British Journal of Pharmacology (2008) 153, S55–S67
 0007–1188/08 $30.00
www.brjpharmacol.org

REVIEW

Ligand discovery and virtual screening using the program LIDAEUS

P Taylor, E Blackburn, YG Sheng, S Harding, K-Y Hsin, D Kan, S Shave and MD Walkinshaw

The Centre for Translational and Chemical Biology, The University of Edinburgh, Michael Swann Building, King's Buildings, Mayfield Road, Edinburgh, UK

This paper discusses advances in docking and scoring approaches with examples from the high-throughput virtual screening program LIDAEUS. We describe the discovery of small molecule inhibitors for the immunophilin CypA, the cyclin-dependent kinase CDK2 and the cyclapolin series of potent Polo-like kinase inhibitors. These results are discussed in the context of advances in massively parallel computing and in the development of annotated databases.
British Journal of Pharmacology (2008) **153,** S55–S67; doi:10.1038/sj.bjp.0707532; published online 26 November 2007

Keywords: virtual screening; LIDAEUS; EDULISS; cyclophilin; cyclin-dependent kinase; Polo-like kinase

Abbreviations: CDK, cyclin-dependent kinase; CypA, human cyclophilin-A; CLogP, the octanol-water partition coefficient, calculated using the Biobyte program (http://biobyte.com.index.html) developed by Hansch and Leo; EDULISS, EDinburgh University LIgand Selection System; LIDAEUS, LIgand Discovery at Edinburgh UniverSity; MlogP, the octanol-water partition coefficient calculated as described by Moriguchi *et al*. (1992); RMSD, root mean square distance

Virtual screening overview: tools and approaches

Ligand discovery can be regarded as a simple matching problem: we would like to find a small molecule (ligand) with the appropriate shape and charge properties to bind effectively to a target protein of interest. High-throughput screening (HTS) provides one possible experimental route to a solution and libraries consisting of over 1 million compounds can be tested in days. Computational screening provides a complementary approach and with massively parallel processing, millions of compounds per week can be tested. Estimates of the number of potential small molecule drug-like compounds vary between 10^{18} and beyond 10^{63} (Lipinski and Hopkins, 2004). Consequently, for any specific target protein, even if the results from each assay and each docking run were totally reliable (which is not the case), it would still be impossible to test binding for every potential ligand. The commonly accepted Lipinski criteria (Lipinski *et al*., 1997) for orally active drug-like molecules set physicochemical property limits to increase the probability of good drug bioavailability. Drug-like molecules are expected to have a molecular weight (MW) $\leqslant$500 Da, $\leqslant$5 hydrogen bond donors (HBDs), $\leqslant$10 hydrogen bond acceptors (HBAs) and a CLog P $\leqslant$5 (the octanol-water partition coefficient calculated as described by Moriguchi *et al*. (1992) (MLogP) $\leqslant$4.15). More stringent criteria have been proposed for initial searches. For example, Lead likeness restricts MW to $<$350 Da and CLogP (the octanol-water partition coefficient, calculated using the Biobyte program (http://biobyte.com.index.html) developed by Hansch and Leo) to $<$3 (Teague *et al*., 1999). Even these more stringent cutoffs do little to reduce the astronomical numbers of potential ligands and exploring such a large-scale-matching problem will require imaginative computational and experimental approaches.

Protein targets

Recent reviews have attempted to estimate the number of druggable proteins in the Protein Data Bank (PDB) (Berman *et al*., 2000). Druggable proteins have structural features that facilitate binding to drug-like molecules. For proteins to progress from intrinsic druggability to becoming a target requires drug binding to modulate the biological role of the protein and to bring about therapeutic benefit (Fishman and Porter, 2005). Currently available literature identifies 1300 studied protein drug targets from humans and infective

Correspondence: Professor MD Walkinshaw, The Centre for Translational and Chemical Biology, The University of Edinburgh, Michael Swann Building, King's Buildings, Mayfield Road, Edinburgh, EH9 3JR, UK.
E-mail: m.walkinshaw@ed.ac.uk
Received 29 July 2007; revised 27 September 2007; accepted 4 October 2007; published online 26 November 2007

organisms (Hopkins and Groom, 2002; Russ and Lampel, 2005; Zheng *et al*., 2006). Estimates of the total number of druggable targets in the human genome have been made based on the number of disease genes; these give a total of up to 1500 targets out of 25 000 in the human genome (Hopkins and Groom, 2002). Bacterial and viral proteins also provide targets; published estimates of the number of targets from infective organisms are well over 1000. This suggests that there should be a pool of about 3000 drug targets in total (Zheng *et al*., 2006).

Of the 1300 currently studied targets, 44% are classified as enzymes, the most populated biochemical class. A total of 557 enzymes are current research targets and 134 have proved to be successful targets. Enzymes represent 50% of all successful targets (Zheng *et al*., 2006). A total of 280 research targets have experimentally determined structures with a specific drug-binding domain (represented by 107-folds), mainly by X-ray crystallography.

Within the PDB, there are about 250 uniquely different (that is $<10\%$ amino acid identity) well-determined structures in complex with 'peptide-ligands'. These represent a subset of protein–protein interactions where the interaction is controlled by a linear peptide on one side of the interface. Table 1 shows some examples of protein–peptide interactions. This group possibly represents the most druggable subset of protein–protein interactions. Short linear peptides are more amenable to replacement by small molecule mimetics. Modulating protein–protein interactions is particularly attractive due to the pivotal role of such interactions in cell signal transduction pathways and cell cycle progression (Fry and Vassilev, 2005; Chene, 2006).

There are a number of publicly accessible sources of protein–ligand binding affinities and web-based tools designed to aid the extraction of information from databases containing structural information on protein targets. For example, the BindingDB is a public, web-accessible database of measured binding affinities for biomolecules and contains data generated by isothermal titration calorimetry and enzyme inhibition (http://www.bindingdb.org/). The Relibase database (http://relibase.ebi.ac.uk/) is a web-based tool for the study of protein–ligand interaction. MSDmotif (http://www.ebi.ac.uk/msd-srv/msdmotif/) provides a tool for summarizing structural information on a database of over 6000 protein–ligand complexes found in the PDB.

Small molecule databases

A number of publicly available small molecule databases have been established over the last few years. The ligand.Info database (http://ligand.info) (Grotthuss *et al*., 2004) is a compilation of a number of publicly available databases providing a Meta-Database of over 1 million entries with calculated three-dimensional (3D) structures and some information about biological activity. The ZINC database (http://blaster.docking.org/zinc/) contains over 4.6 million commercially available compounds in various 2D and 3D formats (Irwin and Shoichet, 2005). Only compounds with MW $\leqslant$700 Da, and calculated LogP values between –4 and 6 are stored. Simple Lipinski filters or other discreet subsets of compounds can be selected.

An ambitious project financed by the National Institutes of Health has the goal of discovering sets of molecules that will specifically modulate the activities of the majority of gene product in the human and other organisms. Fast expanding databases are now being developed that contain results from a number of high-throughput screens, many of which use a set of over 100 000 chemically diverse molecules. These data are available at NCBI's database of small organic molecules at http://www.ncbi.nlm.nih.gov/sites/entrez?db=pcassay.

