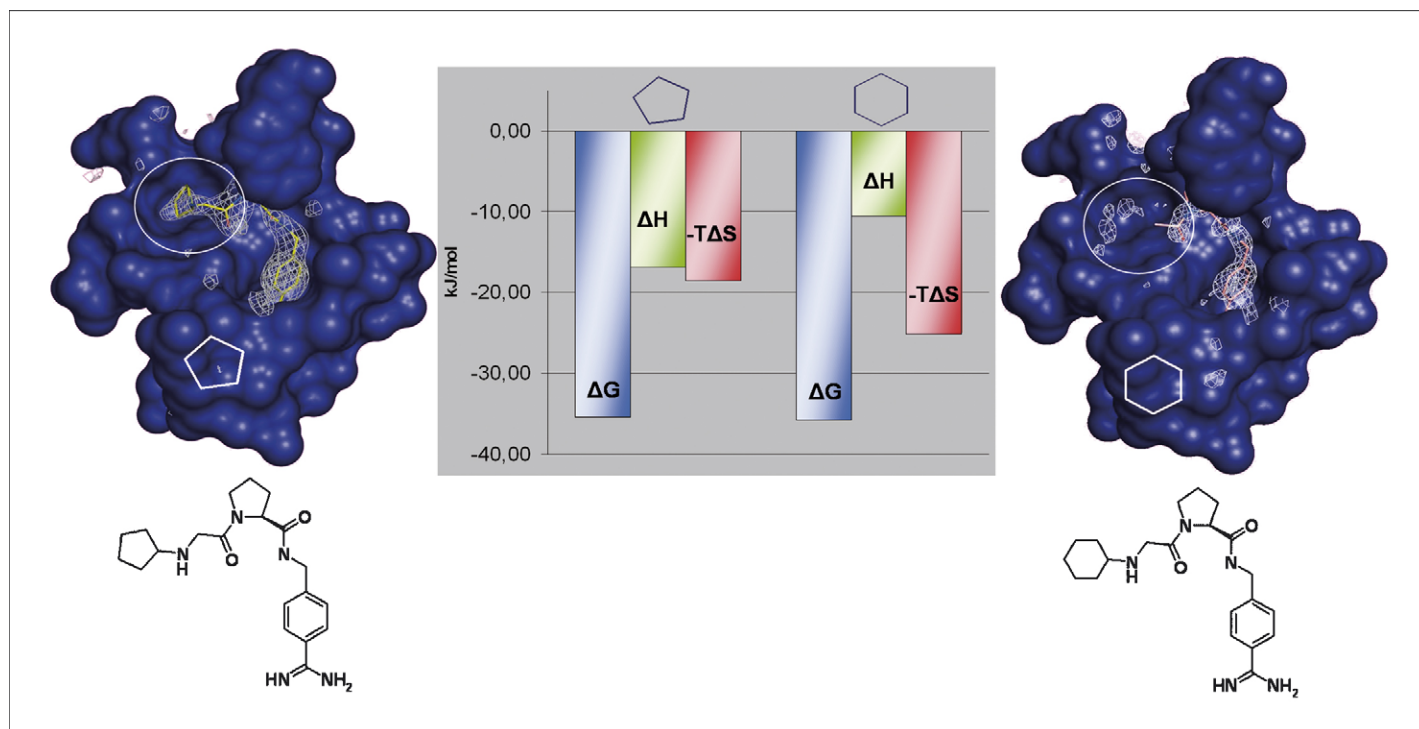
Reviews • FOUNDATION REVIEW

FIGURE 5

Similar ligands decompose differently into enthalpic and entropic binding contributions. Crystal structures of two closely related thrombin inhibitors bearing a cyclopentyl or cyclohexyl moiety as terminal substituent to accommodate the S3/S4 pocket of the catalytic site (surface of the binding pocket is indicated in blue). Whereas the five-membered ring (left) gives rise to a well-defined difference in electron density (white 'chicken-wire' contouring), the six-membered ring (right) cannot be assigned to any density (see inside white circles). It is likely that the latter fragment shows enhanced residual mobility and is scattered around several conformational states. Interestingly, this deviating behaviour is well reflected in the thermodynamic properties (centre). Both compounds exhibit the same free energy of binding (ΔG, blue columns). However, the cyclohexyl derivative (right) with the enhanced residual mobility is entropically (-$T\Delta S$, red columns) more favoured than the 'less-well clamped' five-ring derivative (left). The latter experiences better enthalpic contributions (ΔH, green columns) to binding [124].

