Virtual Screening in Structure-Based Drug Discovery

X. Barril*, R.E. Hubbard and S.D. Morley

Vernalis (R & D) Granta Park. Cambridge. CB1 6GB. UK

Abstract: Recent advances in structure determination and computational methods have encouraged the development of structure-based virtual screening. Here we survey progress in the field and review the most recent methods, validation experiments and real applications, including an in-house example of hit identification for the oncology target Hsp90. These results provide a basis for discussing the current state of structure-based virtual screening and to outline the developments that are expected to have a major impact in the near future.

Key Words: Virtual Screening; In-Silico Screening; Docking; Structure-Based Drug Design; Hsp90; Binding Mode Prediction; Seeded Libraries

INTRODUCTION

Both the hit identification and early lead optimisation stages of drug discovery are typically dominated by the search for compounds with improved inhibitory activity. Historical libraries, as well as newly synthesized or acquired compounds, are tested *in vitro* to identify those showing higher potency. During the last decade or so, the main strategy to maximize the chances of success and speed-up the early phases of drug discovery has been the assay of vast numbers of compounds by means of high-throughput screening (HTS). Although this strategy has certainly provided some successes, it has not fulfilled the initial expectations and it is widely accepted nowadays that a blunt increase in the number of assayed compounds does not guarantee better productivity *per se* [1].

The screening of virtual libraries of compounds by computational means, also known as virtual screening (VS), is conceptually and economically very attractive because it makes possible the evaluation *in silico* of an almost unlimited number of chemical structures, only a subset of which will be selected and subsequently assayed in a low-, medium- or high-throughput screening experiment. Although VS is often presented as an alternative to HTS, in fact both methodologies are highly complementary as noted in recent reviews [2,3].

Many different methodologies can be included under the generic title of VS. Based on the criteria used to evaluate each compound, the following classes can be defined:

i) Chemoinformatics tools. This includes substructural searches as well as the use of chemical descriptors and/or properties (either computationally or experimentally determined) to select molecules based on target-independent criteria such as drug-likeness [4,5], lead-likeness [6] or diversity.

ii) Ligand-based methods. In the case when one or more compounds are known to bind to the biological target, it is possible to identify molecules which share common features with them. The underlying assumption is that similar compounds can have similar effects; therefore the selected compounds are expected to have an increased probability of binding.

iii) Structure-based methods. These methods use the structure of the biological target and knowledge about the molecular recognition process to single out compounds with a good chance of being ligands.

This division is somewhat arbitrary, but is aimed to reflect both the level of insight into the therapeutic target and the complexity and computational cost of the VS process. Whereas the cheminformatics tools are computationally very efficient, only ligand-based and structure-based methods are able to select a small number of structurally diverse compounds with a very significant hit enrichment. The choice between the two latter methods is usually based on the availability of the 3D structure of the therapeutic target.

The acquisition of the structure of the macromolecular target opens up the possibility of understanding the binding process and, hence, to rationalize the drug discovery process and make it more predictable and efficient. This paradigm, together with recent advances in high-throughput crystallography and structure determination [7] has boosted structure-based drug design and the application of structure-based tools to VS.

A VS campaign typically considers a large number of compounds (10^5 to 10^7). This requirement discards the most sophisticated structure-based methods and selects molecular docking as the tool providing the best balance between the quality of the predictions and the computational cost. In this review we will describe how molecular docking works, analyse its performance both in theoretical and real VS experiments, discuss its applicability to the pharmaceutical drug discovery process and highlight present and future methodological challenges.

SUMMARY OF PUBLISHED APPROACHES

Molecular docking was first applied to drug design more than 20 years ago [8] as a computational tool able to generate putative binding modes (search algorithm) and rank them from more to less likely (scoring function). Although

*Address correspondence to this author at Vernalis (R & D), Granta Park. Cambridge. CB1 6GB. UK; Tel: + 44 (0) 1223 89 55 55; Fax:+44 (0) 12 23 89 55 56; E-mail: x.barril@vernalis.com

1389-5575/04 $45.00+.00

the basic principles remain the same, many new algorithms and scoring functions have been developed. A detailed survey of the progress in the field has been presented in recent reviews[9-12] and is out of the scope of this paper. Here we will focus on the similarities and differences between software packages used in virtual screening.

Virtual screening docking software can be broadly categorised according to: i) the scoring function (force field, empirical, knowledge-based), ii) the ligand docking approach (full flexibility, precomputed conformer library, incremental/fragment construction), and iii) the search algorithm (systematic, stochastic, deterministic). Some of the packages support more than one docking algorithm and scoring function. In addition, it is common practice to use a reduced, fast scoring function during the docking process itself to generate reasonable docking poses for each ligand, and to introduce more sophisticated, but slower calculations to rank order the ligands for final selection. A major area of current research is to introduce more realistic treatments of desolvation effects into virtual screening scoring functions of all types.

Empirical scoring functions estimate the binding energy of a ligand conformation in terms of physicochemical interactions such as hydrogen bonding, ionic, and hydrophobic interactions, calibrated against complexes of known affinity. Most of the empirical scores in use (e.g. FlexX [13], ChemScore [14]) today derive ultimately from the pioneering work of Bohm [15,16], as incorporated in the LUDI program [17]. Empirical functions generally perform well in binding mode prediction and hit identification (enrichment), but are less successful at accurately ranking active molecules by binding free energy.

Another approach is to define a docking scoring function in terms of the vdW and Coulombic potentials from a standard molecular mechanics force field (e.g. AutoDock [18], DOCK [19]). The potentials often are precalculated over a receptor grid for performance reasons, making use of the rigid receptor approximation. vdW grids can be also combined with empirical polar potentials to produce so-called semi-empirical scoring functions (e.g. GOLD [20], rDock [21,22], LigScore [23]).

The final scoring function category is that of knowledge-based potentials, exemplified by Potentials of Mean Force (PMFs). PMFs are calculated from observed atom-atom distribution functions across large training sets of protein-ligand structures (no binding affinities are needed). Examples include BLEEP [24], PLP [25], PMFScore [26] and DrugScore [27].

In incremental construction and fragment-based docking methods, the ligand is divided into a series of connected, smaller fragments prior to docking, with the aim of reducing the combinatorial explosion of possible ligand conformations to a more manageable level. The critical step is to identify an appropriate *anchor* fragment that can be correctly docked in the binding site, independently of the intact molecule. Clearly this step must be performed automatically and is not always successful. Subsequent fragments are positioned incrementally relative to the anchor fragment, with local (often systematic) conformational sampling at each stage. FlexX [28], DOCK 4.0 [29], Hammerhead [30] and LUDI [17] all implement variations on this approach.

In contrast, whole-molecule docking methods deal with ligand flexibility either via rigid docking of precomputed conformations, or via stochastic sampling of ligand dihedrals during the docking process. Rigid docking is attractive in principle as it reduces the docking problem to just six degrees of freedom (whole body translation and rotation), thus allowing systematic docking of each conformation. The main drawback is in the assumption that the appropriate binding conformations will be already present in the precomputed library, as there is little or no opportunity for induced fit of the ligand to the binding site. Examples of software that provide rigid docking protocols include EUDOC [31], FLOG [32], and the North Western University version of DOCK [33] (NWU-DOCK). LigandFit, from Accelrys [23], uses a shape discrepancy function to favour the generation of ligand conformations that are complementary to the shape of the docking cavity, prior to rigid body minimisation.