EDULISS, the EDinburgh University Ligand Selection System, is our in-house relational database that stores over 5 million available compounds, containing data from over 25 chemical catalogues. Of the 5.3 million compounds, 3.8 million are unique. 3D coordinates for each molecule are stored with over 1500 topological, geometric, physicochemical and toxicological descriptors per compound (Todeschini and Consonni, 2005). The descriptors can be used interactively to select subgroups of the database and also to provide profiling information. One approach to identify unique compounds is to compare the chemical graph of each compound with the graph of every other. This approach is extremely computationally expensive. An alternative method for identifying unique compounds in EDULISS's large collection has been developed. A small number of descriptors including a 3D-Wiener index, an electronegativity descriptor and a polarizability descriptor are used to group compounds. The resulting small groups of molecules with identical descriptors can then be compared using a

Table 1 Examples of protein-peptide interactions in the Protein Data Bank

PDB	*Protein*	*Peptide*	*Peptide sequence*	*Function*	*Reference*
1YCR	MDM2	p53	SQETFSDLWKLLPEN	Antitumour	(Vassilev *et al*., 2004)
1BXL (NMR)	Bcl-XL	Bak-BH3	GQVQRQLAIIGDDINR	Apoptosis	(Degterev *et al*., 2001)
1EBA	EPO	EPOR	GGTXSCHFGPLTWVCKPQGG	Anaemia	(Qureshi *et al*., 1999)
1EJ4, 1WKW	EiF4e	EiF4e-BP	RIIYDRKFLMECRN	Malignant transformation	(de and Graff, 2004)
1AXC	PCNA	p21	GRKRRQTSMTDFYHSKRRLIFS	Antitumour	(Gulbis *et al*., 1996)
1CKA	c-CRK	C3G	PPPALPPKKR	Oncogene	(Wu *et al*., 1995)
1GUX	Rb tumour suppressor	E7 peptide	DLYCYEQLN	Antitumour	(Lee *et al*., 1998)
1H9O	SH2	Penta -peptide	XVPML	Signal transduction, cancer	(Pauptit *et al*., 2001)
1QZ2	FKBP52	Hsp90	MEEVD	Steroid signalling pathways	(Wu *et al*., 2004)
1ELR	HOP	Hsp90	XMEEVD	Signalling pathways	(Scheufler *et al*., 2000)
1YVH	c-CBL	APS	GRARAVENQXSFY	Oncogene	(Hu and Hubbard, 2005)
1G3F (NMR)	XIAP-Bir3	Smac	AVPIAQKSE	Apoptosis	(Liu *et al*., 2000)

graph-matching program. A web-based interface for EDULISS has been developed; this provides a convenient way of extracting families of compounds with a user-defined set of properties.

Database profiling and compound selection

The EDULISS database comprises 25 different commercial and other smaller specialist compound collections. Of these, some 4.3 million fit the Lipinski 'rule of 5s' (Lipinski *et al*., 1997). A total of 3.2 million fit the Oprea lead-like criteria (Hann and Oprea, 2004). The more stringent Astex Rule of 3 is met by 230 000 compounds (Carr *et al*., 2005) (statistical profiles of some general descriptors are shown in Figure 1, descriptor ranges are shown in Table 2). A study by Oprea *et al*. (2007) investigated recent trends in the property space of leads, drugs and chemical probes. Leads are generally smaller, less complex and have lower LogP than drugs, due to the inevitable modifications involved in the medicinal chemistry optimization process.

It is desirable for a set of compounds for docking or assay to be selected considering both protein target and screening methodology. Solubility is of key importance for both bioavailability and 'screenability'. Experimentally derived aqueous solubility data are not available for the majority of compounds in the EDULISS database. Algorithms for predicting aqueous solubility from structure almost universally rely on a directly proportional relationship between LogP and solubility (Jorgensen and Duffy, 2002; Delaney, 2005). It might be appropriate to have a relaxed solubility requirement (MLogP $\leqslant$4.21) and to include relatively large compounds ($\leqslant$450 Da) with the aim of finding leads of high affinity and high specificity for the target. A greater range of molecular complexity may be explored with a higher MW cutoff (Schuffenhauer *et al*., 2006). However, the application of X-ray crystallography in lead discovery has different property requirements. The technique identifies fragments binding to significant regions of the target protein and then employs fragment growing or linking strategies to improve potency. Fragments are small molecules, 100–250 Da, with few functional groups (Rees *et al*., 2004; Carr *et al*., 2005; Hartshorn *et al*., 2005). In these techniques, virtual hit ligands are soaked into crystals. Relative protein concentrations are high, necessitating high ligand concentrations. Solubility problems can be compounded by the practice of soaking with a fragment cocktail to increase assay throughput. Virtual screening subsets designed for fragment screens

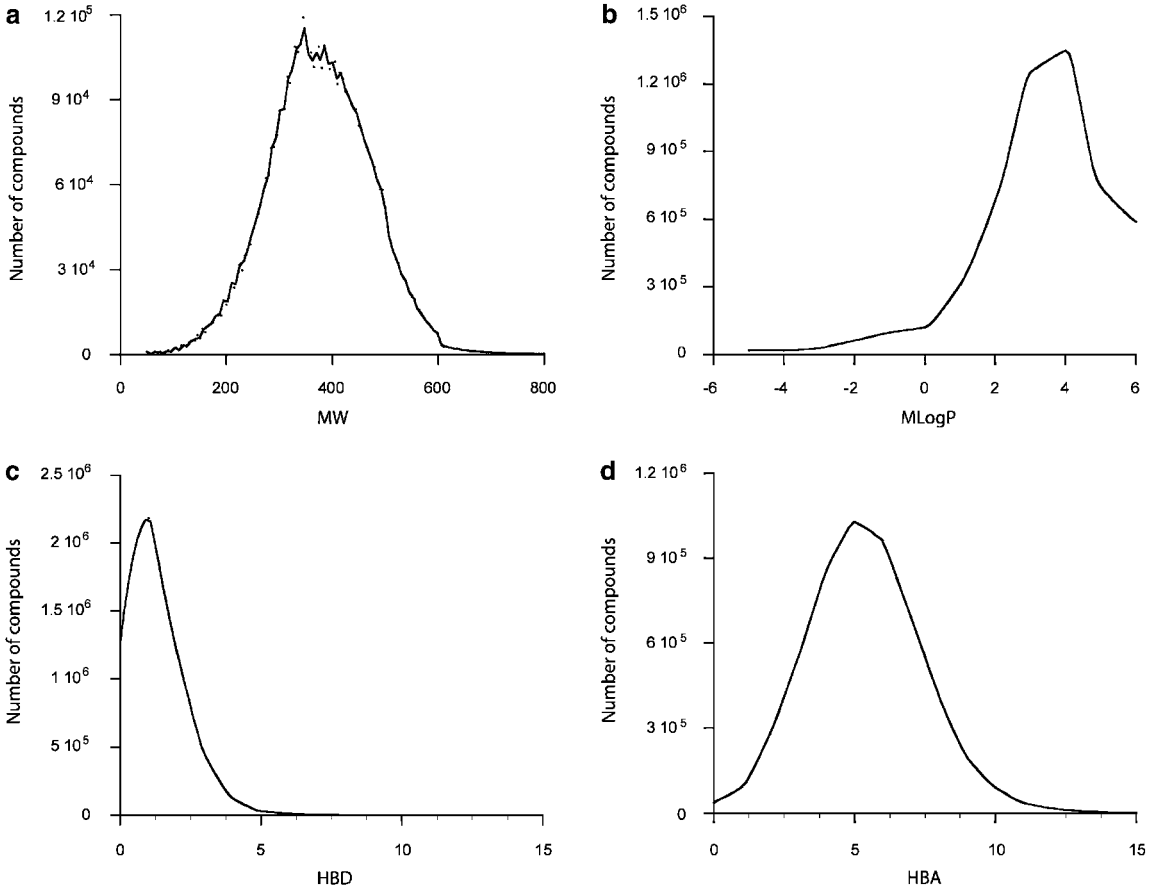


Figure 1 Molecular property profiles of 5.3 million compounds in the EDULISS database. (**a**) MW. (**b**) MLogP. (**c**) Number of HBDs. (**d**) Number of HBAs. Bin sizes for MWs are 5 Da and for MLogP, the number of HBDs and HBAs 1 U. EDULISS, EDinburgh University LIgand Selection System; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; MLogP, the octanol-water partition coefficient calculated as described by Moriguchi *et al*. (1992); MW, molecular weight