Ultimate proof of concept: crystal structure analysis of VS hits

Only rarely, the crystal structure of detected VS hits has been determined and subsequently published. Such case studies have recently been reviewed by Shoichet [24]. Cases have been described for which VS has correctly predicted the subsequently found binding mode. Other examples point to deficiencies arising from the superficially understood phenomena described above. Finally, it has also been reported that VS inappropriately identified an active compound because binding actually occurs in a totally different region of the protein surface.

Conclusions and outlook

VS has been established as a powerful alternative and complement to HTS. When performed optimally, impressive hit rates have been reported, which have been significantly higher (by a factor of 100–1000 [24]) than those for HTS. Comparative studies of HTS and VS indicate that the methods can capture alternative and complementary ligands.

Undoubtedly, VS is not yet a fully mature technology following a well-established process line. Few of the foundations of protein–ligand recognition are understood well enough to be deployed in a large scale, multi-compound effort such as that commonly undertaken in VS. This calls for further indepth research. In particular, protein flexibility and induced-fit adaptations, the role of water in solvation, desolvation and ligand binding, and the electrostatics

involved, including changes in protonation states and an appropriate consideration of entropic changes, will need to be better understood to improve the hit retrieval rates in VS. Frequently, experimental work performed in parallel or as a follow-up to a VS campaign provides a whole bunch of unexpected results pointing towards manifold deficiencies of the concepts applied. Many more experimental studies are required. Nevertheless, VS has proven successful and as a valuable alternative to HTS, in particular if it is used as a tool to support and complement hit discovery. Compared with HTS, it is significantly cheaper and faster to use. It can be easily rerun under modified conditions – for example, if additional information about the target protein under consideration becomes available or novel filter criteria are taken into account.

It is interesting to note that an experienced modeller or medicinal chemist can often figure out whether a particular binding pocket appears druggable or a certain molecule obeys the rules of drug likeness; however, putting such knowledge into computer algorithms makes the multifactorial nature of these rules apparent, complicating attempts to generalize them. The same multifactorial nature holds true for docking and scoring. It is therefore highly advisable to refrain from fully automated strategies in VS. Experience and human intervention are of the utmost importance for keeping control over the various filter steps in a VS run. Used in such a way, VS can be successful; in particular, if information about molecular similarity is considered in terms of generic physicochemical properties and not simply as chemical formulae. Con-

sidering similarity concepts takes the risk that highly diverse molecules remain undetected; however, it probably makes the searches simpler because many parameters that actually matter in VS, and which are not properly considered, simply cancel each other out in a relative comparison. Considering all of the mentioned limitations, and taking molecular similarity as some kind of work-around to evade existing shortcomings, a critical reviewer might raise the question of whether, under such restrained conditions, VS can really contribute something novel that an experienced medicinal chemist would not have thought about. Only successful examples can convince. We applied VS to a tRNA-modifying enzyme for which the inhibitor replaces either guanine or preQ1 (a precursor to the modified base queuine) as a substrate in the enzyme recognition pocket (Figure 6). Initial VS searches retrieved structures that, admittedly, an experienced medicinal