The highest level of docking accuracy in reproducing known ligand binding modes (see later) can be generally achieved with 'on-the-fly' flexible ligand docking, as this does not suffer from either of the problems highlighted above for incremental construction and rigid docking. Monte Carlo (MC), Genetic Algorithm (GA) and 'Tabu' methods are most commonly used to drive the stochastic search towards the global energy minimum. Examples include GOLD [34] (GA), AutoDock [35] (GA/MC), ICM [36] (MC), PRO_LEADS [37] (Tabu), Glide [38] (MC) and rDock [22] (GA/MC). Multiple independent docking runs per ligand are usually required to achieve convergence with a reasonable degree of confidence, particularly for flexible ligands, and this can be an issue in the context of virtual screening of large libraries.

We have developed rDock [21,22] as a modular, high performance virtual screening platform for protein and nucleic acid targets. Efficient docking is achieved through a combination of GA, MC and Simplex algorithms, semi-empirical scoring functions, and a progressive protocol that terminates poorly scoring ligands early in the search. External pharmacophoric, Cartesian and NMR distance restraints can be incorporated into the docking score to conduct more focused searches around experimentally observed or user-defined binding mode requirements. Artificial intelligence-derived post-docking filters have been evolved for the automated removal of likely false positives [39].

VALIDATION OF VS

By definition, the validation of a VS method can only come from comparing the outcome of two experiments carried out in parallel: a) a pure random screening of N compounds; and b) screening of a subset of these compounds selected with the VS method of choice. To date, such a blind test has never been published, and even when direct comparisons between random and rational screenings can be made (see references [40-42] and below), one could argue that it's not VS in general but the specific protocol used that is validated. Certainly, the outcome of a VS is

always the composite of the tools used (methodologies), how these tools have been used (user control) and the introduction of empirical information (subjective input). Hence the analysis of "real and dirty" experiments might not be appropriate for comparison of methodologies, but it is the ultimate test for any screening software. In order to avoid the subjectivity of real VS, several ideal experiments have been designed to assess the performance of docking tools. In the following sections the performance of several docking packages and protocols are compared by means of clean, artificial experiments, and real cases.

Reproduction of the Experimental Binding Mode

One of the main applications of molecular docking programs is the prediction of binding modes of known ligands in the absence of experimental information; accordingly, the so-called 'docking problem' has received much attention. Although in a docking-based VS it could be considered a secondary goal, the correct prediction of the binding mode is important for two main reasons: a) the underlying assumption in docking dictates that the "real" binding mode will have the best score, therefore a real binder will have a better chance of being selected if its binding mode is correctly predicted; and b) in real VS experiments, the predicted binding mode(s) are used to filter out compounds with unsatisfactory interactions with the receptor, which means that a real binder might be discarded if the binding mode is poorly predicted.

In most cases the docking programs have been tested with relatively small number of ligand-receptor complexes, but with the expansion of the PDB [43] bigger test sets are becoming the norm. Table 1 summarizes the published binding mode prediction experiments with large sets of structures [22,34,44-48]; the EUDOC publication [49] has been omitted from the table because it performs rigid docking of the ligand in the crystal structure conformation, whereas all the other programs perform a flexible-ligand rigid-receptor docking. LigandFit has been omitted also as, despite reporting very impressive results [23], the test set consists of just 19 protein-ligand complexes.

A direct comparison between software is hampered by the different size and composition of the test sets. This is illustrated by the different apparent performance of Surflex [48] and rDock [22], from 78% vs. 73% when comparing the full sets, to 77% vs. 82% when comparing only the 66 common structures. In that respect, the recent work by Nissink *et al.* [44] deserves particular consideration because it not only contains the largest set of structures (305), but also the so called "clean lists" of structures, which have been checked for diversity, structural errors and packing effects. Since this test set has been made freely available, it would be desirable for future studies to use it, thus allowing a straight comparison of the different docking packages.

In spite of the qualitative nature of any comparison, it is apparent from Table 1 that some programs, namely GOLD [34,44], Surflex [48] and rDock [22], provide satisfactory binding modes (less than 2Å root mean square deviation; RMSD) for around three quarters of the structures while a second group of programs provides satisfactory predictions in only half the cases. The predictions with RMSD lower

Table 1. Quality of Binding Mode Predictions of Different Docking Programs. As Noted in the Text, Comparisons Across Test Sets and Programs Should be Done Cautiously

Ref.	Program	Mode	N[a]	Success rate (%)	
				RMSD<1.0Å	RMSD<2.0Å
[34]	GOLD	Standard	100	35	66
[44]	GOLD	Standard	305	44	68
	GOLD	Standard	180[b]	51	73
[45]	FlexX	Standard	200	19	49
[46]	DOCK	Chemical Dock	200	7	14
	DOCK	Energy Dock	200	43	53
	DOCK	SCORE Dock	200	41	54
[47]		High Quality	103	31	46
[48]	Surflex	Standard	81	48	78
	Surflex	Standard	66[c]	48	77
[22]	rDock	Standard	157	34	70
	rDock	Solvation docking	157	39	73
	rDock	Solvation docking	66[c]	59	82

[a] Number of crystal structures included in the test set.

[b] Clean set with crystallographic resolution ≤ 2.5Å (recommended test set [44])

[c] Overlapping test sets between Surflex and rDock

than 1Å apparently are much more test-set dependent, as seen by the different GOLD (from 35% to 51%) and rDock sets (from 39% to 59%) and do not allow cross-comparisons. It is difficult to draw conclusions as to why some programs perform better than others, but we consider that the following two factors might be partially responsible: i) conformational search algorithm: all the best-performing programs allow for full flexibility of the ligands during docking, whereas FlexX and DOCK use incremental construction and the program in reference [47] does rigid docking followed by a ligand optimization step; ii) training set: the programs developed in recent years have benefited from larger and more diverse sets of structures to re-weight the different terms of the scoring functions. Based on the data available, there is no conclusive evidence to suggest that empirical scoring functions are necessarily better than force-field based ones. Interestingly, the best-performing programs have been developed in recent years and provide not only accuracy but also enough speed as to screen large virtual libraries in reasonable time (especially Surflex and rDock).

When docking is used as a VS tool, it is not strictly essential for the experimental binding mode to be the best scoring pose but, since the generated poses are often used during postfiltering (forcing interaction with a particular residue, for instance), it is important that the "true" binding mode is at least present amongst the various solutions. All

Table 2. Summary of Published Results with Seeded Libraries. As Noted in the Text, Comparisons Across Publications and Biological Targets Should be Done Cautiously

Ref.	Target	A	N	F(%)	C(%)	EF
[50]	Thrombin	32	1052	10	94	9
	Progesterone Receptor	28	1048	10	64	6
[51]	p38 MAP Kinase	256[a]	10256	0.8	9	11
	IMPDH	91[a]	10091	0.4[b]	13[b]	30[b]
	HIV protease	360[a]	10360	0.3[b]	10[b]	29[b]
[52]	Thrombin	32	1032	10	84	8
[53]	Thymidine Kinase	10	1000	2.9	70	24
	Estrogen Receptor	10	1000	1[c]	70[c]	70
[47]	Plasmepsin II	134	13020	3.8	13	3.3
[54]	T4 Lysozyme (L99A)	51	172118	0.1	14	137
	" (L99A/M102Q)	58	172118	0.1	17	172
[55]	D3 Receptor (antagonist)	10	1000	3.7	70	19
	M1 Receptor	10	1000	5.3	50	9
	V1a Receptor	10	1000	1.3	50	38
	D3 Receptor (agonist)	10	1000	1.8	60	33
	β2-adrenergic Receptor	10	1000	2.6	70	27
	δ-opioid Receptor	10	1000	3.3	90	27
[38]	COX2	128	7656	2	36[b]	18[b]
	Estrogen Receptor	55	7583	2	58[b]	29[b]
	p38 MAP Kinase	72	7600	2	25[b]	12[b]
	Gyrase B	36	7564	2	36[b]	18[b]
	Thrombin	67	7595	2	53[b]	26[b]
	Gelatinase A	43	7571	2	36[b]	18[b]
	Neuramidase	51	7579	2	82[b]	41[b]
[40]	Angiogenin	12	18111	2	42	21
[23]	Thymidine Kinase	9	1002	1	22	22
[48]	Thymidine Kinase	10	1000	1	50	50
	Estrogen Receptor	10	1000	1	80	80
[22]	Thymidine Kinase	10	1010	1	40	40
	Estrogen Receptor	10	1010	1	80	80
	Thrombin	19	1019	1	37	38

[a] Binders with 1.0 to 30.0 μM activity (see text)

[b] Data extracted from figures

[c] Estimated values

the publications report the presence of low RMSD poses (97%, rDock solvation [22]; 91%, Surflex [48]; 86%, Diller's program [47]; 91%, SCORE Dock [46]; 70%, FlexX [45]), which is not surprising given that in most cases an exhaustive docking has been performed. In fact, it would be much more informative to know if a low RMSD conformation is found within the high-scorers, unfortunately this is not always reported and comparison can not be made.

