Lead discovery using molecular docking
Brian K Shoichet*, Susan L McGovern, Binqing Wei and John J Irwin

As the structures of more and more proteins and nucleic acids become available, molecular docking is increasingly considered for lead discovery. Recent studies consider the hit-rate enhancement of docking screens and the accuracy of docking structure predictions. As more structures are determined experimentally, docking against homology-modeled targets also becomes possible for more proteins. With docking studies being undertaken, the ‘drug-likeness’ and specificity of docking hits is also being examined.

Introduction
Given the atomic resolution structure of a macromolecule, such as an enzyme, it should be possible to find novel molecules that bind to it, modulating its activity. This is the premise behind all structure-based ligand design and discovery efforts. Here we consider one aspect of this field: structure-based ligand discovery, emphasizing the screening of compound libraries using molecular docking.

The promise of docking is that the structure of the target will provide a template for the discovery of novel ligands, dissimilar to those previously known. One begins with a database of compounds and the structure of a receptor of interest and asks, ‘Of the compounds in the database, which is most likely to bind to the receptor?’ (Figure 1). This apparently simple question disguises an enormous problem and a clever, if sometimes unsuccessful, strategy. The problem is that of predicting absolute binding affinities of many disparate molecules. Whereas predicting relative binding affinities for related molecules is possible, although time consuming, we have no reason to expect that we can predict absolute binding affinities for so many unrelated compounds. If docking has had an impact, it is because of its strategy of using databases of available compounds. This makes failure cheap; one simply goes to the next compound in a hit list. Docking for discovery is a screening technique: both false positives and false negatives are tolerated as long as true positives are found at a sufficiently high rate to justify the effort. How high should this hit rate be?

Comparing docking and high-throughput screening
Because high-throughput screening (HTS) is the dominant technique for pharmaceutical lead discovery, what level of hit-rate enhancement would be sufficient to justify pursuing structure-based docking? Indeed, if HTS is available, why do docking at all (Figure 1)? Two recent studies begin to consider this question.

Using both HTS and virtual screening, inhibitors were sought for the type 2 diabetes target protein tyrosine phosphatase 1B (PTP1B) [1]. In the HTS experiments, a 400 000 compound in-house library was screened, whereas libraries of commercially available compounds were docked against the X-ray structure of PTP1B. Ultimately, 365 high-scoring docked compounds were tested experimentally. The hit rate for these 365 molecules was 1700-fold higher than that found by HTS and the docking hits were, surprisingly, more drug-like (Figure 2). Intriguingly, there was no overlap among the HTS and docking hit lists, suggesting that the two techniques are complementary.

This PTP1B comparison was imperfect in that the databases being screened and docked were different. A fairer comparison is reported by Paiva et al. [2•], who compared virtual screening and HTS of the Merck chemical collection against the tuberculosis target dihydrodipicolinate reductase. With binding defined as an IC_{50} < 100 µM, they report a hit rate of <0.2% for HTS, and a 6% hit rate for virtual screening. The best docking and HTS hits had K_{i} values of 7.2 µM and 35 µM, respectively.

Although these comparisons are at an early stage, there is some reason to hope that structure-based docking screens can enrich hit rates and that the hits are no less drug-like than those obtained from HTS, even when screening commercially available databases. In the next section, we consider two subsequent questions: how novel are docking hits, and how well do the structural predictions correspond to subsequent experimental results?