Table 2 Descriptor ranges for the EDULISS database of 5.3 million compounds

Descriptor	*Max*	*Average*	*Standard deviation*
Molecule weight	2180.59	374.28	95.77
Number of bonds	306	47.05	13.05
Number of aromatic bonds	69	13.42	5.97
Number of rings	18	3.14	1.18
Sum of atomic van der Waals volumes[a] (Å^3)	172.82	29.29	7.71
Number of rotatable bonds	77	5.31	2.73
MLogP	134.05	3.29	3.35
Topological polar surface area[b] (Å^2)	932.34	73.66	42.23

Abbreviations: EDULISS; LIgand Discovery at Edinburgh UniverSity; MLogP, the octanol-water partition coefficient calculated as described by Moriguchi (Moriguchi *et al*., 1992).
[a]Scaled on Carbon atom.
[b]Using N, O, S, P polar contributions (Todeschini and Consonni, 2005).

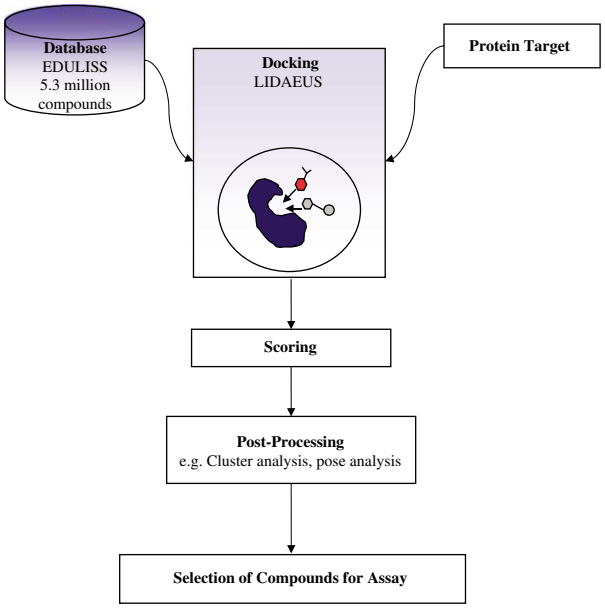


Figure 2 Steps involved in virtual screening using LIDAEUS. LIDAEUS, LIgand Discovery at Edinburgh UniverSity.

need to have stringent solubility requirements, MLogP $\leqslant 3.0$, while containing diverse scaffolds decorated with a broad range of functional groups (Moriguchi *et al*., 1992).

High-throughput virtual screening

High-throughput virtual screening achieves a high throughput of test ligands by using simplified non-quantum mechanical methods without the inclusion of complex molecular dynamics (Woo and Roux, 2005). Typically, the virtual screening process follows the steps outlined in Figure 2. A ligand is selected and positioned into the target protein-binding pocket in a given 'pose' (Muegge and Martin, 1999). The resultant complex is scored on the basis of intermolecular contacts to give a predicted strength of binding interactions (Woo and Roux, 2005). Flexible docking typically allows sampling of ligand and sometimes protein conformations during the docking procedure. Rigid body docking is however much less computationally expensive. Exploring the conformers of relatively simple molecules containing only three or four rotatable bonds (using a broad step size) requires the generation of over 200 starting conformations to be sampled in order to fully consider the majority conformational space (Guner *et al*., 2004). The most widely used flexible docking tools are GOLD (Genetic Optimization for Ligand Docking) (Jones *et al*., 1997), FlexX (Rarey *et al*., 1996; Kramer *et al*., 1999), DOCK (Ewing *et al*., 2001), AutoDock (Goodsell *et al*., 1996), Glide (Friesner *et al*., 2004; Halgren *et al*., 2004) and ICM (Internal Coordinate Mechanics) (Abagyan *et al*., 1994). A variety of different methods are used by the above tools to deal with ligand flexibility such as genetic algorithms, incremental construction, simulated annealing and Monte Carlo methods (Rosenfeld *et al*., 1995; Vieth *et al*., 1998). The diversity exhibited by scoring functions has been used in consensus scoring is implemented in, for example, X-SCORE (Wang *et al*., 2003). Using different but well-performing scoring functions, the accuracy of consensus methods can be greater than any individual scheme (Bissantz *et al*., 2000; Stahl and Rarey, 2001; Jacobsson *et al*., 2003; Raymond *et al*., 2004; Xing *et al*., 2004; Feher, 2006). However, 'artificial enrichment' is a potential pitfall, with scoring functions selected to perform well on a specific protein–ligand complex (Verdonk *et al*., 2004).

Perola *et al*. (2004) have reported that energy minimization can significantly improve the accuracy of docking poses found by GOLD (Jones *et al*., 1997) and ICM (Abagyan *et al*., 1994) programs. Our results also showed that there is better agreement between the docked pose and the crystallographic pose using rigid body refinement. A 'good fit' is defined as an root mean square distance (RMSD) $\leqslant 2$ Å between corresponding heavy atoms of the X-ray structure and the docked ligand pose.

Virtual screening has proved successful in a number of projects (Alvarez, 2004; Kitchen *et al*., 2004; Oprea and Matter, 2004; Ghosh *et al*., 2006) (Table 3). One of the major future challenges is to develop virtual screening methods capable of identifying ligands that will interrupt protein–protein interactions.

LIDAEUS as a tool for high-throughput virtual screening

LIDAEUS, LIgand Discovery at Edinburgh UniverSity, our in-house high-throughput virtual screening program (Wu *et al*., 2003) generates a grid of site points in the binding pocket of the target protein. Each site point is coloured: HBA, HBD or hydrophobic, depending on the preferred protein interaction (Figure 3).

Each molecule selected from the small molecule database is placed in the binding pocket and atoms of the docked molecule are matched to site points. An exhaustive fit of a given number of atoms from the docked molecule onto the site points is undertaken to identify reasonable poses. These are stored for subsequent rigid body energy minimization.

Table 3 Examples of hits from virtual screening experiments

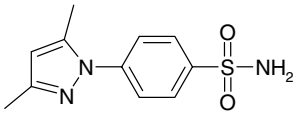
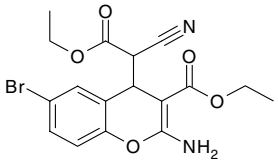
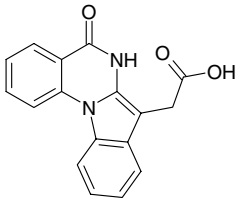
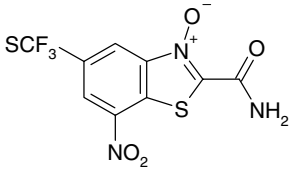
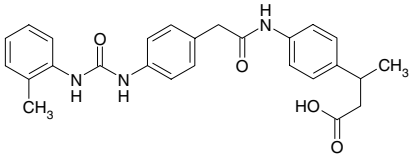
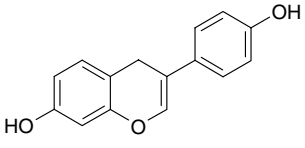
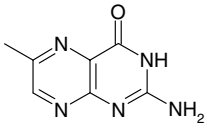
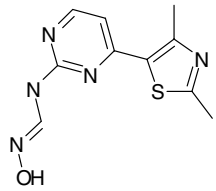
Target	*Virtual screening*	*Example structure**	*Reference*	*Assay*
Carbonic anhydrase II	FlexX (Rarey *et al.*, 1996)		(Gruneberg *et al.*, 2002)	IC_{50} Sub nanomolar
Bcl-2	DOCK (Ewing *et al.*, 2001)		(Wang *et al.*, 2000; Enyedy *et al.*, 2001)	IC_{50} 9 μM
CK2	DOCK (Ewing *et al.*, 2001)	* 	(Vangrevelinghe *et al.*, 2003)	IC_{50} 80 nM
Plk1	LIDAEUS (Wu *et al.*, 2003)		(McInnes *et al.*, 2006)	IC_{50} 20 nM
GPCR	Five targets 5-HT1A, 5-HT4, Dopamine D2, NK1, and CCR3	Compound PRX-93009 scored best for 5-HT1A, no structure shown	(Becker *et al.*, 2004)	Ki 1.0 nM
Integrin α4β1	Catalyst (Greene *et al.*, 1994)	* 	(Singh *et al.*, 2002)	IC_{50} 1.3 nM
ERβ	GOLD 2.0 (Jones *et al.*, 1997)		(Zhao and Brinton, 2005)	IC_{50} 0.68 μM
TGT	Unity/FlexX (Rarey *et al.*, 1996)		(Brenk *et al.*, 2003)	IC_{50} 0.25 μM