Nissink *et al.* [44] compared the success rate of GOLD across protein classes, concluding that the mean RMSD values for aspartic proteases and lyases are significantly higher, but this seems to be more related to the nature of the ligands (large and very flexible) than to the docking cavity itself. In fact, whenever reported, an inverse relationship between the accuracy of the results and the number of rotatable bonds (nRot) of the ligands is observed [45-48]; the impact of this tendency should be relatively small in VS if drug-like (nRot≤8 [5]) or lead like (nRot≤10 [6]) filters are applied upfront.

In conclusion, the performance of most of the programs and scoring functions appears good enough for VS purposes, although it would be advisable to consider not only the top scorer poses but also those energetically similar. It is expected that larger, more diverse and universally available test sets will drive future improvements of the scoring functions and allow critical comparisons across different methods.

Enrichment Factors

The ability of docking to perform VS can be assessed by rank ordering a library of random compounds (allegedly inactive) "seeded" with true binders or, even better, doing a retrospective analysis of a screened library. Recent examples of such experiments are summarized in Table 2 [22,38,40,47,48,50-55]. These "seeded libraries" have often been used to compare docking protocols and scoring functions. The results in Table 2 report the optimal solution for each macromolecular target, thus it represents the maximum performance that can be expected from current methods rather than standard degrees of success. The results are presented in terms of the following parameters: *F* is the Fraction of the library selected at the top of the ranked library; *C* is the Completeness, or the fraction of active compounds in *F* (both in percentage); *EF* is the Enrichment Factor. The corresponding formulae are:

$$F = \frac{n}{N} \text{ x } 100$$

$$C = \frac{a}{A} \text{ x } 100$$

$$EF = \frac{a/n}{A/N} = \frac{C}{F}$$

Where *N* is the total size of the library; *n* the number of compounds selected after screening; *A* the total number of active compounds; and *a* the number of true binders in the selection. Whenever possible, the top 1% of the library (*F*=1) or a similar fraction is used for the statistics; this would translate in the selection of one thousand compounds from a 0.1 million compound library, which is reasonably close to a typical VS experiment.

It is apparent from Table 2 that the results are very dependent on the macromolecular target. In fact, when the same true binders are used, the results are remarkably similar between different methods (e.g. thrombin [50,52]) and between different authors (e.g. thymidine kinase [22,48]; estrogen receptor [22,48,53]). This is in line with the well-known observation that some targets (or rather docking sites) are intrinsically more difficult than others and suggests that different optimal solutions share a common target-dependent limit of success. The data provided by Bissantz *et al.* [55] suggests that homology modelled docking sites can provide very acceptable results. One has to be careful when interpreting the results for the best-behaved receptors, because EF can be dependent on the size of the library. For example, if the random library is increased from 1 to 10 thousand compounds, performance of rDock with the estrogen receptor increases from EF=80 (F=1) to EF=700 (F=0.1) (Morley, S.D., personal communication [56]). This might partially explain the extraordinarily high EF observed with the Lysozyme mutants [54], a system with a very small binding site unable to accommodate the large ligands presumably found in the docking library (the Available Chemical Directory, ACD).

Another factor affecting the results is the nature of the active molecules. Charifson *et al.* [51] tackled this problem by dividing the active compounds into sub-libraries based on range of activity (<0.1μM; 0.1-1μM and 1-30μM). As expected the results are substantially better for libraries seeded with compounds that are more active. As the authors state, the last range of activities (1 to 30μM) are more representative of a real VS situation and the result for that set is reported in Table 2. Unfortunately it is uncommon for such a large and complete set of data to be publicly available, therefore different groups have used different data sets and comparisons across publications can not be made even when the same protein is studied (e.g. p38 [38,51], estrogen receptor [38,53], thrombin [38,50]).

A somewhat less characterized factor affecting the final results is the nature of the so called "random set". Fradera *et al.* [52] made the interesting observation that the quality of the predictions decreased with more drug-like random libraries. A possible explanation for this behaviour is that drugs have more chemical functionalities and, therefore, a higher probability of making favourable contacts with the binding site than non-drugs. The apparently poor performance of Diller's and Merz's program [47] can be explained by the nature of the test set. Both the active and inactive compounds are elaborated from the core of pepstatin (a binder of plasmepsinII). This is in agreement with previous observations [50] that docking cannot differentiate active compounds from structurally similar but inactive compounds.

Several papers [38,40,51,53,55] show comparisons of results obtained with different docking programs, the overall conclusion being that the scoring function has a far deeper impact than the software in the final performance. As

Table 3. Summary of Recently Published Docking-based VS. The Table is Provided to Allow Comparison of Protocols; as Noted in the Text, Success Rates Cannot be Compared

Ref	Target	Database	Cpds	Pre-filters	Cpds	Program	Protocol	Cpds	Post-filters	Cpds	Hit Def.	Hits
56	Thymidylate synthase	ACD	153516	None	153516	NWU-DOCK	Rigid docking	400	• Visual Inspection • Chemical tractability	5	IC_{50}<500μM	2
57	FKBP	ACD	n.a.	None	n.a.	SANDOCK	Rigid docking	n.a.	• Visual Inspection	n.a.	K_D<100μM	7
58	Protein Tyrosine Phosphatase 1B	ACD	150000	None	150000	DOCK3.5	Rigid docking	2000	• Visual Inspection • Diversity • Solubility • Chemical stability • Availability	25	K_i<65μM	5
59	HIV-1 Integrase	NCI	246182	• 5≤NHA≤35 • nRot≤10	n.a.	DOCK	Flexible docking in two stages	582	• Visual Inspection • Diversity • Solubility • Availability • QC	22	IC_{50}<200μM	13
60	Pyruvate Kinase	ACD	370000	• 10≤NHA≤35	n.a.	DOCK	Rigid docking with critical spheres	5698	• Visual Inspection (automated)	42	Measurable effect	14
61	TAR RNA	ACD	153000	None	153000	DOCK ICM	Three stages: rigid docking, ligand opt. and full opt.	350	• Score • Solubility • Non-toxic • Availability	8	Measurable effect	3
62	Farnesyl-transferase	ACD	219390	• q=0 • No zwitterions • ≥1 AR • 300≤MW≤700	67928	EUDOC	Rigid docking in two phases	313	• Solubility • Visual Inspection • QC	21	IC_{50}<100μM	4
63	Hypoxantine phosphoribosyl-transferase	ACD	150000	None	150000	NWU-DOCK	Rigid docking of conformational ensembles	500	• Visual Inspection	22	K_i≤17μM	16
64	Aldose Reductase	ACD	120000	None	120000	ADAM&EVE	Flexible docking forcing a minimum number of HB	718	• MW > 250 • ≥1 ring • Score • Visual Inspection	36	IC_{50}≤20μM	7
65	Bcl-2	NCI	206876	• nRot≤10 • 10≤NHA≤50	n.a.	DOCK	Flexible Docking	500	• Diversity • No peptides • No organomet. • Availability	35	IC_{50}<15μM	7
66	Retinoic Acid Receptor-α	ACD	150000	None	150000	ICM	Flexible docking followed by full opt.	300	• Visual Inspection	30	EC_{50}≤4μM	2
67	tRNA-guanine transglycosylase	ACD	120000	None	120000	LUDI	Flexible docking	6	None	6	Measurable activity	3
41	Dihydrodipico-linate reductase	Merck Chemical collection	n.a.	None	n.a.	FLOG	Rigid docking of pre-generated conformers	n.a.	None	n.a.	IC_{50}<100μM	n.a.
68	Matriptase	NCI	450000	• nRot≤10 • 10≤NHA≤50	n.a.	DOCK	Flexible docking in two stages	2000	• q≥1in solution	69	>95% I at 75μM	15

