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 more docking studies being undertaken, the 'drug-likeness' and specificity of docking hits is also being examined.

Addresses
Department of Molecular Pharmacology & Biological Chemistry, Northwestern University, 303 East Chicago Avenue, Chicago, IL 60611-3008, USA
*e-mail: b-shoichet@northwestern.edu

Current Opinion in Chemical Biology 2002, 6:439–446

1367-5931/02/$ – see front matter

Abbreviations

HTS	high-throughput screening
PTP1B	protein tyrosine phosphotase 1B
TGT	tRNA guanine transglycosylase

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.

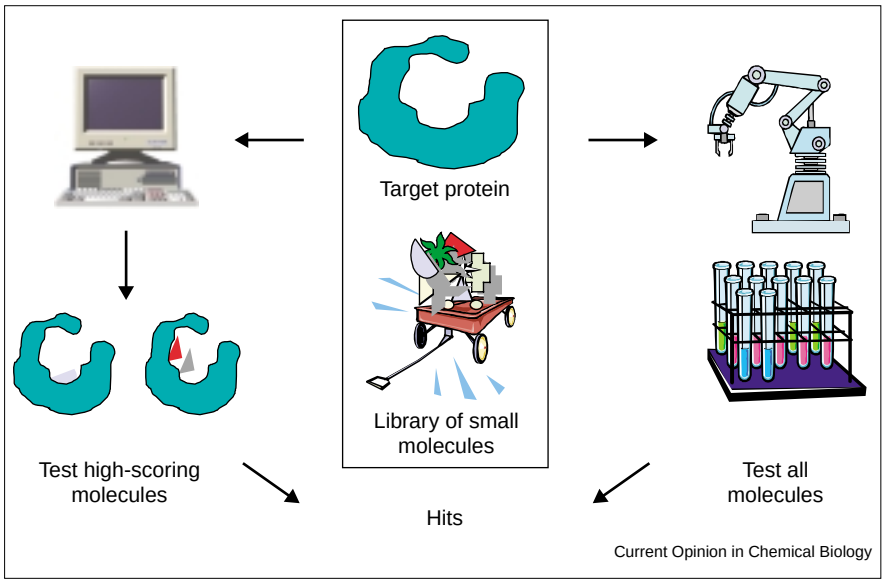
This PTP1B comparison was imperfect in that the databases being screened and docked were different. A fairer comparison is reported by Paiva *et al.* [2•], who compared virtual screening and HTS of the Merck chemical collection against the tuberculosis target dihydrodipicolinate reductase. With binding 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).

Figure 1

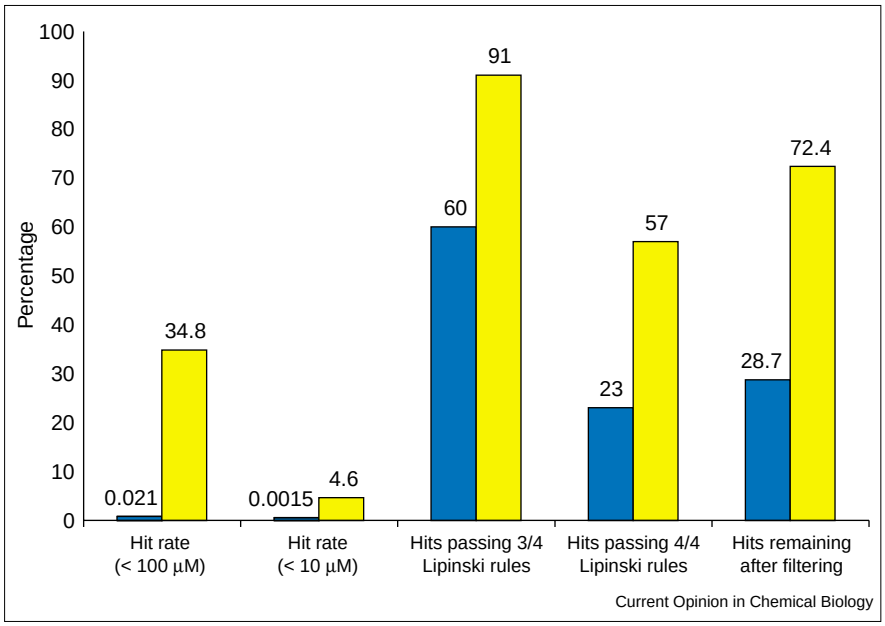


Docking (left) and HTS (right) to discover new leads for drug discovery.