Recent applications of docking
Docking has been used to discover novel ligands for well over 30 targets. Work in the past year has continued to focus on enzymes (Table 1). The inhibitors discovered were novel, having little similarity to the known ligands. Most initial leads had affinities in the low-micromolar range. The new sulfonamide inhibitors of carbonic anhydrase II [3] are the exception — they are much more potent, but not completely novel. There has been an efflorescence of new docking methods in the past several years (see below), and several docking programs were used in these studies (Table 1).
Many of the docking studies included structure determination of a lead bound to the enzyme ([3,4,5*]; R Powers, B Wei, BK Shoichet, unpublished data; Table 1). In these papers, the docking predictions captured the X-ray results well (Figure 3). In the case of tRNA guanine transglycosylase (TGT), the deviations between the crystallographic and predicted geometries were less than 0.5Å. Whereas in 1996 we described docking methods predicting ligand-complexes to ‘low resolution’ [6], the field has since progressed sufficiently that, at least in circumstances where the target remains relatively rigid and the ligand has only a few rotatable bonds, docking can suggest new ligands and also their bound geometries with some accuracy (but see [7] for a counter-example).

In several projects, the docking results were followed with biological studies ([8,9]; R Powers, BK Shoichet, unpublished data). An example is the discovery of micromolar ligands for the retinoic acid receptor [8], where the ligands were agonists and their efficacy was reported in cell culture. Both this effort and that against Bcl-2 [9] were also methodologically noteworthy: the structures of the targets were not enzymes, and both were homology modeled.

Figure 1

Characteristics of PTP1B inhibitors discovered from HTS (blue) and virtual screening (yellow) [1]. The numerical percentages are given at the top of each column.
Integration of docking with design and virtual libraries

Can docking hits be turned into leads through synthetic elaboration? In several studies [5, 10], the affinity of hits was improved by 10- to 1000-fold (Table 1), often through ‘classical’ structure-based techniques (i.e. beginning with a lead and using the complexed structure and chemical intuition to improve it). More sophisticated efforts have tried to include synthetic accessibility in the virtual screening from the beginning. In an effort to design novel CDK4 inhibitors, the de novo design program LEGEND was used to suggest possible new inhibitor scaffolds [4]. The problem of synthetic inaccessibility, a common complaint of de novo design methods, was overcome by using the program SEEDS to search the Available Chemicals Directory (MDL, Inc., San Leandro, CA) for scaffolds that were components of the designed ligands. These scaffolds were tested and one class of active molecules was elaborated using the docked geometries as a guide. A scaffold-based approach was also used to discover and elaborate novel inhibitors of thrombin [11], Factor Xa [12] and, earlier, DNA gyrase [13]. These scaffold-based approaches are methodologically tractable because the ligands are typically rigid and allow for synthetic elaboration — more work in this area may be expected in the future.

There are ongoing efforts to include combinatorial chemistry in docking virtual libraries [14, 15]. These methods yield compounds that are synthetically accessible and cover much more chemical space than simple database docking, but the high-scoring docked molecules cannot be immediately acquired and tested. To focus the chemical space that is screened, the virtual libraries may be tailored for a particular target, for example using pharmacophoric constraints [16–18], or may be restricted to elaborations of particular scaffolds [4, 19]. Iterative cycles of combinatorial library design, virtual screening, synthesis and biological testing have been used [12]. Hybrid methods employing

<table>
<thead>
<tr>
<th>Target</th>
<th>Representative hit</th>
<th>Lead inhibitor IC_{50} (µM)</th>
<th>Follow up inhibitor IC_{50} (µM)</th>
<th>Docking program</th>
<th>Type of structure used?</th>
<th>Complex structure solved?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aldose reductase [10]</td>
<td><img src="image" alt="Image" /></td>
<td>4.3</td>
<td>0.21</td>
<td>Adam&amp;Eve</td>
<td>X-ray</td>
<td>No</td>
</tr>
<tr>
<td>CDK4 [4]</td>
<td><img src="image" alt="Image" /></td>
<td>44</td>
<td>0.011</td>
<td>Legend/SEEDS</td>
<td>Homology model</td>
<td>Yes*</td>
</tr>
<tr>
<td>Matriptase [43]</td>
<td><img src="image" alt="Image" /></td>
<td>0.92</td>
<td>0.21</td>
<td>DOCK</td>
<td>Homology model</td>
<td>No</td>
</tr>
<tr>
<td>Bcl-2 [9]</td>
<td><img src="image" alt="Image" /></td>
<td>10.4</td>
<td>NR</td>
<td>DOCK</td>
<td>Homology model</td>
<td>No</td>
</tr>
<tr>
<td>Adenovirus protease [58]</td>
<td><img src="image" alt="Image" /></td>
<td>3.1</td>
<td>NR</td>
<td>EUDOC</td>
<td>X-ray</td>
<td>No</td>
</tr>
<tr>
<td>AmpC ß-lactamase (unpublished data)</td>
<td><img src="image" alt="Image" /></td>
<td>26</td>
<td>NR</td>
<td>DOCK</td>
<td>X-ray</td>
<td>Yes</td>
</tr>
</tbody>
</table>
**Table 1 continued**