Table 3 *Continued*

Target	*Virtual screening*	*Example structure**	*Reference*	*Assay*
CDK2	LIDAEUS (Wu *et al*., 2003)		(Wu *et al*., 2003)	IC_{50} 2.2 μM

Structures marked with an asterisk do not represent those initially identified by *in silico* screening. Minor chemical modifications have been made and from these compounds the experimental data determined.

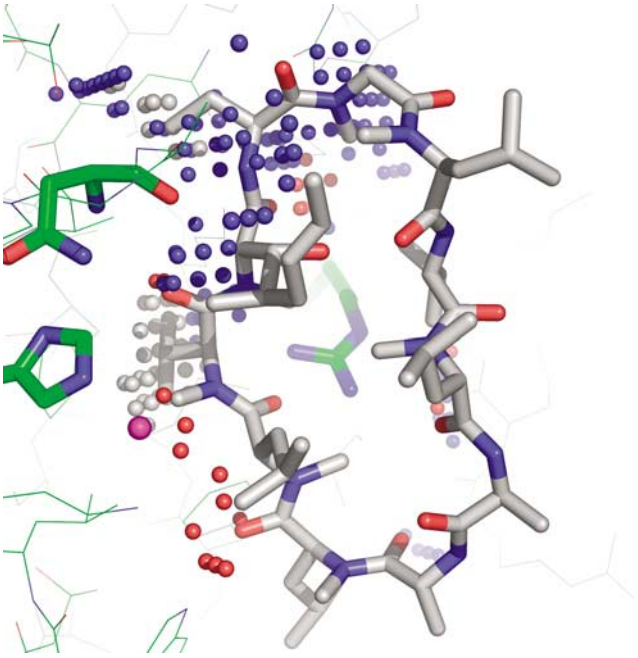


Figure 3 LIDAEUS site points in the binding pocket of CypA in complex with cyclosporine-A (PDB code 1cwa). Each site point is coloured depending on the preferred protein interaction (HBA, red; HBD, blue; hydrophobic, grey). The magenta sphere represents a key water molecule. Key residues are shown in bold. CypA, human cyclophilin-A; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; LIDAEUS, LIgand Discovery at Edinburgh UniverSity; PDB, Protein Data Bank.

There are various tunable parameters that influence how thoroughly the binding pocket is explored and hence the time required to dock a series of compounds. The precision with which an atom matches the site point, called 'resolution' in this program, is usually set between 0.02 and 0.06 Å and plays an important role in determining the number of allowed starting poses. Resolution values greater than 0.06 Å lead to an exponential growth in the number of starting poses. Increasing the number and density of the site points has a similar effect of dramatically increasing the number of allowed starting poses.

In LIDAEUS, there are two built-in scoring functions, a force field-based energy function and pose interaction profile (PIP) a knowledge-based function (Kan, 2007). The energy function is essentially a linear combination of van der Waals and hydrogen bonding energies. The geometry-dependent hydrogen bonding term incorporates salt bridges and obviates the need for calculating hydrogen atom positions. The program assigns fixed formal charges to identify ionizable groups.

The PIP score uses a protein interaction profile that can be assigned for specific regions of the binding pocket where explicit types of ligand interactions, for example, a particular hydrogen bond, are required. The PIP string is a hexadecimal code containing information about the interactions made between a given set of protein residues and the docked ligand. The target PIP string is usually based on a known X-ray crystal structure in which key features of the protein–ligand binding interaction have been identified. It is a very efficient process to match and score the bit strings of the target interactions against those calculated for the docked ligand pose. Thus, the PIP score can be used as a component of the final score to ensure that docked ligands have both favourable energies of interaction and satisfy specific interactions in their pose.

While LIDAEUS is broadly similar in function to many docking programs, it differs in two major areas: the extent to which the fitting protocol can be modified by the user and the modularity of the system. All definitions within the program in the way of atom and site point types are soft, that is, can be customized by the user. This happens at two levels: one can initially type individual atoms according to connectivity criteria and then group many or one of these types into colours used in the pose generation or scoring process. While using the default definitions allows for standard searches using atoms grouped into broad classes such as hydrophobic, HBA and HBD, it is possible to add specific restrictions.

LIDAEUS exists as a series of modules that run as a UNIX pipeline, so that initial typing of molecules, posing, scoring and sorting are all separate programs. It allows us to easily develop experimental modules and test different scoring methods. The program is being developed in two areas. The current flexible docking module is too slow to be used in a high-throughput mode and this is being addressed. Secondly, a front end is being written that allows intermediate users the ability to easily configure customizable features.

The examples discussed in the paper were run on a modest seven-node cluster. Run times are dependent on the target protein, the ligand complexity and the site point resolution set for LIDAEUS. However, representative times for a database of 50 000 ligands would be 6 h. Using an IBM Blue Gene/L supercomputer, run times have been reduced from 8 days on a seven-node cluster to 62 min on 1024 processors using a standard data set of 1.67 million small molecules.

Validation of LIDAEUS docking and scoring performance

The immunophilin proteins FKBP (FK506 Binding Protein) and human cyclophilin-A (CypA) have been used as test systems to develop and test the results from the database mining program LIDAEUS. Despite having similar biological profiles, the structure and active site of the two proteins are very different. Both proteins have peptidyl-prolyl isomerase activity and speed up the *cis–trans* equilibration of proline residues by lowering the barrier to rotation about the imide bond (Fischer *et al.*, 1993). Inhibition of the enzymatic turnover of an immunophilin substrate provides a functional assay for screening potential inhibitors (Fischer *et al.*, 1984). X-ray crystallographic, surface plasmon resonance, isothermal titration calorimetry and mass spectrometry results provide complementary techniques for characterizing ligand binding.

A set of nine chemically related ligands of human CypA with IC_{50} values between 2 and 100 μM were used to test LIDAEUS docking performance (referred to as the test set). For each ligand, the X-ray structure is known and the RMSD between corresponding atoms of the X-ray structure and the ligand structure is used as a measure of fit. Correct docking poses (RMSD $\leqslant 2$ Å) of ligand 1 in the test set (Figure 4 shows the correlation of PIP score and energy score were E, with RMSD from X-ray structure for a ligand in the test set) were all scored >0.93 by the PIP function and their energy scores, $E < -11$ kcal mol^{-1}. PIP scores are normalized between 0 and 1: a high PIP score indicates conserved interactions between those in the X-ray structure and the docked pose. The other ligands in the CypA test set have similar results, showing that a combined total score of energy function and PIP (matching a defined pose iinteraction profile) ranks the correct docking binding mode higher than alternative poses (Equation 1).

$$S = W_1 E + W_2 \sum_i PIP_i \quad (1)$$

S, total score; *E*, force field-based energy score; *PIP*, knowledge-based PIP score; W_1, weighting factor specific to protein system; W_2, weighting factor specific to protein system.