chemist might also have suggested as being 'substrate-like'. However, a follow-up VS screen in which several water molecules were allowed to be replaced in the binding site suggested a cyclic urea structure with a different orientation in the binding pocket [126]. Synthetically, this lead appeared tempting and presently serves as a starting point for a new series of compounds (Figure 6). This unbiased approach, resulting, in chemical terms, in an unexpected lead, demonstrates that VS can make important contributions to drug discovery, providing some unexpected candidates. Nevertheless, there is still a long way to go until it becomes an established tool for routine lead discovery. Interestingly, nowadays, most aspects of contemporary drug discovery are optimized towards high throughput; by contrast, our current increase in the knowledge and understanding of protein–ligand recognition principles is still proceeding at a low-output rate.
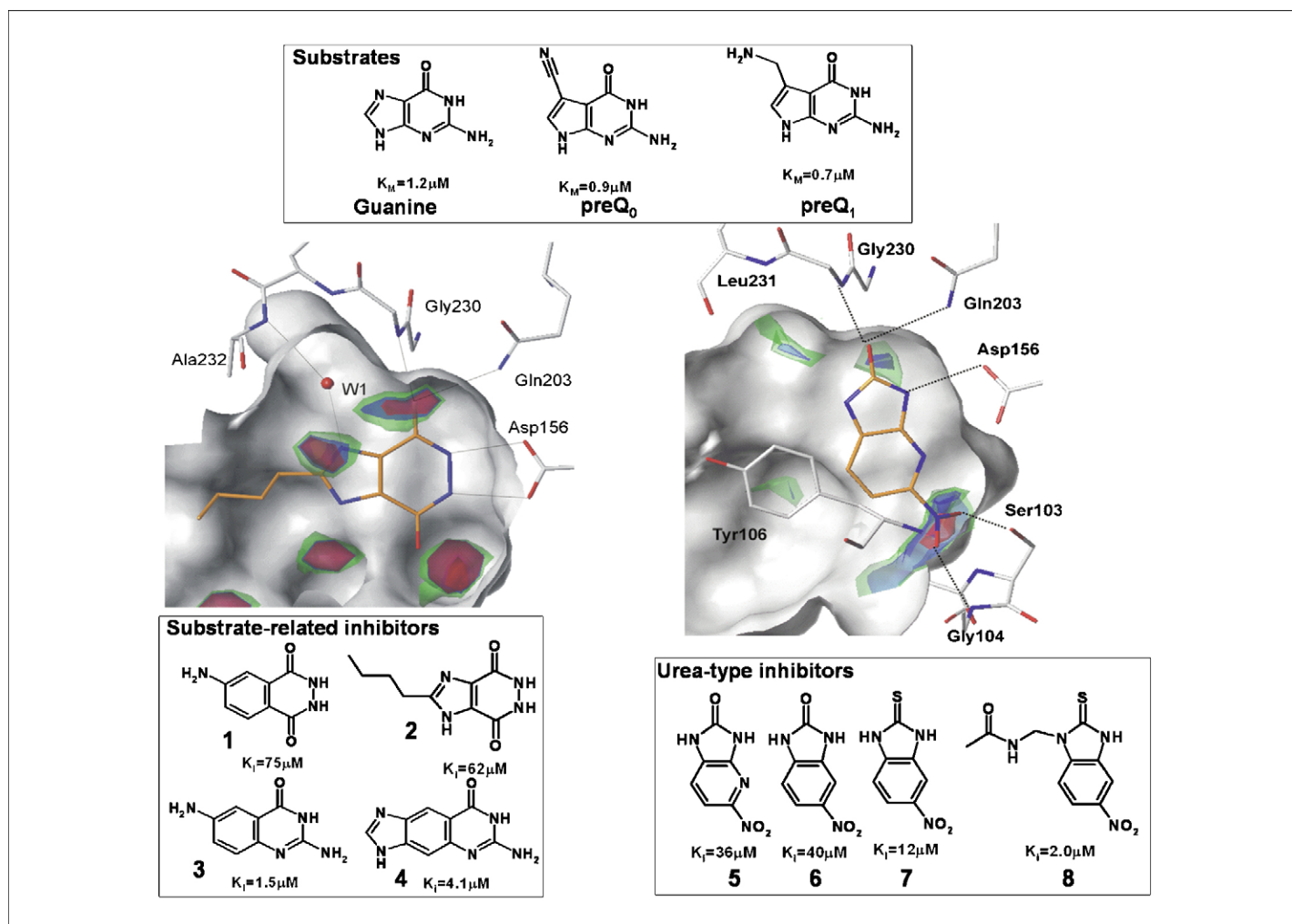