<table>
<thead>
<tr>
<th>Target</th>
<th>Representative hit</th>
<th>Lead inhibitor IC₅₀ (µM)</th>
<th>Follow up inhibitor IC₅₀ (µM)</th>
<th>Docking program</th>
<th>Type of structure used?</th>
<th>Complex structure solved?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retinoic acid receptor [8]</td>
<td><img src="image" alt="Structure" /></td>
<td>2ᵃ</td>
<td>NR</td>
<td>ICM</td>
<td>Homology model</td>
<td>No</td>
</tr>
<tr>
<td>TGT [5⁺]</td>
<td><img src="image" alt="Structure" /></td>
<td>8.3</td>
<td>0.2</td>
<td>LUDI</td>
<td>X-ray</td>
<td>Yes</td>
</tr>
<tr>
<td>Carbonic anhydrase II [3]</td>
<td><img src="image" alt="Structure" /></td>
<td>0.0008</td>
<td>NR</td>
<td>flexx</td>
<td>X-ray</td>
<td>Yes</td>
</tr>
<tr>
<td>HPRTase [59]</td>
<td><img src="image" alt="Structure" /></td>
<td>2.2</td>
<td>NR</td>
<td>DOCK⁺</td>
<td>X-ray</td>
<td>No</td>
</tr>
<tr>
<td>Lysozyme cavity site</td>
<td><img src="image" alt="Structure" /></td>
<td>56⁺</td>
<td>NR</td>
<td>DOCK⁺</td>
<td>X-ray</td>
<td>Yes</td>
</tr>
<tr>
<td>Dihydro-dipicolinate [2⁺]</td>
<td><img src="image" alt="Structure" /></td>
<td>7.2</td>
<td>NR</td>
<td>FLOG</td>
<td>X-ray</td>
<td>No</td>
</tr>
</tbody>
</table>

ᵃStructure determined with the analogous CDK2 enzyme. ᵇCanonical DOCK from UCSF. ᶜNot Reported. ᵈThe Northwestern University version of DOCK, a derivative of UCSF DOCK. ᵉEC₅₀ in cell culture. ᶠKᵢ. ᵍThe Northwestern University version of DOCK, a derivative of UCSF DOCK. ʰEC₅₀ in cell culture. ｉKᵢ.

* Technical advances in docking algorithms

Molecular docking continues to witness the introduction of new algorithms and programs — these are much needed given the weaknesses in conformational sampling and scoring. Along with efforts to improve established docking programs, such as AutoDock, DOCK, Ecepp/ProDock, FlexX, FLOG, GOLD, GREEN, ICM, LUDI, ProLeads, QXP and SLIDE (reviewed in [21] and [22]), new docking programs have been published in the past year, including the EUDOC algorithm [23], SEED [24⁺], SEEDS [4⁺] and MM [25].

Two ongoing methodological challenges are adequate sampling of ligand–receptor configurations and accurate evaluation of their complementarity. The treatment of ligand conformations has been incorporated into most of the current docking algorithms. Treatment of receptor flexibility remains a major challenge [22,26]. Abagyan [22] has pointed out that if the receptor conformation sampled is high-energy, and this is not accounted for in the docking calculation, adding receptor flexibility may actually make docking calculations worse.