(Table 3). contd.....

Ref	Target	Database	Cpds	Pre-filters	Cpds	Program	Protocol	Cpds	Post-filters	Cpds	Hit Def.	Hits
30	Adenovirus proteinase	ACD	219390	• q=0 • No zwitterions • Non-peptidic • 300≤ MW≤700	67928	EUDOC	Rigid docking in two stages	30	• Score	3	Measurable activity	1
69	Aldose Reductase	NCI	127000	None	127000	DOCK	Rigid docking	1270	• Chemical clustering • Visual Inspection	25	IC_{50}<50μM	5
40	Protein Tyrosine Phosphatase 1B	• ACD • BioSpecs • Maybridge	235000	• 17≤NHA≤60	165581	NWU-DOCK	Rigid docking of conformational ensembles Chemical labelling	1000	• Diversity • Visual Inspection • Availability	365	IC_{50}<100μM	127
70	TAR RNA	ACD	181000	• MW<600 • -5≤q≤5 • nRot<20	128000	DOCK ICM	Three stages: rigid docking, ligand opt. and full opt.	500	• Score • Visual Inspection • Price • Availability	43	Measurable effect	11
71	AmpC β_Lactamase	ACD	229810	None	229810	NWU-DOCK	2 independent runs of rigid docking. Chemical labelling. HB artificially exaggerated	1000	• Visual Inspection	56	K_i<650μM	3
72	Carbonic Anhidrase	• Maybridge • LeadQuest	99027	• Rule of 5 [b] • Zn^{2+} binding groups • Pharmacophore • Similarity	100	FlexX	Flexible docking	100	• Score • Visual Inspection	13	IC_{50}<0.5μM	9
73	Thymidylate Synthase-Dihidrofolate Reductase	ACD	152571	None	152571	NWU-DOCK	Rigid docking of conformational ensembles Chemical labelling	400	• Visual Inspection	14	Significant activity	1
74	tRNA-guanine transglycosylase	• ACD • IBS • ChemStar • AEGC • LeadQuest • Ambinter	800000	• nRot < 8 • MW < 450 • ≥2 HB Donors • ≥1 HB Acceptors • ≥1 5-6MR • Pharmacophore • No duplicates	856	FlexX	Flexible docking Particle concept [c]	856	• Pharmacophore • Minimization • Visual Inspection	9	K_i<10μM	5
75	CDK2	Maybridge	50000	None	50000	LIDAEUS	Flexible docking	n.a.	• Score	120	IC_{50}<20μM	6
	CDK2	Maybridge	50000	None	50000	LIDAEUS	Flexible docking refined parameters	n.a.	• Score	28	>30% I at 30μM	8
[a]	Hsp90	rCAT	1400000	• 250<MW<550 • nRot≤6 • Chemical stability	700000	rDock	Flexible docking to 2 conformations of the protein in 3 stages	9000	• Score • Pharmacophore • Balance of polar/apolar interactions • Diversity • Availability	719	IC_{50}<10μM	7

a This work. b Reference [4] c Reference [76]

Abbreviations: Ref., Reference; Cpds, Compounds; Hit Def., Hit Definition; n.a., Not Available; nRot, number of rotatable bonds; MW, molecular weight; AR, Aromatic Ring; HB, Hydrogen-bond; NHA, Non-hydrogen atoms; 5-6MR, 5 or 6 memberd ring; q, electrostatic charge; QC, Quality Control.

expected [57], consensus scoring has repeatedly shown to provide the best EF [38,40,51,53,55], but given the trade-off between EF and completeness, this method tends to miss a large proportion of true binders.

One of the main conclusions drawn from all these studies is that it is not possible to know *a priori* the best docking protocol for a given site, thus it is highly advisable to construct "seeded libraries" and fine-tune the VS protocol whenever enough experimental data is available. On the other hand, if this is done properly, it is possible to identify one (or more) optimum protocols able to provide high enrichment factors (and hit rates) for most targets and libraries. The downside to this approach is that a significant number of true binders will be missed (false negatives). Although most of the best EF published range from 10 to 70, as we will see in the next section, reported EF in real cases are significantly better. Possible reasons for this will be discussed.

Real Cases

Table 3 summarizes the most recently published VS experiments [31,41,42,58-77]. In this section we will describe the different steps common to any VS campaign, referencing real applications.

Definition of the receptor

Once the biomolecule of pharmacological interest has been decided, the first step is to obtain the coordinates, generally by X-ray crystallography, but homology models have also been used [67,70]. Since most programs need a predefined volume to dock against, a subset of the biomolecule has to be selected at that stage. In most cases this is a straightforward step (e.g. the active site of an enzyme), but if the target is poorly characterized this decision will require thorough analysis of the structure and function of the target. This potential weakness of the docking methodology can become a strong point if the interest is to find binders to a specific site rather than general inhibitors [75].

The conformational flexibility of the biomolecule around the docking site has to be evaluated next. Docking programs can only consider the flexibility of the receptor to a limited extent and, even then, this comes with a major increase in computational cost, thus limiting its application to a small number of compounds [63,68,72,74]. With this limitation in mind, it is important to select a conformation of the docking site as representative as possible. This is generally done by using a structure with a bound ligand and, whenever possible, testing the definition of the cavity with other ligands for which the binding mode is known (cross-docking). When the protein is known to adopt very different conformations, the use of several representative structures would be advisable.

In nearly all cases, the resolution of the crystal structure is not high enough to identify the position of hydrogen atoms. Unless a united-atom model is used, hydrogen atoms will have to be added. Polar hydrogens are particularly important for hydrogen bonds and often require examination of the surrounding to determine the most likely position. The protonation state of ionisable residues should also be considered. Some programs allow conformational flexibility to polar hydrogens to guarantee the best possible fit with any ligand.

Ions, ligands and solvent molecules found in the crystal structure are generally stripped from the docking site, nonetheless it is important to detect interstitial waters that might play an important role for the molecular recognition of the ligands, as worse results are generally obtained when key water molecules are not considered [44]. Some water molecules can either be displaced or interact with a ligand. This dual behaviour can be simulated using the particle concept [78]. If cofactors or non-standard residues are found in the active site, it might be necessary to assign some parameters manually. Similarly, if particular atoms are known to be crucial for the interaction with the ligands, the parameters of these atoms could be artificially modulated [73] to reproduce the observed interaction affinity.

Preparation of the Small Molecule DataBase

Thanks to its ready accessibility, the ACD and the National Cancer Institute collection (NCI) are the most popular databases in the published approaches, but other commercial, in-house and virtual databases have also been used. Generally the libraries contain 2D representations of the molecules and have to be converted to 3D. Reference [79] provides a comparison of various 3D structure generators.