FIGURE 6

An unexpected ligand skeleton from virtual screening. Virtual screening (VS) has been used to search for putative inhibitors of t-RNA guanine transglycosylase [88]. Initial hits such as 1–4 (lower left box), which were followed up by chemical synthesis, showed structural similarity with the natural substrates of this enzyme guanine, preQ$_0$ and preQ$_1$ (upper box). These searches were based on the protein-based pharmacophore shown on the left (contours for hydrogen bond acceptor group) and in Figure 4. An experienced medicinal chemist could possibly have come up with similar suggestions for potential leads such as 1–4. However, in a second VS campaign we focused on the replacement of two water molecules at the lower right rim of the binding pocket of t-RNA guanine transglycosylase (right, contours for hydrogen bond acceptor group [126]). This screen suggested 5 as a potential hit. Its cyclic urea-type skeleton is distinct in its structure and binding mode from any known natural substrate and it can serve as a synthetically easily accessible lead. Several derivatives (e.g. 6–8) have been synthesized and show a reasonable structure–activity relationship (Stengl et al., unpublished). It is unlikely that this novel lead structure would have been suggested by comparative substrate considerations. This example underlines the power of VS as an alternative source for novel lead discovery.

References

1 Bolten, B.M. and DeGregorio, T. (2002) Trends in development cycles. *Nat. Rev. Drug Discov.* 1, 335–336
2 Fishman, M.C. and Porter, J.A. (2005) A new grammar for drug discovery. *Nature* 437, 491–493
3 Lahana, R. (1999) How many leads from HTS? *Drug Discov. Today* 4, 447–448
4 Ramesha, C.S. (2000) Comment: How many leads from HTS? *Drug Discov. Today* 5, 43–44
5 Langer, T. and Krovat, E.M. (2003) Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr. Opin. Drug Discov. Devel.* 6, 370–376
6 Jenkins, J.L. *et al.* (2003) Virtual screening to enrich hit lists from high-throughput screening: a case study on small-molecule inhibitors of angiogenin. *Proteins* 50, 81–93
7 Doman, T.N. *et al.* (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* 45, 2213–2221
8 Evensen, E. *et al.* (2003) Comparing performance of computational tools for combinatorial library design. *J. Med. Chem.* 46, 5125–5128
9 Paiva, A.M. *et al.* (2001) Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of *Mycobacterium tuberculosis*. *Biochim. Biophys. Acta* 1545, 67–77
10 Zolli-Juran, M. *et al.* (2003) High throughput screening identifies novel inhibitors of *Escherichia coli* dihydrofolate reductase that are competitive with dihydrofolate. *Bioorg. Med. Chem. Lett.* 13, 2493–2496
11 Elowe, N.H. *et al.* (2005) Experimental screening of dihydrofolate reductase yields a ''test set'' of 50,000 small molecules for a computational data-mining and docking competition. *J. Biomol. Screen* 10, 653–657
12 Parker, C.N. (2005) McMaster University Data-mining And Docking Competition: computational models on the catwalk. *J. Biomol. Screen* 10, 647–648
13 Lang, P.T. *et al.* (2005) Evaluating the high-throughput screening computations. *J. Biomol. Screen* 10, 649–652
14 Bender, A. *et al.* (2005) Screening for dihydrofolate reductase inhibitors using MOLPRINT 2D, a fast fragment-based method employing the naïve Bayesian classifier: limitations of the descriptor and the importance of balanced chemistry in training and test sets. *J. Biomol. Screen* 10, 658–666
15 Bernacki, K. *et al.* (2005) Virtual ligand screening against *Escherichia coli* dihydrofolate reductase: improving docking enrichment using physics-based methods. *J. Biomol. Screen* 10, 675–681
16 Brenk, R. *et al.* (2005) Here be dragons: docking and screening in an uncharted region of chemical space. *J. Biomol. Screen* 10, 667–674
17 Kuntz, I.D. *et al.* (1982) A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 161, 269–288
18 DesJarlais, R.L. *et al.* (1986) Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.* 29, 2149–2153
19 DesJarlais, R.L. *et al.* (1990) Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. U. S. A.* 87, 6644–6648
20 Rutenber, E. *et al.* (1993) Structure of a nonpeptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design. *J. Biol. Chem.* 268, 15343–15346