For a set of nine related cyclophilin inhibitors, the effect of changing the weights W_1 and W_2 was examined by systematically trialling different values. For this series of compounds, the values that gave the best RMSD fit of the docked pose compared to the crystallographic pose were weights of 1 and -40 for W_1 and W_2, respectively. These values proved useful in identifying a new series of cyclophilin ligands (Kan, 2007).

The role of water in accurate docking

It has been reported in several studies that water-mediated protein–ligand interactions are an important factor in the docking process. Ligands can displace water in the active site or incorporate them as an extension of the protein surface. The presence of key water molecules can significantly improve docking performance (Pospisil *et al.*, 2002; Yang and Chen, 2004). Our results show that eight out of nine ligands in the test set were correctly docked into near-native positions by LIDAEUS, while re-docking of six of them was significantly improved when key water molecules were included in the protein–ligand binding system. The presence of the key waters enables the LIDAEUS program to identify several important interactions involved in the complex and construct the significant HBA or HBD site points at the binding atom locations. (Key waters are those that form bridging H bonds to both protein and ligand molecules). Ligand 7 (Figure 5) is an example where including water improved docking performance. The presence of the key waters enables the LIDAEUS program to construct the significant HBA or HBD site points at the binding atom

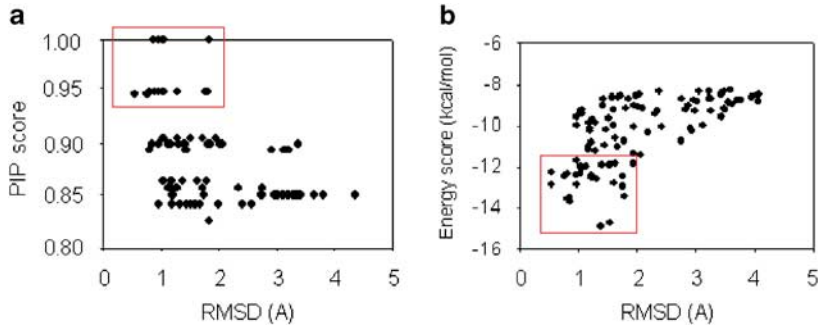


Figure 4 Correlation of PIP score and energy score, E, with RMSD from X-ray structure for ligand 1 of the CypA test set. (**a**) Plot of PIP score against RMSD of docking poses with respect to X-ray structure. (**b**) Plot of energy score, E, against RMSD with respect to X-ray structure. Red boxes highlight 'good poses' that meet scoring cutoffs: PIP score >0.93, energy score <-11 kcal mol^{-1}. CypA, human cyclophilin-A; PIP, pose interaction profile; RMSD, root mean square distance.

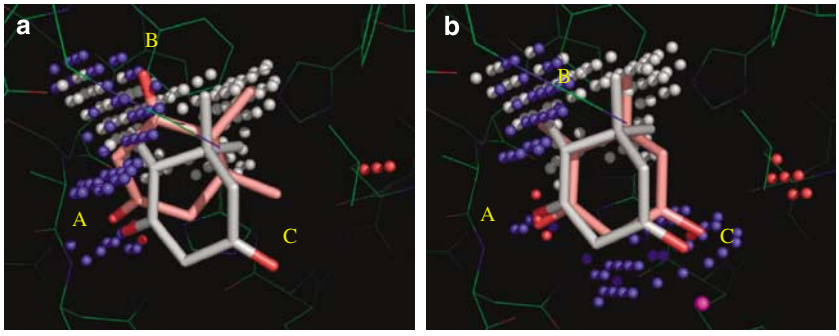


Figure 5 Including water in site point generation. Poses were improved when an essential water molecule was used in the calculation of energy maps and in site point construction. X-ray structure of the ligand is shown in white and docked poses in pink. (**a**) Illustrates how one oxygen atom (O1) from the ligand was put into the experimental position (position A), but the other oxygen atom (O2) was put into position B instead of position C, as revealed from the X-ray structure. (**b**) When the important water molecule (in magenta) is included in site point generation, competent site points set helped bring the oxygen atom to position C.

locations. Furthermore, energy maps contoured with the presence of key waters better represent the energy distribution in the local area. Thus, those maps would give the correct fits lower energy scores than maps generated without key waters.

Discovery of ligands for immunophilins

In a test to find CypA ligands, the ZINC database (Irwin and Shoichet, 2005) of 2 million compounds was used as an input for a LIDAEUS screen looking for compounds that would match site points in the active site of CypA (two parallel runs using 60 site points with a resolution of 0.06 Å and 170 site points with a resolution of 0.04 Å). The top 2000 poses were re-ranked, specifying that hydrophobic interactions with Phe113 and a hydrogen bond to Arg55 were satisfied, using PIP scoring (Figure 6).

The combined energy and PIP scores ranged from −164 to −80 (arbitrary units). The top 360 unique compounds were grouped according to chemical similarity (using molecular fingerprinting and Tanimoto coefficients) and binding mode (visually using Pymol). From this analysis, 14 compounds (all chemically distinct from known cyclophilin inhibitors) were purchased and tested for inhibition and binding by peptidyl-prolyl isomerase assay (Kofron *et al*., 1991). Eleven compounds showed a statistically significant reduction in peptidyl-prolyl isomerase activity. Six of the 14 compounds were 'hits' in the peptidyl-prolyl isomerase enzymatic assay; they inhibited CypA with IC_{50} values ranging from 27 to 135 μM. Subsequent isothermal titration calorimetry studies for the best three compounds gave K_d values of 2 to 8 μM.

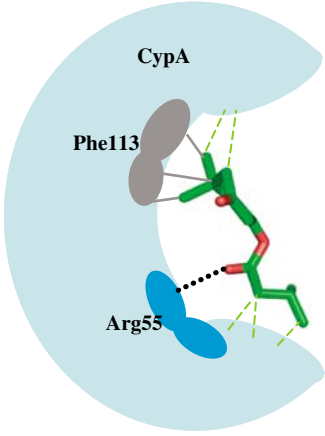


Figure 6 PIP used in the CypA virtual screening experiment. The top 2000 poses (rank ordered using the energy score, E,) were re-ranked, specifying that there were predicted hydrophobic interactions with Phe113 (grey lines) and a predicted hydrogen bond to Arg55 (black dotted lines) using PIP scoring. The diagram shows a molecule making interactions specified in the PIP. The green dashed lines are non-covalent interactions not specified in the PIP. CypA, human cyclophilin-A; PIP, pose interaction profile;

Virtual screening for CDK inhibitors using LIDAEUS

Cyclin-dependent kinases (CDKs) are key regulators in all steps of the cell cycle and as such are interesting targets for anticancer therapies. There are already a number of clinical trials underway with CDK2 and CDK4 inhibitors for a range of cancers (Collins and Garrett, 2005). The small molecule inhibitors, roscovotine (Seliciclib) and flavopiridol, are CDK2 inhibitors and show promising activity in lung cancer. These drugs target the ATP binding site of the CDKs. This is a problem in the design of CDK selective drugs, as all nine CDKs show some homology and most of the active site residues are well conserved. Another complicating factor in the design of specific inhibitors is that the active form of the kinase is induced by complex formation with a partner cyclin and phosphorylation of a specific threonine residue located on the T-loop of the kinase. These events cause subtle changes in active site geometry, which may be important for inhibitor design.