Whenever possible, undesirable compounds should be removed upfront to avoid wasting computational time. For this reason, drug-like [4] and other target-independent filters are commonly applied at that stage. These filters include descriptors such as molecular weight, presence of a given chemical moiety [74], number of rotatable bonds, rings or formal charges, as well as molecular properties (solubility, logP, etc).

Subject to the availability of ligand-based information or detailed analysis of the 'hot spots' of the binding site, receptor-specific filters based on pharmacophoric models [74,76] or similarity [74] can also be applied. In order to reduce significantly the number of compounds to be docked, the pharmacophore(s) will have to be fairly elaborate. Although very useful to reduce the initial number of compounds and focus on the potentially more interesting ones, these approaches have the drawbacks of introducing a labour-intensive extra step and having to combine several programs not necessarily compatible in terms of input-output format.

Either for efficiency or simplicity, some docking programs do not treat the flexibility of the ligands explicitly. In these cases, a library of conformers is usually generated before docking.

Docking

The references in Table 3 include a fairly diverse set of docking programs. Besides the differences in search and scoring protocols, the programs also differ in their ability to include additional constraints during the docking run. In certain cases, modifications of the potential has been used to force chemical complementarity [41,73] or occupancy [62] of certain regions of interest within the active site, to establish hydrogen bonds with a particular atom of the receptor [66] or to use the similarity to a ligand of reference [74]. These

constraints are a powerful mean of introducing empirical information during docking and can both speed-up the calculation and provide a better chance of success but, if not carefully chosen, the opposite effect might arise.

If the time scale needed to dock the full library is a realistic one, the most accurate docking protocol will be used. However, if this is computationally too expensive (either because the size of the library is too big or the software is not fast enough), alternative protocols can be constructed to vary the accuracy of docking as the calculations proceed (see Table 3 for examples).

Postfilter

Filters such as the ones described in preparation of the database can also be applied here. Although at this stage it does not save computational time, it allows visualization of the putative binding modes of the excluded compounds, the use of more accurate (i.e. computationally more expensive) methods to evaluate molecular properties, and a more informed decision.

Many of the published VS make extensive use of visual inspection of the docking output to select the final set of compounds. It is worth noting that, since only one (or a few) poses per ligand will be visualized, this step relies on the ability of the docking program to identify the "real" binding mode. For some programs the proportion of compounds correctly docked is only 50%, and even the best performers can only predict correctly up to 80% of the structures (see above). It would be advisable then to explore visually several binding poses. Either way the labour intensive character of this post-filter makes it unsuitable for large scale applications and it would be preferable to use automatable rules, which could be applied to all the poses and/or large collections of compounds.

Other post-filter criteria include the final score, a proper balance between polar and apolar terms, chemical tractability, diversity, solubility, and practical issues such as price, availability and quality control of the compounds.

Once the final selection of compounds has been made, the compounds have to be purchased, synthesized or picked out from the in-house library of compounds. Often this process results in an attrition rate of 10% to 50%, depending on the library.

Outcome andProgression

The ultimate goal of any screening campaign is to identify new chemical entities for progression, therefore, a VS is only successful if the hits are: 1) reasonably potent (generally below 10-30μM); 2) active through binding to the docked site (binding confirmed by competitive assays, NMR, X-ray, or other methods); 3) amenable to evolution in a medicinal chemistry program; 4) novel (i.e. have a favourable intellectual property position). Depending on the project, additional criteria will have to be met (e.g. antimicrobial activity for antibiotics). In short, the quality of the hits is a much better measure of success than the quantity. Most of the published works report only the number of hits, but in a few cases extra steps have been taken to confirm binding to the docking site, either by NMR [59,67,70] or X-ray crystallography [69,73,74,77]. In all but one case [77] the computationally predicted binding mode was in very good agreement with the experimentally determined one. Other authors report various degrees of hit progression [58,66,69,71] and even biological activity [61,64,65,67,72,73,75]. But overall it is not possible to know how many of the hits became leads or if the attrition rate is better for VS-generated hits than for the HTS-generated ones.

When considering only the number of hits, the reported values seem well above what one would expect at random, but just in five cases a control experiment is reported. Perola *et al.* [64] and Wu *et al.* [77] included a control group of the same size as the VS hit list (21 and 28 respectively). This can be useful to show that VS works better than random selection of compounds, but statistically it is not significant enough to calculate enrichment factors. Direct comparison between a HTS experiment and VS has only been previously reported in two cases [41,42] and we provide a third example (see below).

The Merck paper [42] does not provide the absolute number of compounds and hits, but give hit rates of 6% and ≤0.2% for VS and HTS respectively, resulting in a ~30-fold improvement. This comparison is very relevant because both HTS and VS screened the Merck chemical collection (although different subsets of it) and used exactly the same experimental protocol. Interestingly this EF is within the range reported with seeded libraries (see Table 2), as should be expected from a VS protocol which relies solely on docking.

The work by Shoichet's group [41] does not provide such a clean comparison between HTS and VS because 1) different libraries were screened, the one used for HTS (Pharmacia's in-house library) not being very drug-like; 2) the assay conditions used for VS hits were more permissive than the ones used for HTS; and 3) the VS hits were screened using a medium-throughput assay probably more accurate and sensitive than the HTS. As a result, the authors reported a value of 1700 as an upper estimate of the EF.

In spite of the limitations stated above, it seems clear that the EFs reported by Doman *et al.* [41] and here, as well as the hit rates of most of the VS experiments in Table 3, are well above reasonable expectations for pure docking VS. In our opinion this is due to several factors: i) pre-filters applied to the commercial libraries to guarantee a minimum drug-like character or appropriate chemical composition of the docking library; ii) use of empirical information during or after docking (e.g. chemical labelling of the docking site, visual inspection, fulfilment of a pharmacophore) to bias the results; iii) the intrinsic value of low- or medium-throughput screening assays, generally more accurate that HTS. Obviously, the goal of a VS experiment is not to prove the methodology but to obtain good hits. Accordingly, an inspection of Table 3 shows that it is common practice to assist and modify the docking protocols with all sorts of complements, although this blurs the contribution of each of the components.

Finally, we would like to emphasize that the different works referenced in Table 3 cannot be compared in term of hit rates, because different targets have different hit propensities and because the hit rate is heavily dependent on the screening method and the definition of hit (ranging from "measurable activity" to <500nM in the examples provided).

Moreover, given the small number of compounds assayed in most of the reported VS, the hit rates are not statistically very significant.

VS AGAINST HSP90, AN INDUSTRIAL EXAMPLE

Most of the published examples of VS involve very labour intensive processes and the selection of a very limited number of compounds. In an industrial environment, where the pressure is on the rapid identification of new hits, more automated VS protocols are necessary and a greater number of compounds are typically assayed. In this section we exemplify this with a VS campaign against Hsp90 done in our company.

Definition of the Receptor

The identification of small ligands of Hsp90's ATP binding site through our proprietary SeeDs technology [80], allowed the design of enrichment experiments to identify the definition of the cavity yielding best results. This included 3 interstitial water molecules consistently found around Oδ2 of Asp93.

Following the finding that the conformation of the 110-115 loop can be altered significantly upon binding of PU3 [81], two structures of Hsp90 were used: 1YET [82] and an in-house structure of the PU3-Hsp90 complex [81].

Preparation of the Small Molecule Database

rCat is our proprietary catalogue of 3.5 million compounds. A docking library of 0.7 million compounds was selected from rCat based on: i) crude drug-like filters (molecular weight 250-550Da, and six or less rotatable bonds); ii) removal of reactive groups (a list of unstable chemical moieties was compiled based on chemical expertise and substructural searches were performed to clean the docking library of molecules containing reactive groups); and iii) vendor delivery timelines.