21 Lam, P.Y. *et al.* (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263, 380–384
22 Lyne, P.D. (2002) Structure-based virtual screening: an overview. *Drug Discov. Today* 7, 1047–1055
23 Barril, X. *et al.* (2004) Virtual screening in structure-based drug design. *Mini Rev. Med. Chem.* 4, 779–791
24 Shoichet, B.K. (2004) Virtual screening of chemical libraries. *Nature* 432, 862–865
25 Chin, D.N. *et al.* (2004) Integration of virtual screening into the drug discovery process. *Mini Rev. Med. Chem.* 4, 1053–1065
26 Jain, A.N. (2004) Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discov. Devel.* 7, 396–403
27 Stahura, F.L. and Bajorath, J. (2004) Virtual screening methods that complement HTS. *Comb. Chem. High Throughput Screen* 7, 259–269
28 Hou, T. and Xu, X. (2004) Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.* 10, 1011–1033
29 Kitchen, D.B. *et al.* (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949
30 Bleicher, K. *et al.* (2003) Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discov.* 2, 369–378
31 Kubinyi, H. (2006) *Success Stories of Computer-Aided Design, in: Computer Applications in Pharmaceutical Research and Development* (Ekins, S. Ed.) [*Wiley Series in Drug Discovery and Development* (Wang, B. Ed.)], Wiley-Interscience, pp. 377–424.
32 Morley, S.D. and Afshar, M. (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock. *J. Comput. Aided Mol. Des.* 18, 189–208
33 Filikov, A.V. *et al.* (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J. Comput. Aided Mol. Des.* 14, 593–610
34 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730
35 Hajduk, P.J. *et al.* (2005) Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* 48, 2518–2525
36 Fiser, A. and Sali, A. (2003) Comparative protein structure modelling. In *Protein Structure* (Chasman, D., ed.), pp. 167–206, Marcel Dekker, Inc.
37 Wallner, B. and Elofsson, A. (2005) All are not equal: a benchmark of different homology modelling programs. *Protein Sci.* 14, 1315–1327
38 Hillisch, A. *et al.* (2004) Utility of homology models in the drug discovery process. *Drug Discov. Today* 9, 659–669
39 Bissantz, C. *et al.* (2003) Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins* 50, 5–25
40 Evers, A. and Klebe, G. (2004) Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J. Med. Chem.* 47, 5381–5392
41 Evers, A. and Klabunde, T. (2005) Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* 48, 1088–1097
42 Schafferhans, A. and Klebe, G. (2001) Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* 307, 407–427
43 Evers, A. *et al.* (2003) Ligand-supported homology modelling of protein binding sites using knowledge-based potentials. *J. Mol. Biol.* 334, 327–345
44 Teague, S.J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* 2, 527–541
45 Knegtel, R.M. *et al.* (1997) Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266, 424–440
46 Leach, A.R. (1994) Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 235, 345–356
47 Claussen, H. *et al.* (2001) FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308, 377–395
48 Wei, B.Q. *et al.* (2004) Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* 337, 1161–1182
49 Carlson, H.A. and McCammon, J.A. (2000) Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* 57, 213–218
50 Meagher, K.L. and Carlson, H.A. (2004) Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case. *J. Am. Chem. Soc.* 126, 13276–13281
51 Sotriffer, C.A. *et al.* (2004) Probing flexibility and ''induced-fit'' phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins* 56, 52–66
52 Barril, X. and Morley, S.D. (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* 48, 4432–4443
53 McGovern, S.L. and Shoichet, B.K. (2003) Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* 46, 2895–2907

54 Steuber *et al.* Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying conditions. *J. Med. Chem.* (In press).

55 Stubbs, M.T. *et al.* (2002) pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. *ChemBioChem* 3, 246–249

56 Rauh, D. *et al.* (2003) Understanding protein–ligand interactions: the price of protein flexibility. *J. Mol. Biol.* 335, 1325–1341

57 Rauh, D. *et al.* (2003) ZZ Made EZ: influence of inhibitor configuration on enzyme selectivity. *J. Mol. Biol.* 330, 761–770