The program FlexE [27⁺], an extension to the FlexX algorithm, explicitly samples a predefined ensemble of receptor structures. The structures are superposed and alternative conformations are recombined to create complete structures of the receptor. Docking against this ensemble is twofold faster than explicit docking against all conformations. In another approach, Goodsell and co-workers [28] adopted a single averaged interaction energy grid to implicitly account for an ensemble of receptor conformations in the AutoDock program (see also [29]). In the SLIDE program, Schnecke and Kuhn [30] introduced a step to optimize the conformations of receptor side chains after initial placement of ligand. These methods are new and their effect on database docking is still being evaluated.

* de novo design, combinatorial chemistry and molecular docking draw on the strengths of each, and blur the distinction among these methods [20].
Scoring functions for database docking continue to be actively researched. These functions may be loosely divided into three categories: force-field methods, knowledge-based methods and empirical-regression methods. Force-field methods use potentials similar to those found in molecular mechanics and can be linked to more quantitatively reliable techniques such as molecular dynamics and thermodynamic integration. On the other hand, they are prone to calculating high magnitude, and high error, interaction energies. Both knowledge-based and empirical scoring functions derive from experimental data; the former from patterns of atom contacts observed in structures, the latter from fits to binding energies. Both are less subject to calculating overly large interaction energies, but can suffer from problems of induction, including error in the data from which they are derived. In force-field-based scoring functions, much recent work has focused on efficient methods to incorporate solvation energies in docking scores. Many methods use a generalized Born/surface area model or other approximations of continuum electrostatics, and these continue to be explored [24•,31]. Several studies suggest that docking screens may be improved by using better partial atomic charges for the ligands [32] (B Wei, BK Shoichet, unpublished data). Recent improvements in both empirical and knowledge-based scoring functions [33] have focused on improving the balance between polar and non-polar interactions [34], considering the role of solvent [35], and correcting for intra-ligand contacts in the structures from which the knowledge-based potentials are derived [36].

There is, at present, no consensus as to which type of scoring function is the best for docking screens; good results have been achieved with all three approaches. Similarly, whereas all three can exclude unreasonable molecules from hit lists, for instance molecules that are too big or too charged for a target site, none can reliably rank ‘reasonable’ docking hits, except in special cases. It is perhaps a testament to the weakness of all current scoring functions that some of the most reliable rankings are achieved by ‘consensus’ scoring schemes that combine weighted scores from several, often fundamentally different, approaches [37•,38–41].

The use of homology models in docking
For many interesting targets, an experimental structure is unavailable. In principle, homology modeling can calculate a structure for use in drug discovery [42], thereby dramatically increasing the number of targets to which docking might be applied. How reliable are these models for docking? This question cannot be answered definitively, but some tentative points may be made.

Several groups have used homology models to design or discover ligands for target proteins. Indeed, of the 12 docking screens summarized in Table 1, four were based on homology models of the target structures [4•,8,9,43]. Docking against homology-modeled structures has also been used to improve the pharmacokinetic properties of known inhibitors [44] and to develop or expand structure–activity relationships [4•,45]. De novo design methods have been used with modeled structures to develop new inhibitors [4•,46]. At least one group has developed a docking algorithm specifically for modeled structures [47].

Some guides are available for how much sequence identity there should be for virtual screening to be applied; estimates of 60% sequence identity between the template and the model proteins have been suggested [48•]. As it happens, models with 45–56% sequence identity were used in the studies considered here [4•,9,45]. Intriguingly, many of the homology models used in successful database screening or de novo design projects were modeled on ligand-bound conformations of the template protein [4•,8,9,46].

Hit conformation and promiscuous inhibitors
As is well known to screeners and medicinal chemists, many HTS hits are promiscuous and non ‘drug-like’. This can also be true of docking hits. Whereas detailed testing
Pathological inhibitors have a single mechanism of action both virtual and experimental screening databases. These are widely reported as hits. They may, indeed, be common in protein databases, appearing to act as inhibitors. Aggregating molecules are welcome in hit lists and in docking databases. A second class of promiscuous inhibitors includes molecules that chemically react with proteins. Filters to remove reactive and non-'drug-like' compounds from both virtual and experimental screening databases are under constant development.