These pre-filters were carried out using Isis/Base [83] and MOE [84]. The program CORINA [85], version 2.63, was used to convert the docking library from 2D to 3D. The same program was used to generate multiple ring conformations; the internal energy threshold was set to 7 KJ/mol. This allows the treatment of cyclic groups as rigid bodies. The total number of docked conformers was 1.7 million.

Docking

rDock [21] was used to screen the docking library against Hsp90 using two different crystal structures: Geldanamycin-bound [82] (PDB entry 1YET) and PU3-bound Hsp90 [81]. Binding to PU3 produces a very significant conformational change of the backbone of the protein in the region of the active site and it was deemed necessary to consider both conformations. rDock was used for docking in its high-throughput mode [22]. 4300 compounds were selected using cavity 1 and 6000 using cavity 2, totalling approximately 9000 non-redundant compounds.

Postfilter

Those compounds binding almost exclusively through polar or apolar interactions were removed to guarantee an adequate balance of these terms, in agreement with the composition of the targeted site. All the Hsp90 inhibitors for which a structure is available [82,86,87] form a hydrogen bond with one carboxylic oxygen of Asp93 and accept a second hydrogen bond from an interstitial water molecule, as shown in Fig. (**1**). Using a Perl script created in our lab, the binding poses not satisfying this donor-acceptor motif were discarded.

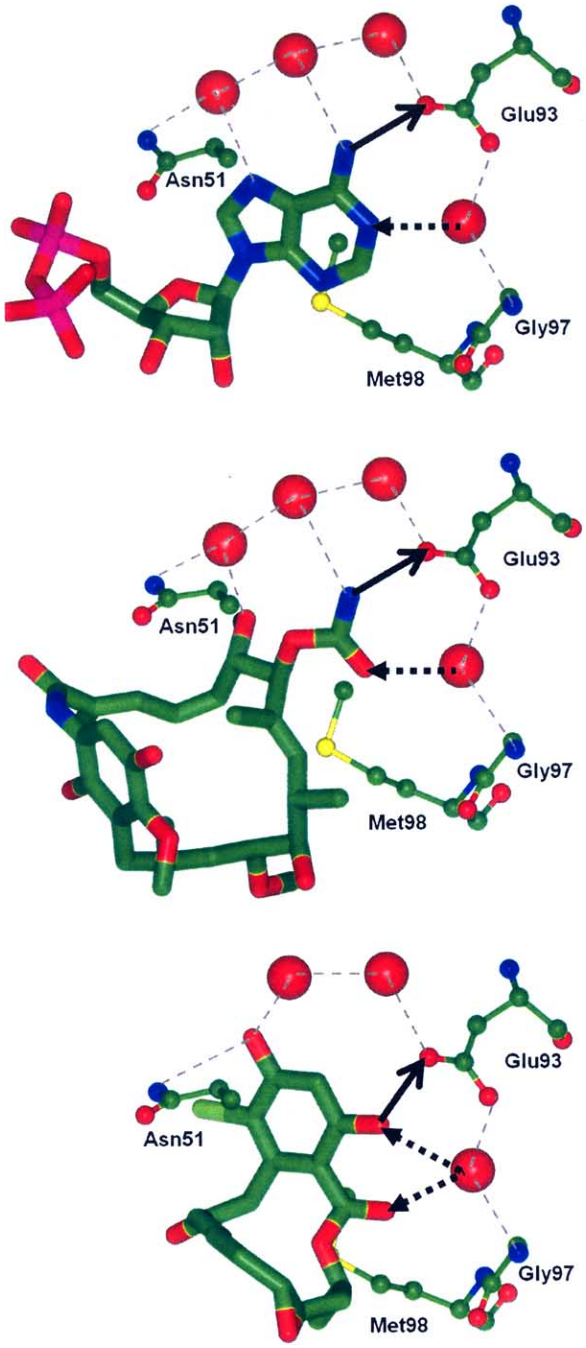

Fig. (1). Binding mode of ADP (top), Geldanamycin (middle) and radicicol (bottom) to Hsp90 (PDB entries 1BYQ [87], 1YET [82] and 1BGQ [86], respectively). These three binders display a common hydrogen-bond donor/acceptor motif, represented by solid (donor) and dashed (acceptor) lines.

The remaining compounds were clustered in chemical families to assess their diversity. Over-represented families

were purged with MOE [84] based on chemical diversity. The top 1000 remaining scorers were selected for purchase. Of those, 719 compounds were actually available and assayed.

Outcome and Progression

Out of 719 compounds assayed, those inhibiting the ATPase activity of Hsp90 by more than 50% were selected for titration measurements, providing 13 compounds with IC_{50}<100μM (1.8% hit rate) and 7 with IC_{50}<10μM (1.0% hit rate). Interestingly, two hits are closely related to a compound identified by HTS by a research partner (Workman, P., oral presentation at the 94th annual meeting of the American Association for Cancer Research, abstract published in reference [88]) and show very similar activity. The hit rate in the HTS, using a very similar assay protocol, was approximately 0.002%. Crystallographic studies confirmed that these compounds share a common binding mode. This series of compounds have progressed [89] and currently are in late lead optimisation. The rest of the hits cluster in 5 chemical families, all of which have been confirmed to bind to the docking site by NMR or X-ray crystallography. These compounds have provided back-up series.

Summary

The HSP90 results show that rDock is capable of rapidly identifying high quality, tractable hits from large compound libraries. This is ultimately a more important and practical validation of the platform for industrial VS than the idealised test results for docking accuracy and seeded libraries reported above. Whilst it is natural to focus on the overall assay hit rate as a measure of success, a much more meaningful metric is the quality, diversity and novelty of the hit compounds identified, as this will dictate the ability of VS to properly fuel chemistry lead optimisation programs.

CHALLENGES IN STRUCTURE-BASED VIRTUAL SCREENING

The evolution of docking programs, together with the steady increase of computational power and the universal access to chemical collections have resulted in real applications of docking-based VS both in academic and industrial environments. We have described some success stories, but molecular docking can and needs to be improved further to establish itself as a cornerstone of drug discovery. In our opinion, the following points will have a major impact in the future of docking-based VS.

Inclusion of the Effect of the Solvent

Water, the most abundant component of living organism, has a profound effect on biochemical systems. Particularly in the process of molecular association, both the ligand and the receptor have to be partially desolvated in order to form a complex. A wide range of theoretical methods have been developed to account for the solvent effect [90], but only a handful are fast enough to be of use for docking. Given that in empirical and knowledge-based scoring functions the effect of solvent is implicitly accounted for, the groups using force-field based scoring functions have been more actively experimenting with methods to include the effect of solvent [35,54,91,92]. Recently rDock has combined an empirical scoring function with a solvation term based on the change of solvent accessible surface area [22]. In spite of the use of rough approximations, these studies confirm that explicit consideration of the effect of the solvent provides improved results and is the way forward.

Flexibility of the Receptor

The treatment of the binding site of the macromolecular target as a rigid body is an oversimplification used in the vast majority of molecular docking applications. Moreover, in most cases only one structure of the receptor is used. The quality of this approximation will depend very much on the particular flexibility of the docking site and is considered to be responsible for the poor results obtained with specific macromolecular targets (see Table 2). This well-known limitation has stimulated the development of strategies to take into account the flexibility of the receptor, which have been reviewed recently [93-95]. So far, no consensus has been reached as for which methods will provide the best results but, given the extent of ligand-induced changes in the binding sites of proteins [96], it is expected that it will have a major impact on the quality of docking results.