58 Jelesarov, I. and Bosshard, H.R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* 12, 3–18

59 Gohlke, H. and Klebe, G. (2002) Approaches to the description and prediction of binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed. Engl.* 41, 2644–2676

60 Reyda, S. *et al.* (2003) Reconstructing the binding site of factor Xa in trypsin reveals ligand-induced structural plasticity. *J. Mol. Biol.* 325, 963–977

61 Gohlke, H. *et al.* (2000) Predicting binding modes, binding affinities and ''hot spots'' for protein–ligand complexes using a knowledge-based scoring function. *Perspect. Drug Discov. Des.* 20, 115–144

62 Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857

63 Verdonk, M.L. *et al.* (1999) SuperStar: a knowledge based approach for identifying interaction sites in proteins. *J. Mol. Biol.* 289, 1093–1108

64 Gohlke, H. *et al.* (2000) Knowledge-based scoring function to predict protein–ligand interactions. *J. Mol. Biol.* 295, 337–356

65 Hendlich, M. *et al.* (2003) Relibase: design and development of a database for comprehensive analysis of protein–ligand interactions. *J. Mol. Biol.* 326, 607–620

66 Günther, J. *et al.* (2003) Utilising structural knowledge in drug design strategies: applications using Relibase. *J. Mol. Biol.* 326, 621–636

67 Irwin, J.J. and Shoichet, B.K. (2005) ZINC- a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model* 45, 177–182

68 Bohacek, R.S. *et al.* (1996) The art and practice of structure-based drug design: a molecular modelling perspective. *Med. Res. Rev.* 16, 3–50

69 Erlanson, D.A. *et al.* (2004) Fragment-based drug discovery. *J. Med. Chem.* 47, 3463–3482

70 Rees, D.C. *et al.* (2004) Fragment-based lead discovery. *Nat. Rev. Drug Discov.* 3, 660–672

71 Oprea, T.I. (2002) Current trends in lead discovery: are we looking for the appropriate properties? *J. Comput. Aided Mol. Des.* 16, 325–334

72 Verdonk, M.L. *et al.* (2004) Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* 44, 793–806

73 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26

74 Congreve, M. *et al.* (2003) A rule-of-three for fragment-based lead discovery. *Drug Discov. Today* 8, 876–877

75 Van de Waterbeemd, H. and Gifford, E. (2003) ADMET *in silico* modelling: towards prediction paradise? *Nat Rev. Drug Discov.* 2, 192–204

76 Martin, Y.C. (2005) A bioavailability score. *J. Med. Chem.* 48, 3164–3170

77 Baurin, N. *et al.* (2004) Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. *J. Chem. Inf. Comput. Sci.* 44, 643–651

78 Fink, T. *et al.* (2005) Virtual exploration of the small-molecule chemical universe below 160 daltons. *Angew. Chem. Int. Ed. Engl.* 44, 1504–1508

79 Shoichet, B.K. *et al.* (2002) Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* 6, 439–446

80 Muegge, I. and Rarey, M. (2001) Small molecule docking and scoring. In *Reviews in Computational Chemistry* (Lipkowitz, K.B. and Boyd, D.B., eds), pp. 1–60, Wiley-VCH

81 Abagyan, R. and Totrov, M. (2001) High-throughput docking for lead generation.. *Curr. Opin. Chem. Biol* 5, 375–382

82 Mohan, V. *et al.* (2005) Docking: successes and challenges. *Curr. Pharm. Des.* 11, 323–333

83 Alvarez, J.C. (2004) High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.* 8, 365–370

84 Cole, J.C. *et al.* (2005) Comparing protein-ligand docking programs is difficult. *Proteins* 60, 325–332

85 Perola, E. *et al.* (2004) Detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* 56, 235–249

86 Jansen, J.M. and Martin, E.J. (2004) Target-based scoring approaches and expert systems in structure-based virtual screening. *Curr. Opin. Chem. Biol.* 8, 359–364