Small molecules that aggregate in solution at low micro-molar concentrations constitute a third class of promiscuous inhibitors ([57*]) (Figure 4). These aggregates are typically 50 to 400 nm in size — several compounds may form even larger particles — and may adsorb or absorb the target proteins, appearing to act as inhibitors. Aggregating molecules are common in docking and HTS hit lists, and have been widely reported as hits. They may, indeed, be common in both virtual and experimental screening databases. These ‘pathological’ inhibitors have a single mechanism of action and a distinctive behavioral fingerprint by which they can be recognized.

To detect aggregating ‘inhibitors’, we suggest re-testing hits at their apparent IC_{50} using a 10-fold higher concentration of the target protein than was used in the original assay. Well-behaved inhibitors should not be affected, but aggregate-based inhibitors show a dramatic decrease in potency ([57*]). Testing hits in a second counter-screen assay is also sensible ([13**,57*]). Perhaps the most definitive test, if also the slowest, is to use dynamic light scattering to test for large particles in the reaction buffer. These experiments are easily performed and can save much grief by removing these ‘pathological’ hits early, allowing one to focus on more interesting genuine inhibitors that can be produced by molecular docking.

Conclusions
The recent explosion of protein structures, and the advent of the genome projects, has renewed interest in using structure-based docking for early-phase lead discovery. Database docking has made considerable progress in the past decade, but it remains a screening technique. As such, current docking programs will dissatisfy investigators interested in definitive predictions of new inhibitors, and predictions of geometries can still go wildly wrong. In favorable circumstances, docking screens can substantially enrich hit rates and predict the structures of hits bound to their targets in sufficient detail to be useful for the synthetic elaboration of leads. With all of its weaknesses, structure-based screening through docking is mature enough to be considered as a first-line technique in pharmaceutical discovery research.

Acknowledgements
Supported by GM59957 (to BKS). SLM is a Northwestern University Presidential Scholar, and is partly supported by T32-GM08152 (K Mayo, PI) and a PhRMA Foundation Medical Student Fellowship. We are grateful to David Lorber for many discussions.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest

Using a homology model of Cdk4 based on the structure of Cdk2, new scaffolds for Cdk4 inhibitors were discovered. After synthetic exploration, a 42 nM inhibitor was obtained. The bound structure was confirmed by X-ray crystallography of the complex with Cdk2.


The docking screen was followed by X-ray structure determination of the lead bound to the enzyme. The predicted structure was within 0.5 Å of the crystallographic structure.


An admirable aspect of this study was the care taken to confirm that all hits were genuine. Initial hits from virtual screening were discarded if they showed a flat structure-activity relationship, if specific binding was not observed by NMR, or if the compound showed peculiar properties in other biophysical experiments. Ultimately, the authors determined the crystallographic structure of a new inhibitor in complex with DNA gyrase.


Using a Monte Carlo incremental growth algorithm and a knowledge-based scoring function, picomolar inhibitors were successfully discovered based on the structure of a known submicromolar inhibitor.


An extension to the Flex algorithm that takes into account receptor flexibility.


Several of the popular docking scoring functions are compared. The authors describe a consensus function based on many of the individual functions to improve hit rates.


57. McGovern SL, Caselli E, Grigorieff N, Shoichet BK: A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem 2002, 45:1712-1722. The physical and enzymological properties of promiscuous screening hits were examined. The non-specific inhibitors appear to act by forming aggregates of 100-400 nm diameter. Such aggregate-forming molecules may be common in screening libraries.


Now in press
The work referred to in the text as (R. Powers, B.K. Shoichet, unpublished data) is now published. The work referred to in the text as (BW & BKS, unpublished data) is now in press.