We used LIDAEUS to carry out a virtual screen of 50 000 commercially available compounds from the Maybridge catalogue (www.maybridge.com) docked into the active site of CDK2 (taken from the X-ray structure of the CDK2–staurosporine complex; PDB code 1AQ1). The predicted top 120 poses based on the docking score were screened at a fixed concentration of 30 μM using an assay to monitor the

inhibition of phosphorylation by CDK2/cyclinE. Twenty-nine percent of the compounds were classed as active by showing more than 30% inhibition. The most active four compounds all had a heteroaryl-2-amino-pyrimidine core and measured IC_{50} values between 0.9 and 17 μM (Table 4). X-ray crystal structures of the four hits were obtained and each was clearly identified in the ATP binding site. A comparison was made of the calculated docked pose (without any PIP influence) against the experimentally determined ligand structures. The four ligands were all found to dock in twisted conformations with a twist of 35° around the bond between the two aromatic rings. The RMSD atom against atom fit of the three top scoring docked ligands versus the experimental structure were 1.6, 1.58 and 3.42 Å with scores of −24, −23 and −20 kcal mol^{-1}, respectively. Despite the chemical similarity of these four ligands, they adopt different binding modes (Table 4) CYC1 and CYC2 form identical hydrogen-bond interactions to ATP: NH….O (Glu81), N….HN(Leu83) and CH….O(Leu83). When the amine group is substituted as in CYC3 and CYC4, the ATP binding mode is precluded and the ligand flips over to allow the bulky substituent to point out of the pocket. An alternative hydrogen bonding pattern is made CH…O (Glu81), N….HN(Leu83) and NH….O(Leu83). These four structures provided an excellent starting point for the design of chemical modifications. Over 40 related structures have been synthesized to optimize *in vivo* potency. The tightest binding ligand of this series, an amino derivative (CYC5), has a $K_i = 2$ nM and was shown to induce cell death in cultured HeLa cells (Wang *et al*., 2004a, b)

The importance of fine tuning a template structure in virtual screening

The shape and surface of the target pocket is clearly one of the most important factors in successful virtual screening runs. The search for CDK2-specific inhibitors highlighted the importance of understanding the biological role of the target protein. A crystallographic study was used to analyse the structures of six inhibitor ligands belonging to the thiazole-pyrimidine class, identified by LIDAEUS; both in complex with monomeric CDK2 and also with the binary CDK2/cyclinA active complex (Wu *et al*., 2003; Kontopidis *et al*., 2006). The activation of CDK2 by phosphorylation and cyclin binding causes significant loop and helix movements but leaves the shape of the ATP binding site relatively unchanged with a maximum side-chain movement between 1 and 2 Å for residues comprising this pocket. However, these small differences in pocket shape play a major role in the relative binding strengths of inhibitors. In some cases, the same ligands can adopt significantly different poses in the monomeric and active complexes. Binding enthalpies of the ligands have been estimated based on calculated van der Waals and hydrogen bond contacts measured in the crystal. The measured IC_{50} values correlate well with the calculated interaction energy (energy score) for the binary complex, but show poor correlation with the inactive complex. This fits with the way the assay has been carried out—using the active complex. It also suggests that the enthalpic energy-scoring scheme, using van der Waals and hydrogen bonding terms, provides a self-consistent measure of binding strength (Kontopidis *et al*., 2006).

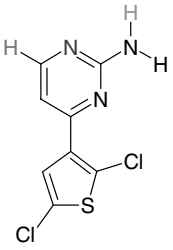
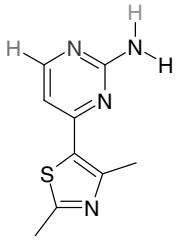
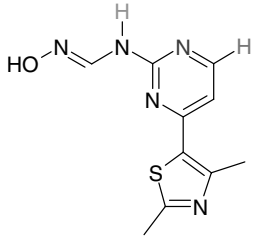
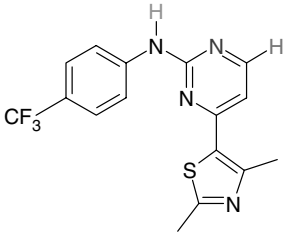
The discovery of cyclapolin, a potent Polo-like kinase inhibitor

Polo-like kinase 1 (Plk1) controls the G2/M transition of the cell cycle by phosphorylating a number of substrates that function in mitotic progression. Overexpression of Plk1 is frequently observed in tumours and in downregulation, using small interfering RNA, has been shown to inhibit cancer cell proliferation (McInnes *et al*., 2005). Small molecule Plk-specific inhibitors are valuable biological tools and can be used as leads for antitumour agents. A number of general kinase inhibitors, such as staurosporine and purvalanol, are known to inhibit Plk1 (McInnes *et al*., 2005). After years of intensive effort by academic and pharmaceutical research groups, the X-ray structure of the kinase domain has recently been published (Kothe *et al*., 2007) in complex with a pyrazole inhibitor (PHA-680626), which has an IC_{50} value of 0.5 μM (PDB code 2owb). Before this structure became available, we had developed a homology model of the kinase domain of Plk based on the staurosporine-bound conformation of protein kinase A, which has a 31% sequence identity (PDB code 1stc). The model was shown to be consistent with the known structure-activity properties of a series of ligands which were docked into the binding pocket in a similar manner to that found in CDK2. LIDAEUS was used to dock a library of 200 000 commercially available compounds into the modelled active site of Plk1. A total of 350 of the top-ranked compounds were then assayed by measuring inhibition of Plk1 phosphorylation of Cdc25C. A number of Plk1 inhibitors were identified with potencies ranging between 0.5 and 20 μM. A series of compounds (named the cyclapolins) based on the benziathole N-oxide core of the most active hit were synthesized and provide a consistent structure-activity relationships for the inhibition of Plk1 (Figure 7). The most active compound in this series showed significant improvement in potency and has an IC_{50} value of 2 nM. For this series, there is a good correlation between the docking score and potency. Treatment of HeLa cells with cyclapolin1 leads to mitotic cells that show severe spindle abnormalities (McInnes *et al*., 2006).

Outlook

The evolution of structure-based lead discovery has been guided by fashion and by some interesting technological advances. Twenty-five years ago, we had the first useful molecular graphics systems that could help medicinal chemists visualize molecular properties. This technology along with fast data collection and structure determination of protein X-ray structures opened the path to structure-based methods. Ironically, in the mid 1990s, just as this approach was beginning to bear fruit, the fashion swung to robotics and high-throughput screening, possibly spurred by the newly founded discipline of Combinatorial Chemistry,

Table 4 CDK2 inhibitors discovered using LIDAEUS

Compound	*Kinase inhibition (CDK2/ cyclin E) IC_{50} (μM)*	*Hydrogen-bonding pattern*	*Reference*
CYC1 	17	ATP hydrogen-bonding pattern: NH….O(Glu81), 2.96 Å, N….HN(Leu83), 3.64 Å, CH….O(Leu83), 3.36 Å	(Wu *et al*., 2003)
CYC2 	13	ATP hydrogen-bonding pattern: NH….O(Glu81), 2.86 Å, N….HN(Leu83), 3.30 Å, CH….O(Leu83), 3.25 Å	(Wu *et al*., 2003)
CYC3 	2.2	Alternative hydrogen bonding pattern: CH….O(Glu81), 3.31 Å, N….HN(Leu83),2.82 Å, NH….O(Leu83), 2.54 Å Ligand flips over to allow the bulky substituent to point out of the pocket	(Wu *et al*., 2003)
CYC4 	0.9	Alternative hydrogen bonding pattern: CH….O(Glu81), 3.31 Å, N….HN(Leu83), 2.92 Å, NH….O(Leu83), 2.58 Å Ligand flips over to allow the bulky substituent to point out of the pocket	(Wu *et al*., 2003)

Colour coding denotes atoms involved in key hydrogen-bonding interactions.

which made it possible to generate very large libraries of compounds. Now in larger organizations high-throughput screening, *in silico* and structure-based approaches are quite well integrated.