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 0.5 Å. 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.

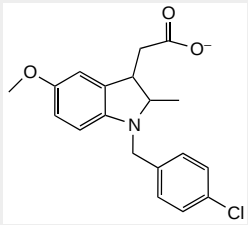
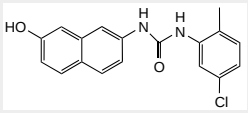
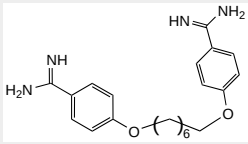
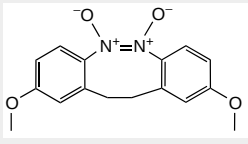
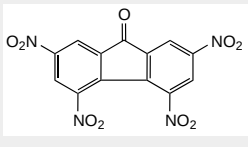
Figure 1



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

Table 1

Recent examples of novel inhibitor discovery using molecular docking.

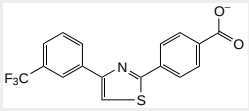
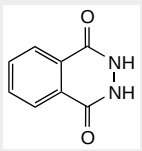
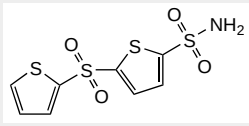
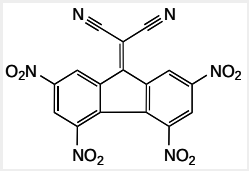
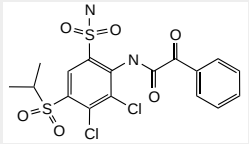
Target	Representative hit	Lead inhibitor IC_{50} (μM)	Follow up inhibitor IC_{50} (μM)	Docking program	Type of structure used?	Complex structure solved?
Aldose reductase [10]		4.3	0.21	Adam&Eve	X-ray	No
CDK4 [4'] Matriptase [43]		44 0.92	0.011 0.21	Legend/SEEDS DOCK[†]	Homology model Homology model	Yes* No
Bcl-2 [9]		10.4	NR[‡]	DOCK[†]	Homology model	No
Adenovirus protease [58]		3.1	NR	EUDOC	X-ray	No
AmpC β-lactamase (unpublished data)		26[§]	NR	DOCK[#]	X-ray	Yes

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 pharmacaphoric 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 1 continued

Target	Representative hit	Lead inhibitor IC_{50} (μM)	Follow up inhibitor IC_{50} (μM)	Docking program	Type of structure used?	Complex structure solved?
Retinoic acid receptor [8]		2[¶]	NR	ICM	Homology model	No
TGT [5•]		8.3	0.2	LUDI	X-ray	Yes
Carbonic anhydrase II [3]		0.0008	NR	FlexX	X-ray	Yes
HPRTase [59]		2.2	NR	DOCK[#]	X-ray	No
Lysozyme cavity site (unpublished data)		56[¥]	NR	DOCK[#]	X-ray	Yes
Dihydro-dipicolinate [2•]		7.2	NR	FLOG	X-ray	No

*Structure determined with the analogous CDK2 enzyme. [†]Canonical DOCK from UCSF. [‡]Not Reported [$]$K_i$. [#]The Northwestern University version of DOCK, a derivative of UCSF DOCK. [¶]EC_{50} in cell culture. [¥]K_d.

de novo design, combinatorial chemistry and molecular docking draw on the strengths of each, and blur the distinction among these methods [20].