87 Grüneberg, S. *et al.* (2002) Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J. Med. Chem.* 45, 3588–3602

88 Brenk, R. *et al.* (2003) Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.* 46, 1133–1143

89 Moebius, F.F. *et al.* (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr. Med. Chem.* 8, 1035–1055

90 Rarey, M. and Dixon, J.S. (1998) Feature trees: a new molecular similarity measure based on tree matching. *J. Comput. Aided Mol. Des.* 12, 471–490

91 Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748

92 Rarey, M. *et al.* (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489

93 Friesner, R.A. *et al.* (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749

94 Halgren, T.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* 47, 1750–1759

95 Morris, G.M. *et al.* (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662

96 Abagyan, R. and Totrov, M. (2001) High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* 5, 375–382

97 Halperin, I. *et al.* (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 47, 409–443

98 Brooijmans, N. and Kuntz, I.D. (2003) Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* 32, 335–373

99 Cummings, M.D. *et al.* (2005) Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* 48, 962–976

100 Schulz-Gasch, T. and Stahl, M. (2003) Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J. Mol. Model. (Online)* 9, 47–57

101 Kellenberger, E. *et al.* (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* 57, 225–242

102 Perola, E. *et al.* (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* 56, 235–249

103 Joseph-McCarthy, D. *et al.* (2003) Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins* 51, 172–188

104 Schnecke, V. and Kuhn, L.A. (1999) Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discov. Des.* 20, 171–190

105 Rarey, M. (1999) The particle concept: placing discrete water molecules during protein–ligand docking predictions. *Proteins* 34, 17–28

106 Verdonk, M.L. *et al.* (2005) Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* 48, 6504–6515

107 Hindle, S.A. *et al.* (2002) Flexible docking under pharmacophore type constraints. *J. Comput. Aided Mol. Des.* 16, 129–149

108 Fradera, X. *et al.* (2000) Similarity-driven flexible ligand docking. *Proteins* 40, 623–636

109 Österberg, F. *et al.* (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 46, 34–40

110 Sotriffer, C.A. *et al.* (2002) Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. *J. Med. Chem.* 45, 1967–1970

111 Gohlke, H. and Klebe, G. (2002) DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.* 45, 4153–4170

112 Radestock, S. *et al.* (2005) Improving binding mode predictions by docking into protein-specifically adapted potential fields. *J. Med. Chem.* 48, 5466–5479

113 Sotriffer, C.A. *et al.* (1996) Comparative docking studies on ligand binding to the multispecific antibodies IgE-La2 and IgE-Lb4. *J. Comput. Aided Mol. Des.* 10, 305–320

114 Stahl, M. and Rarey, M. (2001) Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* 44, 1035–1042

115 Wang, R. *et al.* (2003) Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* 46, 2287–2303

116 Ferrara, P. *et al.* (2004) Assessing scoring functions for protein–ligand interactions. *J. Med. Chem.* 47, 3032–3047

117 Charifson, P.S. *et al.* (1999) Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* 42, 5100–5109

118 Wang, R. and Wang, S. (2001) How does consensus scoring work for virtual library screening? An idealized computer experiment *J. Chem. Inf. Comput. Sci.* 41, 1422–1426

Reviews • FOUNDATION REVIEW

119 Raha, K. and Merz, K.M.FJr (2005) Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein–ligand complexes. *J. Med. Chem.* 48, 4558–4575

120 Dixon, J.S. (1997) Evaluation of the CASP2 docking section. *Proteins* (Suppl.), 198–204

121 Lipinski, C.A. (2003) Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discov. Today* 8, 12–16

122 Velec, H.F.G. *et al.* (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* 48, 6296–6303

123 Luque, I. and Freire, E. (2002) Structural parameterization of the binding enthalpy of small ligands. *Proteins* 49, 181–190

124 Gerlach, C. *et al.* (2005). Library design based on privileged scaffolds through docking and direct design in the protein binding pocket. Abstracts of Papers, 230th ACS National Meeting, Washington, DC, USA, 28th August–1st September, COMP-185.

125 Graves, A.P. *et al.* (2005) Decoys for docking. *J. Med. Chem.* 48, 3714–3728

126 Brenk, R. *et al.* (2004) Crystallographic study of inhibitors of tRNA-guanine transglycosylase suggests a new structure-based pharmacophore for virtual screening. *J. Mol. Biol.* 338, 55–75