The main challenges in docking are still the old problems of how to efficiently model effects including dielectrics, entropy, water and flexibility. Advances in computer architectures may help tackle such problems. However, we also need to design methods that allow efficient simulation of these effects. Possibilities include using cliques of side chain conformations around the active site, and hybrid molecular modelling/quantum mechanical calculations. High-throughput virtual screening, using simplified methods (non-quantum mechanical or complex molecular dynamics), can already achieve docking rates of over 1 M compounds an hour (Shave *et al*., 2008).

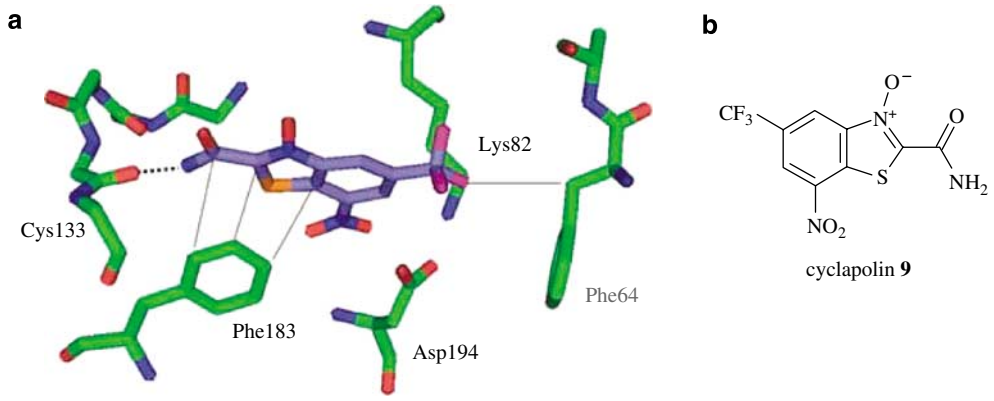


Figure 7 Cyclapolin 9 in the active site of Plk1. (**a**) Proposed binding mode of cyclapolin 9 in the ATP-binding site of Plk1, the black dotted line represents a hydrogen bonding interaction between cyclapolin and Cys133 and the grey lines hydrophobic interactions between cyclapolin and Phe183. Diagram reproduced from (McInnes *et al.*, 2006). (**b**) Cyclapolin 9 is the top hit from virtual screening, IC_{50} 500 nM. Plk1, Polo-like kinase 1.

Technical advances in miniaturization (396-well plates) and sensitive ligand-binding assays are already generating very large amounts of binding data, which contribute to structure-activity relationships. Accurate protein–ligand binding data can add to our understanding of how proteins recognize ligands. Identifying the key features of successful virtual screening calculations can only enhance the chances of discovering new ligands.

Acknowledgements

Elizabeth Blackburn acknowledges support from Organon, UK as part of a CASE studentship with the Biotechnology and Biological Sciences Research Council, UK (BBSRC). Simon Harding and Steven Shave acknowledge support from the BBSRC.

Conflict of interest

The authors state no conflict of interest.

References

Abagyan R, Totrov M, Kuznetsov D (1994). ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* **15**: 488–506.

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

Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, Kalid O *et al.* (2004). G protein-coupled receptors: *in silico* drug discovery in 3D. *Proc Natl Acad Sci USA* **101**: 11304–11309.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res* **28**: 235–242.

Bissantz C, Folkers G, Rognan D (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **43**: 4759–4767.

Brenk R, Naerum L, Gradler U, Gerber HD, Garcia GA, Reuter K *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.

Carr RA, Congreve M, Murray CW, Rees DC (2005). Fragment-based lead discovery: leads by design. *Drug Discov Today* **10**: 987–992.

Chene P (2006). Drugs targeting protein–protein interactions. *ChemMedChem* **1**: 400–411.

Collins I, Garrett MD (2005). Targetting the cell division cycle in cancer. CDK and cell cycle checkpoint kinase inhibitors. *Curr Opin Pharmacol* **4**: 366–373.

de BA, Graff JR (2004). eIF-4E expression and its role in malignancies and metastases. *Oncogene* **23**: 3189–3199.

Degterev A, Lugovskoy A, Cardone M, Mulley B, Wagner G, Mitchison T *et al.* (2001). Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat Cell Biol* **3**: 173–182.

Delaney JS (2005). Predicting aqueous solubility from structure. *Drug Discov Today* **10**: 289–295.

Enyedy IJ, Ling Y, Nacro K, Tomita Y, Wu X, Cao Y *et al.* (2001). Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J Med Chem* **44**: 4313–4324.

Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **15**: 411–428.

Feher M (2006). Consensus scoring for protein–ligand interactions. *Drug Discov Today* **11**: 421–428.

Fischer G, Bang H, Mech C (1984). Determination of enzymatic catalysis for the *cis–trans*-isomerization of peptide binding in proline-containing peptides]. *Biomed Biochim Acta* **43**: 1101–1111.

Fischer S, Michnick S, Karplus M (1993). A mechanism for rotamase catalysis by the FK506 binding protein (FKBP). *Biochemistry (Moscow)* **32**: 13830–13837.

Fishman MC, Porter JA (2005). Pharmaceuticals: a new grammar for drug discovery. *Nature* **437**: 491–493.

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT *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.

Fry DC, Vassilev LT (2005). Targeting protein–protein interactions for cancer therapy. *J Mol Med* **83**: 955–963.

Ghosh S, Nie A, An J, Huang Z (2006). Structure-based virtual screening of chemical libraries for drug discovery. *Curr Opin Chem Biol* **10**: 194–202.

Goodsell DS, Morris GM, Olson AJ (1996). Automated docking of flexible ligands: applications of AutoDock. *J Comput Aided Mol Des* **9**: 1–5.

Greene J, Kahn S, Savoj H, Sprague P, Teig S (1994). Chemical function queries for 3D database search. *J Chem Inf Comput Sci* **34**: 1297–1308.

Grotthuss MV, Pas J, Koczyk G, Wyrwicz LS (2004). Ligand. Info small-molecule Meta-Database. *Comb Chem High Throughput Screen* **7**: 757–761.

Gruneberg S, Stubbs MT, Klebe G (2002). Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem* **45**: 3588–3602.

Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* **87**: 297–306.

Guner O, Clement O, Kurogi Y (2004). Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr Med Chem* **11**: 2991–3005.

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT *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.

Hann MM, Oprea TI (2004). Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* **8**: 255–263.

Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005). Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **48**: 403–413.

Hopkins AL, Groom CR (2002). The druggable genome. *Nat Rev Drug Discov* **1**: 727–730.

Hu J, Hubbard SR (2005). Structural characterization of a novel Cbl phosphotyrosine recognition motif in the APS family of adapter proteins. *J Biol Chem* **280**: 18943–18949.

Irwin JJ, Shoichet BK (2005). ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **45**: 177–182.

Jacobsson M, Liden P, Stjernschantz E, Bostrom H, Norinder U (2003). Improving structure-based virtual screening by multivariate analysis of scoring data. *J Med Chem* **46**: 5781–5789.

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997). Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**: 727–748.

Jorgensen WL, Duffy EM (2002). Prediction of drug solubility from structure. *Adv Drug Deliv Rev* **54**: 355–366.

Kan D (2007). Studies of protein-ligand interactions and the discovery of new cyclophilin inhibitors. PhD Thesis, University of Edinburgh, Edinburgh.

Kitchen DB, Decornez H, Furr JR, Bajorath J (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **3**: 935–949.