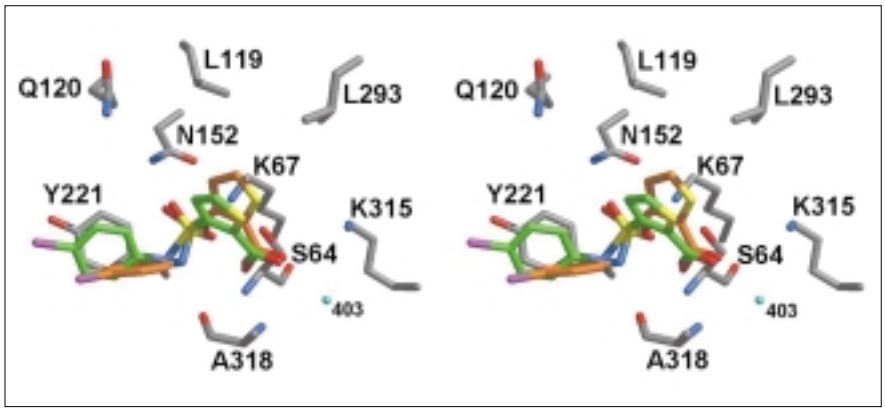
Technical advances in docking algorithms

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, Pro_LEADS, 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.

Figure 3



Overlay of the predicted (carbon atoms in green) and crystallographic (carbon atoms in orange) conformations of a docking-derived inhibitor of AmpC β-lactamase (stereo view) (our unpublished data). The X-ray structure was determined to 1.94 Å resolution, nitrogen atoms in blue, oxygens in red, sulfurs in yellow. Reprinted with permission from *Structure*. Copyright 2002 Elsevier Science.

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.

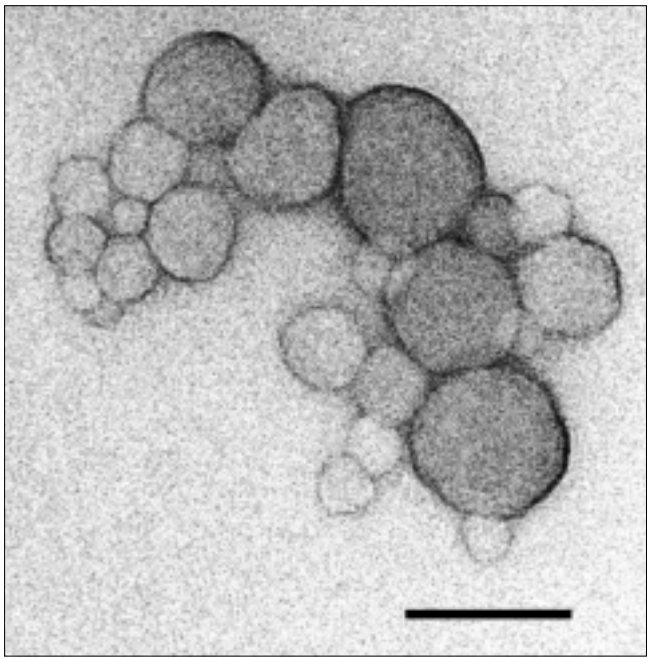
Several groups have used homology models to design or discover ligands for target proteins. Indeed, of the 12 docking screens summarized in Table 1, four were based on homology models of the target structures [4•,8,9,43]. Docking against homology-modeled structures has also been used to improve the pharmacokinetic 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

Figure 4



Aggregates formed by tetraiodophenolphthalein, a promiscuous inhibitor, as visualized by transmission electron microscopy. Bar = 100 nm. Reprinted with permission from the *Journal of Medicinal Chemistry* [57•]. Copyright 2002 American Chemical Society.

to confirm a hit routinely occurs in HTS projects, docking hits are not always confirmed as carefully. In our own painful experience, this can lead to artifacts that are confusing and time consuming. Every effort should be made to remove what we have come to call 'pathological' inhibitors from docking hit lists.

We can consider three classes of non-specific molecules emerging from docking screens. The first is those molecules that present 'privileged structures' [49•] (i.e. whose innate properties make them genuinely more likely to bind to many targets). Such 'privileged structures' typically are welcome in hit lists and in docking databases. A second class of promiscuous inhibitors includes molecules that chemically react with proteins [50]. Filters to remove reactive and non-'drug-like' compounds from both virtual and experimental screening databases are under constant development [49•,51–56].