Scoring Functions

In spite of its evolution, the scoring functions continue to be one of the factors limiting the success of VS. This limitation is more apparent in the hit-to-lead and lead optimisation phases of drug discovery, when libraries are designed around common scaffolds and the interaction energy gap between the best inhibitors and the rest of the compounds is relatively small. It is expected that as larger sets of structures and binding data become available (through initiatives as the Ligand-Protein Database [97], for example), empirical and knowledge-based scoring functions will continue to improve. Force-field based scoring functions, on the other hand, are expected to improve via application of "first principles" to incorporate as many terms of the binding free energy as possible (solvation, internal energy of the receptor, entropy). One-window free energy grids represent a very interesting example of this tendency [98,99].

As a medicinal chemistry project evolves, more and more inhibitors of the therapeutic target become available. This provides the possibility to create training sets, which can be used to check the performance and/or to optimize the scoring function. In the first case, several scoring functions can be tested on their own or in combination (consensus scoring [51]) to find which one performs best. Going one step beyond, the combination of structural information with QSAR tools allows the scoring functions to be tailored to the specific site of interest [100-102].

Integration with other Drug Discovery Technologies

Experiments with seeded libraries (see above) demonstrate that, in real applications, docking can provide up to 30-fold enrichment factors in the first half percent of the virtual library [51]. This means that, in the best scenario,

if an initial library of one million compounds is docked, approximately 5000 compounds would have to be assayed and only 10-15% of the real hits would actually be found, which is likely to translate into relatively low hit rates. Future developments of docking will certainly improve these figures but, to make VS a real alternative to HTS, docking has to be integrated with other computational techniques able to make use of empirical information (e.g. similarity, pharmacophoric constraints, substructural searches, etc.). This is illustrated by the real VS examples provided in Table 3. New programs, integrating docking with similarity [48,52], pharmacophoric restraints [103,104] or other tools might not be conceptually ground-breaking, but are likely to prove themselves extremely useful.

One of the pitfalls of HTS is the hit attrition rate, due to low quality hits. Thanks to its *in-silico* nature, VS is ideally suited to avoid this problem by means of "quality filters" applied upfront in order to guarantee an adequate profile of the docking library. This should include not only solubility, permeability and chemical stability, but also good chemical tractability to facilitate a rapid progression in the subsequent phases of drug discovery. Even if these prefilters imply lower hit rates, the parallel optimization of all these parameters since the early stages will accelerate the whole process.

CONCLUSION

Both the validation tests and real VS experiments have shown the capacity of docking tools to predict the binding mode of a ligand to its receptor and to gather most of the binders at the top of a ranked library of compounds. Nevertheless, it is also accepted that the greatest degree of success is obtained when docking-based VS benefits from target-dependant "fine-tuning", which requires experimental information and expert user input.

The low cost of VS is one of its main appeals, but it provides other benefits, such as more flexibility, faster timelines and better integration with the discovery platform, but the decision of going into the VS or the HTS route will ultimately be very dependant on the particular circumstances of the project and the research organization. While big pharmaceutical companies with HTS facilities in place might prefer to explore every single compound to avoid VS false negatives, smaller companies and academic institutions might be able to benefit from the lower cost, better flexibility and faster timelines associated with VS.

LIST OF ABBREVIATIONS

VS	=	Virtual Screening
HTS	=	high-throughput screening
RMSD	=	root-mean square deviation
nRot	=	number of rotatable bonds
EF	=	enrichment factor
ACD	=	Available Chemical Directory
NCI	=	National Cancer Institute
NWU-DOCK	=	North Western University version of DOCK
PMF	=	Potential of Mean Force

REFERENCES

[1] Fox, S.; Wang, H.; Sopchak, L.; Farr-Jones, S. *J. Biomol. Screen.* **2002**, *7*, 313-6.
[2] Mestres, J. *Biochem. Soc. Trans.* **2002**, *30*, 797-9.
[3] Bajorath, J. *Nat. Rev. Drug Discov.* **2002**, *1*, 882-94.
[4] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3-26.
[5] Oprea, T. I. *J. Comput. Aided Mol. Des.* **2000**, *14*, 251-64.
[6] Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308-15.
[7] Stewart, L.; Clark, R.; Behnke, C. *Drug Discov. Today* **2002**, *7*, 187-96.
[8] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269-88.
[9] Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *28*, 28.
[10] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151-66.
[11] Glen, R. C.; Allen, S. C. *Curr. Med. Chem.* **2003**, *10*, 763-7.
[12] Böhm, H.-J.; Stahl, M. In *Reviews in Computational Chemistry*; Boyd, K. B. L. a. D. B., Ed.; Wiley-VCH: Hoboken, NJ, **2002**; Vol. *18*.
[13] Rarey, M.; Wefing, S.; Lengauer, T. *J. Comput. Aided Mol. Des.* **1996**, *10*, 41-54.
[14] Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425-45.
[15] Bohm, H. J. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243-56.
[16] Bohm, H. J. *J. Comput. Aided Mol. Des.* **1998**, *12*, 309-23.
[17] Bohm, H. J. *J. Comput. Aided Mol. Des.* **1992**, *6*, 61-78.
[18] Goodsell, D. S.; Olson, A. J. *Proteins* **1990**, *8*, 195-202.
[19] Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 505-524.
[20] Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43-53.
[21] Morley, S. D.; Afshar, M. accepted, **2004**.
[22] Morley, S. D.; Juhos, S.; Garmendia-Doval, A. B. *in preparation* **2004**.
[23] Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graph Model* **2003**, *21*, 289-307.
[24] Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. *J. Comput. Chem.* **1999**, *20*, 1165-1176.
[25] Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317-24.
[26] Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791-804.
[27] Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337-56.
[28] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470-89.
[29] Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411-28.
[30] Welch, W.; Ruppert, J.; Jain, A. N. *Chem. Biol.* **1996**, *3*, 449-62.
[31] Pang, Y. P.; Xu, K.; Kollmeyer, T. M.; Perola, E.; McGrath, W. J.; Green, D. T.; Mangel, W. F. *FEBS Lett.* **2001**, *502*, 93-7.
[32] Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. *J. Comput. Aided Mol. Des.* **1994**, *8*, 153-74.
[33] Lorber, D. M.; Shoichet, B. K. *Protein Sci.* **1998**, *7*, 938-50.
[34] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727-48.
[35] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
[36] Totrov, M.; Abagyan, R. *Proteins* **1997**, *Suppl. 1*, 215-20.
[37] Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins* **1998**, *33*, 367-82.
[38] Schulz-Gasch, T.; Stahl, M. *J. Mol. Model. (Online)* **2003**, *9*, 47-57.
[39] Garmendia-Doval, A. B.; Morley, S. D.; Juhos, S. *Proceedings of the 6th International Conference on Artificial Evolution*, Marseilles (France), **2003**.
[40] Jenkins, J. L.; Kao, R. Y.; Shapiro, R. *Proteins* **2003**, *50*, 81-93.
[41] 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*, 2213-21.