Kofron JL, Kuzmic P, Kishore V, Colon-Bonilla E, Rich DH (1991). Determination of kinetic constants for peptidyl prolyl *cis–trans* isomerases by an improved spectrophotometric assay. *Biochemistry (Moscow)* **30**: 6127–6134.

Kontopidis G, McInnes C, Pandalaneni SR, McNae I, Gibson D, Mezna M *et al*. (2006). Differential binding of inhibitors to active and inactive CDK2 provides insights for drug design. *Chem Biol* **13**: 201–211.

Kothe M, Kohls D, Low S, Coli R, Cheng AC, Jacques SL *et al*. (2007). Structure of the catalytic domain of human polo-like kinase 1. *Biochemistry (Moscow)* **46**: 5960–5971.

Kramer B, Rarey M, Lengauer T (1999). Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins* **37**: 228–241.

Lee JO, Russo AA, Pavletich NP (1998). Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature* **391**: 859–865.

Lipinski C, Hopkins A (2004). Navigating chemical space for biology and medicine. *Nature* **432**: 855–861.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **23**: 3–25.

Liu Z, Sun C, Olejniczak ET, Meadows RP, Betz SF, Oost T *et al*. (2000). Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* **408**: 1004–1008.

McInnes C, Mazumdar A, Mezna M, Meades C, Midgley C, Scaerou F *et al*. (2006). Inhibitors of Polo-like kinase reveal roles in spindle-pole maintenance. *Nat Chem Biol* **2**: 608–617.

McInnes C, Mezna M, Fischer PM (2005). Progress in the discovery of polo-like kinase inhibitors. *Curr Top Med Chem* **5**: 181–197.

Moriguchi I, Hirono S, Liu Q, Nakagome Y, Matsushita Y (1992). Simple methods of calculating octanol water partition coefficient. *Chem Pharm Bull* **40**: 127–130.

Muegge I, Martin YC (1999). A general and fast scoring function for protein–ligand interactions: a simplified potential approach. *J Med Chem* **42**: 791–804.

Oprea TI, Allu TK, Fara DC, Rad RF, Ostopovici L, Bologa CG (2007). Lead-like, drug-like or 'Pub-like': how different are they? *J Comput Aided Mol Des* **21**: 113–119.

Oprea TI, Matter H (2004). Integrating virtual screening in lead discovery. *Curr Opin Chem Biol* **8**: 349–358.

Pauptit RA, Dennis CA, Derbyshire DJ, Breeze AL, Weston SA, Rowsell S *et al*. (2001). NMR trial models: experiences with the colicin immunity protein Im7 and the p85alpha C-terminal SH2-peptide complex. *Acta Crystallogr D Biol Crystallogr* **57**: 1397–1404.

Perola E, Walters WP, Charifson PS (2004). A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **56**: 235–249.

Pospisil P, Kuoni T, Scapozza L, Folkers G (2002). Methodology and problems of protein–ligand docking: case study of dihydroorotate dehydrogenase, thymidine kinase, and phosphodiesterase 4. *J Recept Signal Transduct Res* **22**: 141–154.

Qureshi SA, Kim RM, Konteatis Z, Biazzo DE, Motamedi H, Rodrigues R *et al*. (1999). Mimicry of erythropoietin by a nonpeptide molecule. *Proc Natl Acad Sci USA* **96**: 12156–12161.

Rarey M, Kramer B, Lengauer T, Klebe G (1996). A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **261**: 470–489.

Raymond JW, Jalaie M, Bradley MP (2004). Conditional probability: a new fusion method for merging disparate virtual screening results. *J Chem Inf Comput Sci* **44**: 601–609.

Rees DC, Congreve M, Murray CW, Carr R (2004). Fragment-based lead discovery. *Nat Rev Drug Discov* **3**: 660–672.

Rosenfeld R, Vajda S, DeLisi C (1995). Flexible docking and design. *Annu Rev Biophys Biomol Struct* **24**: 677–700.

Russ AP, Lampel S (2005). The druggable genome: an update. *Drug Discov Today* **10**: 1607–1610.

Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H *et al*. (2000). Structure of TPR domain–peptide complexes: critical elements in the assembly of the Hsp70–Hsp90 multichaperone machine. *Cell* **101**: 199–210.

Schuffenhauer A, Brown N, Selzer P, Ertl P, Jacoby E (2006). Relationships between molecular complexity, biological activity, and structural diversity. *J Chem Inf Model* **46**: 525–535.

Shave SR, Taylor P, Walkinshaw M, Smith L, Hardy J, Trew A (2008). Ligand discovery on massively parallel systems. *IBM Journal of Research and Development* **52** 1/2 January/March.

Singh J, Van VH, Liao Y, Lee WC, Cornebise M, Harris M *et al*. (2002). Identification of potent and novel alpha4beta1 antagonists using *in silico* screening. *J Med Chem* **45**: 2988–2993.

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

Teague SJ, Davis AM, Leeson PD, Oprea T (1999). The Design of Leadlike Combinatorial Libraries. *Angew Chem Int Ed Engl* **38**: 3743–3748.

Todeschini R, Consonni V (2005). *Handbook of Molecular Descriptors*. Wiley-VCH: UK.

Vangrevelinghe E, Zimmermann K, Schoepfer J, Portmann R, Fabbro D, Furet P (2003). Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J Med Chem* **46**: 2656–2662.

Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z *et al*. (2004). *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**: 844–848.

Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD *et al*. (2004). Virtual screening using protein–ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci* **44**: 793–806.

Vieth M, Hirst JD, Dominy BN, Daigler H, Brooks CL (1998). Assessing search strategies for flexible docking. *J Comput Chem* **19**: 1623–1631.

Wang JL, Liu DX, Zhang ZJ, Shan SM, Han XB, Srinivasula SM *et al*. (2000). Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proc Natl Acad Sci USA* **97**: 7124–7129.

Wang R, Lu Y, Wang S (2003). Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* **46**: 2287–2303.

Wang S, Meades C, Wood G, Osnowski A, Anderson S, Yuill R *et al*. (2004a). 2-Anilino-4-(thiazol-5-yl)pyrimidine CDK inhibitors: synthesis, SAR analysis, X-ray crystallography, and biological activity. *J Med Chem* **47**: 1662–1675.

Wang S, Wood G, Meades C, Griffiths G, Midgley C, McNae I *et al*. (2004b). Synthesis and biological activity of 2-anilino-4-(1H-pyrrol-3-yl) pyrimidine CDK inhibitors. *Bioorg Med Chem Lett* **14**: 4237–4240.

Woo HJ, Roux B (2005). Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc Natl Acad Sci USA* **102**: 6825–6830.

Wu B, Li P, Liu Y, Lou Z, Ding Y, Shu C *et al*. (2004). 3D structure of human FK506-binding protein 52: implications for the assembly of the glucocorticoid receptor/Hsp90/immunophilin heterocomplex. *Proc Natl Acad Sci USA* **101**: 8348–8353.

Wu SY, McNae I, Kontopidis G, McClue SJ, McInnes C, Stewart KJ *et al*. (2003). Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure* **11**: 399–410.

Wu X, Knudsen B, Feller SM, Zheng J, Sali A, Cowburn D *et al*. (1995). Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure* **3**: 215–226.

Xing L, Hodgkin E, Liu Q, Sedlock D (2004). Evaluation and application of multiple scoring functions for a virtual screening experiment. *J Comput Aided Mol Design* **18**: 333–344.

Yang JM, Chen CC (2004). GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* **55**: 288–304.

Zhao L, Brinton RD (2005). Structure-based virtual screening for plant-based ERbeta-selective ligands as potential preventative therapy against age-related neurodegenerative diseases. *J Med Chem* **48**: 3463–3466.

Zheng CJ, Han LY, Yap CW, Ji ZL, Cao ZW, Chen YZ (2006). Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol Rev* **58**: 259–279.