Small molecules that aggregate in solution at low micromolar 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

1. Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, Stallings WC, Connolly DT, Shoichet BK: Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 2002, 45:2213-2221.

2. • Paiva AM, Vanderwall DE, Blanchard JS, Kozarich JW, Williamson JM, Kelly TM: Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of ***Mycobacterium tuberculosis***. *Biochim Biophys Acta* 2001, 1545:67-77.

The same library of compounds was screened by experimental and docking methods to discover novel inhibitors, and the results compared. Each method found different inhibitors; the virtual screen had a higher hit rate and discovered more-potent inhibitors.

3. Gruneberg S, Wendt B, Klebe G: Subnanomolar inhibitors from computer screening: a model study using human carbonic anhydrase II. *Angew Chem Int Ed Engl* 2001, 40:389-393.

4. • Honma T, Hayashi K, Aoyama T, Hashimoto N, Machida T, Fukasawa K, Iwama T, Ikeura C, Ikuta M, Suzuki-Takahashi I *et al.*: Structure-based generation of a new class of potent Cdk4

inhibitors: new ***de novo*** design strategy and library design. *J Med Chem* 2001, 44:4615-4627.
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.

5. • Gradler U, Gerber HD, Goodenough-Lashua DM, Garcia GA, Ficner R, Reuter K, Stubbs MT, Klebe G: A new target for shigellosis: rational design and crystallographic studies of inhibitors of tRNA-guanine transglycosylase. *J Mol Biol* 2001, 306:455-467.
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.

6. Shoichet BK, Kuntz ID: Predicting the structures of protein complexes: a step in the right direction. *Chem Biol* 1996, 3:151-156.

7. Fritz TA, Tondi D, Finer-Moore JS, Costi MP, Stroud RM: Predicting and harnessing protein flexibility in the design of species-specific inhibitors of thymidylate synthase. *Chem Biol* 2001, 8:981-995.

8. Schapira M, Raaka BM, Samuels HH, Abagyan R: ***In silico*** discovery of novel retinoic acid receptor agonist structures. *BMC Struct Biol* 2001, 1:1.

9. Enyedy IJ, Ling Y, Nacro K, Tomita Y, Wu X, Cao Y, Guo R, Li B, Zhu X, Huang Y *et al.*: Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J Med Chem* 2001, 44:4313-4324.

10. Iwata Y, Arisawa M, Hamada R, Kita Y, Mizutani MY, Tomioka N, Itai A, Miyamoto S: Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J Med Chem* 2001, 44:1718-1728.

11. Makino S, Kayahara T, Tashiro K, Takahashi M, Tsuji T, Shoji M: Discovery of a novel serine protease inhibitor utilizing a structure-based and experimental selection of fragments technique. *J Comput Aided Mol Des* 2001, 15:553-559.

12. Liebeschuetz JW, Jones SD, Morgan PJ, Murray CW, Rimmer AD, Roscoe JM, Waszkowycz B, Welsh PM, Wylie WA, Young SC *et al.*: PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors. *J Med Chem* 2002, 45:1221-1232.

13. •• Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, Kostrewa D, Kuehne H, Luebbers T, Meunier-Keller N *et al.*: Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J Med Chem* 2000, 43:2664-2674.
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.

14. • Grzybowski BA, Ishchenko AV, Kim CY, Topalov G, Chapman R, Christianson DW, Whitesides GM, Shakhnovich EI: Combinatorial computational method gives new picomolar ligands for a known enzyme. *Proc Natl Acad Sci USA* 2002, 99:1270-1273.
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.

15. Siani MA, Skillman AG, Carreras CW, Ashley G, Kuntz ID, Santi DV: Development and screening of a polyketide virtual library for drug leads against a motilide pharmacophore. *J Mol Graph Model* 2000, 18:497-511, 539-440.

16. Joseph-McCarthy D, Tsang SK, Filman DJ, Hogle JM, Karplus M: Use of MCSS to design small targeted libraries: application to picornavirus ligands. *J Am Chem Soc* 2001, 123:12758-12769.