[42] 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.
[43] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235-42.
[44] Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. *Proteins* **2002**, *49*, 457-71.
[45] Kramer, B.; Rarey, M.; Lengauer, T. *Proteins* **1999**, *37*, 228-41.
[46] Tao, P.; Lai, L. *J. Comput. Aided Mol. Des.* **2001**, *15*, 429-46.
[47] Diller, D. J.; Merz, K. M. Jr. *Proteins* **2001**, *43*, 113-24.
[48] Jain, A. N. *J. Med. Chem.* **2003**, *46*, 499-511.
[49] Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. *J. Comput. Chem.* **2001**, *22*, 1750-1771.
[50] Knegtel, R. M.; Wagener, M. *Proteins* **1999**, *37*, 334-45.
[51] Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.* **1999**, *42*, 5100-9.
[52] Fradera, X.; Knegtel, R. M.; Mestres, J. *Proteins* **2000**, *40*, 623-36.
[53] Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759-67.
[54] Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. *J. Mol. Biol.* **2002**, *322*, 339-55.
[55] Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. *Proteins* **2003**, *50*, 5-25.
[56] Morley, S.D. personal communication.
[57] Wang, R.; Wang, S. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422-6.
[58] Tondi, D.; Slomczynska, U.; Costi, M. P.; Watterson, D. M.; Ghelli, S.; Shoichet, B. K. *Chem. Biol.* **1999**, *6*, 319-31.
[59] Burkhard, P.; Hommel, U.; Sanner, M.; Walkinshaw, M. D. *J. Mol. Biol.* **1999**, *287*, 853-8.
[60] Sarmiento, M.; Wu, L.; Keng, Y. F.; Song, L.; Luo, Z.; Huang, Z.; Wu, G. Z.; Yuan, A. K.; Zhang, Z. Y. *J. Med. Chem.* **2000**, *43*, 146-55.
[61] Chen, I. J.; Neamati, N.; Nicklaus, M. C.; Orr, A.; Anderson, L.; Barchi, J. J., Jr.; Kelley, J. A.; Pommier, Y.; MacKerell, A. D. Jr. *Bioorg. Med. Chem.* **2000**, *8*, 2385-98.
[62] Bond, C. J.; Jurica, M. S.; Mesecar, A.; Stoddard, B. L. *Biochemistry* **2000**, *39*, 15333-43.
[63] Filikov, A. V.; Mohan, V.; Vickers, T. A.; Griffey, R. H.; Cook, P. D.; Abagyan, R. A.; James, T. L. *J. Comput. Aided Mol. Des.* **2000**, *14*, 593-610.
[64] Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y. P. *J. Med. Chem.* **2000**, *43*, 401-8.
[65] Freymann, D. M.; Wenck, M. A.; Engel, J. C.; Feng, J.; Focia, P. J.; Eakin, A. E.; Craig, S. P. *Chem. Biol.* **2000**, *7*, 957-68.
[66] Iwata, Y.; Naito, S.; Itai, A.; Miyamoto, S. *Drug Des. Discov.* **2001**, *17*, 349-59.
[67] 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*, 4313-24.
[68] Schapira, M.; Raaka, B. M.; Samuels, H. H.; Abagyan, R. *BMC Struct. Biol.* **2001**, *1*, 1.
[69] Gradler, U.; Gerber, H. D.; Goodenough-Lashua, D. M.; Garcia, G. A.; Ficner, R.; Reuter, K.; Stubbs, M. T.; Klebe, G. *J. Mol. Biol.* **2001**, *306*, 455-67.
[70] Enyedy, I. J.; Lee, S. L.; Kuo, A. H.; Dickson, R. B.; Lin, C. Y.; Wang, S. *J. Med. Chem.* **2001**, *44*, 1349-55.
[71] Rastelli, G.; Ferrari, A. M.; Costantino, L.; Gamberini, M. C. *Bioorg. Med. Chem.* **2002**, *10*, 1437-50.
[72] Lind, K. E.; Du, Z.; Fujinaga, K.; Peterlin, B. M.; James, T. L. *Chem. Biol.* **2002**, *9*, 185-93.
[73] Powers, R. A.; Morandi, F.; Shoichet, B. K. *Structure (Camb.)* **2002**, *10*, 1013-23.
[74] Gruneberg, S.; Stubbs, M. T.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 3588-602.
[75] Atreya, C. E.; Johnson, E. F.; Irwin, J. J.; Dow, A.; Massimine, K. M.; Coppens, I.; Stempliuk, V.; Beverley, S.; Joiner, K. A.; Shoichet, B. K.; Anderson, K. S. *J. Biol. Chem.* **2003**, *29*, 29.
[76] Brenk, R.; Naerum, L.; Gradler, U.; Gerber, H. D.; Garcia, G. A.; Reuter, K.; Stubbs, M. T.; Klebe, G. *J. Med. Chem.* **2003**, *46*, 1133-43.
[77] Wu, S. Y.; McNae, I.; Kontopidis, G.; McClue, S. J.; McInnes, C.; Stewart, K. J.; Wang, S.; Zheleva, D. I.; Marriage, H.; Lane, D. P.; Taylor, P.; Fischer, P. M.; Walkinshaw, M. D. *Structure (Camb.)* **2003**, *11*, 399-410.
[78] Rarey, M.; Kramer, B.; Lengauer, T. *Proteins* **1999**, *34*, 17-28.
[79] Sadowski, J.; Gasteiger, J.; Klebe, G. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000-1008.
[80] Knowles, D. *Curr. Drug Discov.* **2002**, 31-35.
[81] Dymock, B.; Barril, X.; Beswick, M.; Collier, A.; Davies, N.; Drysdale, M.; Fink, A.; Fromont, C.; Hubbard, R. E.; Massey, A.; Surgenor, A.; Wright, L. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 325-8.
[82] Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P. *Cell* **1997**, *89*, 239-50.
[83] MDL Information Systems, Inc.
[84] Chemical Computing Group, Inc.
[85] Gasteiger, J.; Rudolph, C.; Sadowski, J. *Tetrahedron Comp. Method* **1990**, *3*, 537-547.
[86] Roe, S. M.; Prodromou, C.; O'Brien, R.; Ladbury, J. E.; Piper, P. W.; Pearl, L. H. *J. Med. Chem.* **1999**, *42*, 260-6.
[87] Obermann, W. M.; Sondermann, H.; Russo, A. A.; Pavletich, N. P.; Hartl, F. U. *J. Cell. Biol.* **1998**, *143*, 901-10.
[88] Aherne, W.; Maloney, A.; Sharp, S.; Clarke, P.; Walton, M.; Hardcastle, A.; Boxall, K.; Richards, J.; Rowlands, M.; Newbatt, Y.; Raynaud, F.; Smith, N.; Nutley, B.; McDonald, E.; James, K.; Cheung, J.; Matthews, T.; Roe, M.; Prodromou, C.; Pearl, L.; Workman, P. *Proc. Amer. Assoc. Cancer Res.* **2003**, *44* (2nd ed.), 4002.
[89] Drysdale, M. J.; Dymock, B. W.; Barril-Alonso, X.; Workman, P.; Pearl, L. H.; Prodromou, C.; MacDonald, E. *PCT Int. Appl.* WO2003055860, **2003**.
[90] Orozco, M.; Luque, F. J. *Chem. Rev.* **2001**, *101*, 203-4.
[91] Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins* **1999**, *34*, 4-16.
[92] Zou, X.; Sun, Y.; Kuntz, I. D. *J. Am. Chem. Soc.* **1999**, *121*, 8033-8043.
[93] Carlson, H. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447-52.
[94] Carlson, H. A. *Curr. Pharm. Des.* **2002**, *8*, 1571-8.
[95] Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213-8.
[96] Fradera, X.; De La Cruz, X.; Silva, C. H.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Bioinformatics* **2002**, *18*, 939-48.
[97] Roche, O.; Kiyama, R.; Brooks, C. L. 3rd. *J. Med. Chem.* **2001**, *44*, 3592-8.
[98] Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 502-11.
[99] Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 3417-23.
[100] Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1995**, *38*, 2681-91.
[101] Wang, T.; Wade, R. C. *J. Med. Chem.* **2002**, *45*, 4828-37.
[102] Gohlke, H.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 4153-70.
[103] Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. *J. Comput. Aided Mol. Des.* **2002**, *16*, 129-49.
[104] Joseph-McCarthy, D.; Thomas, B. E.; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. *Proteins* **2003**, *51*, 172-88.