17. Mason JS, Cheney DL: Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac Symp Biocomput* 2000:576-587.

18. Olender R, Rosenfeld R: A fast algorithm for searching for molecules containing a pharmacophore in very large virtual combinatorial libraries. *J Chem Inf Comput Sci* 2001, 41:731-738.

19. Aronov AM, Munagala NR, Kuntz ID, Wang CC: Virtual screening of combinatorial libraries across a gene family: in search of inhibitors of ***Giardia lamblia*** guanine phosphoribosyltransferase. *Antimicrob Agents Chemother* 2001, 45:2571-2576.

20. Lamb ML, Burdick KW, Toba S, Young MM, Skillman AG, Zou X, Arnold JR, Kuntz ID: Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* 2001, 42:296-318.

21. Walters WP, Stahl MT, Murcko MA: Virtual screening — an overview. *Drug Discovery Today* 1998, 3:160-178.

22. Abagyan R, Totrov M: High-throughput docking for lead generation. *Curr Opin Chem Biol* 2001, 5:375-382.

23. Pang YP, Perola E, Xu K, Prendergast FG: EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J Comput Chem* 2001, 22:1750-1771.

24. • Majeux N, Scarsi M, Caflisch A: Efficient electrostatic solvation model for protein-fragment docking. *Proteins* 2001, 42:256-268.
A new method to treat electrostatic solvation is developed for ligand-fragment docking.

25. David L, Luo R, Gilson MK: Ligand-receptor docking with the Mining Minima optimizer. *J Comput Aided Mol Des* 2001, 15:157-171.

26. Carlson HA, McCammon JA: Accommodating protein flexibility in computational drug design. *Mol Pharmacol* 2000, 57:213-218.

27. • Claussen H, Buning C, Rarey M, Lengauer T: FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* 2001, 308:377-395.
An extension to the Flexx algorithm that takes into account receptor flexibility.

28. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS: Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 2002, 46:34-40.

29. Knegtel RM, Kuntz ID, Oshiro CM: Molecular docking to ensembles of protein structures. *J Mol Biol* 1997, 266:424-440.

30. Schnecke V, Kuhn LA: Virtual screening with solvation and ligand-induced complementarity. *Perspectives Drug Discov Des* 2000, 20:171-190.

31. Arora N, Bashford D: Solvation energy density occlusion approximation for evaluation of desolvation penalties in biomolecular interactions. *Proteins* 2001, 43:12-27.

32. Jakalian A, Bush BL, Jack DB, Bayly CI: Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J Comp Chem* 2000, 21:132-146.

33. Shimada J, Ishchenko AV, Shakhnovich EI: Analysis of knowledge-based protein–ligand potentials using a self-consistent method. *Prot Sci* 2000, 9:765-775.

34. Stahl M: Modifications of the scoring function in FlexX for virtual screening applications. *Perspectives Drug Discov Des* 2000, 20:83-98.

35. Gohlke H, Hendlich M, Klebe G: Knowledge-based scoring function to predict protein–ligand interactions. *J Mol Biol* 2000, 295:337-356.

36. Muegge I: Effect of ligand volume correction on PMF scoring. *J Comput Chem* 2001, 22:418-425.

37. • Charifson PS, Corkery JJ, Murcko MA, Walters WP: Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 1999, 42:5100-5109.
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.

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

39. Perez C, Ortiz AR: Evaluation of docking functions for protein-ligand docking. *J Med Chem* 2001, 44:3768-3785.

40. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB: Consensus scoring for ligand/protein interactions. *J Mol Graph Model* 2002, 20:281-295.

41. Terp GE, Johansen BN, Christensen IT, Jorgensen FS: A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein–ligand binding affinities. *J Med Chem* 2001, 44:2333-2343.

42. Al-Lazikani B, Jung J, Xiang Z, Honig B: Protein structure prediction. *Curr Opin Chem Biol* 2001, 5:51-56.

43. Enyedy IJ, Lee SL, Kuo AH, Dickson RB, Lin CY, Wang S: Structure-based approach for the discovery of bis-benzamidines as novel inhibitors of matriptase. *J Med Chem* 2001, 44:1349-1355.

44. Rabinowitz MH, Andrews RC, Becherer JD, Bickett DM, Bubacz DG, Conway JG, Cowan DJ, Gaul M, Glennon K, Lambert MH *et al.*: Design of selective and soluble inhibitors of tumor necrosis factor alpha converting enzyme (TACE). *J Med Chem* 2001, 44:4252-4267.

45. Marhefka CA, Moore BM, 2nd, Bishop TC, Kirkovsky L, Mukherjee A, Dalton JT, Miller DD: Homology modeling using multiple molecular dynamics simulations and docking studies of the human androgen receptor ligand binding domain bound to testosterone and nonsteroidal ligands. *J Med Chem* 2001, 44:1729-1740.

46. Shakespeare W, Yang M, Bohacek R, Cerasoli F, Stebbins K, Sundaramoorthi R, Azimioara M, Vu C, Pradeepan S, Metcalf C III *et al.*: Structure-based design of an osteoclast-selective, nonpeptide src homology 2 inhibitor with ***in vivo*** antiresorptive activity. *Proc Natl Acad Sci USA* 2000, 97:9373-9378.

47. Schafferhans A, Klebe G: Docking ligands onto binding site representations derived from proteins built by homology modelling. *J Mol Biol* 2001, 307:407-427.

48. • Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A: Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 2000, 29:291-325.
A thorough review of many of the issues involved in comparative modeling.

49. • Roche O, Schneider P, Zuegge J, Guba W, Kansy M, Alanine A, Bleicher K, Danel F, Gutknecht EM, Rogers-Evans M *et al.*: Development of a virtual screening method for identification of 'frequent hitters' in compound libraries. *J Med Chem* 2002, 45:137-142.
A computational method for filtering compounds likely to appear as active in multiple screens. The filtering method correctly identified 90% of the test set as either frequent or non-frequent hitters.

50. Rishton GM: Reactive compounds and ***in vitro*** false positives in HTS. *Drug Discov Today* 1997, 2:382-384.

51. Muegge I, Heald SL, Brittelli D: Simple selection criteria for drug-like chemical matter. *J Med Chem* 2001, 44:1841-1846.

52. Frimurer TM, Bywater R, Naerum L, Lauritsen LN, Brunak S: Improving the odds in discriminating 'drug-like' from 'non drug-like' compounds. *J Chem Inf Comput Sci* 2000, 40:1315-1324.

53. Hull RD, Fluder EM, Singh SB, Nachbar RB, Kearsley SK, Sheridan RP: Chemical similarity searches using latent semantic structural indexing (LaSSI) and comparison to TOPOSIM. *J Med Chem* 2001, 44:1185-1191.

54. Gaive J, De Julian-Ortiz JV, Garcia-Domenech R: General topological patterns of known drugs. *J Mol Graph Model* 2001, 20:84-94.

55. Xu J, Stevenson J: Drug-like index: a new approach to measure drug-like compounds and their diversity. *J Chem Inf Comput Sci* 2000, 40:1177-1187.

56. Oprea TI: Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* 2000, 14:251-264.

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.

58. Pang YP, Xu K, Kollmeyer TM, Perola E, McGrath WJ, Green DT, Mangel WF: Discovery of a new inhibitor lead of adenovirus proteinase: steps toward selective, irreversible inhibitors of cysteine proteinases. *FEBS Lett* 2001, 502:93-97.

59. Freymann DM, Wenck MA, Engel JC, Feng J, Focia PJ, Eakin AE, Craig SP: Efficient identification of inhibitors targeting the closed active site conformation of the HPRT from ***Trypanosoma cruzi***. *Chem Biol* 2000, 7:957-968.

Now in press

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

60. Powers RA, Morandi F, Shoichet BK: Structure-based discovery of a novel, non-covalent inhibitor of AmpC -lactamase. *Structure* 2002, 10:in press.

61. Wei BW, Baase WA, Weaver LH, Matthews BW, Shoichet BK: A model binding site for testing scoring functions in molecular docking. *J Mol Biol* 2002, in press.