HITS, LEADS, & ARTIFACTS FROM VIRTUAL & HIGH THROUGHPUT SCREENING

BRIAN K. SHOICHET, SUSAN L. MCGOVERN, BINQING Q. WEI, AND JOHN J. IRWIN

Northwestern University, Department of Molecular Pharmacology & Biological Chemistry,
303 E. Chicago Ave, Chicago, IL 60611-3008, USA

E-Mail: b-shoichet@northwestern.edu

Received: 31st July 2002 / Published: 15th May 2003

Molecular docking attempts to find complementary fits for two molecules, typically a candidate ligand and a macromolecular receptor. Among the most popular applications of docking computer programs is that of screening a database of small molecules for those that might act as ligands for a biological receptor of known or modeled structure. The motivating idea is that the receptor structure can act as a template to select database molecules that will complement it structurally and chemically, and so bind to it, modulating its function. The hope is that this will allow novel families of ligands to be found, allowing one to escape from the tedium of substrate-based or analog-based design (Figure 1).

Although simple in principle, docking screens are shot through with uncertainty. Even small molecule ligands have several rotatable bonds, six is not uncommon, and the receptor site has many more. The number of conformations to be explored in docking rises exponentially with the rotatable bonds, so that even for a small molecule ligand this can be a daunting problem. Whereas most docking programs sample small molecule flexibility, the protein is often left rigid. There are some reasons, moreover, to worry that introducing conformational flexibility into the enzyme could, if not done carefully, make docking performance worse, not better (1).

If sampling is challenging, ranking the database molecules for fit in the site is harder still. Calculating absolute binding energies for a protein and a small molecule ligand is notoriously difficult even for very detailed, time consuming techniques, such as Free Energy Perturbation (FEP).

Published in "Molecular Informatics: Confronting Complexity", Martin G. Hicks & Carsten Kettner (Eds.), Proceedings of the Beilstein-Institut Workshop, May 13th - 16th 2002, Bozen, Italy

http://www.beilstein-institut.de/bozen2002/proceedings/Shoichet/Shoichet.pdf

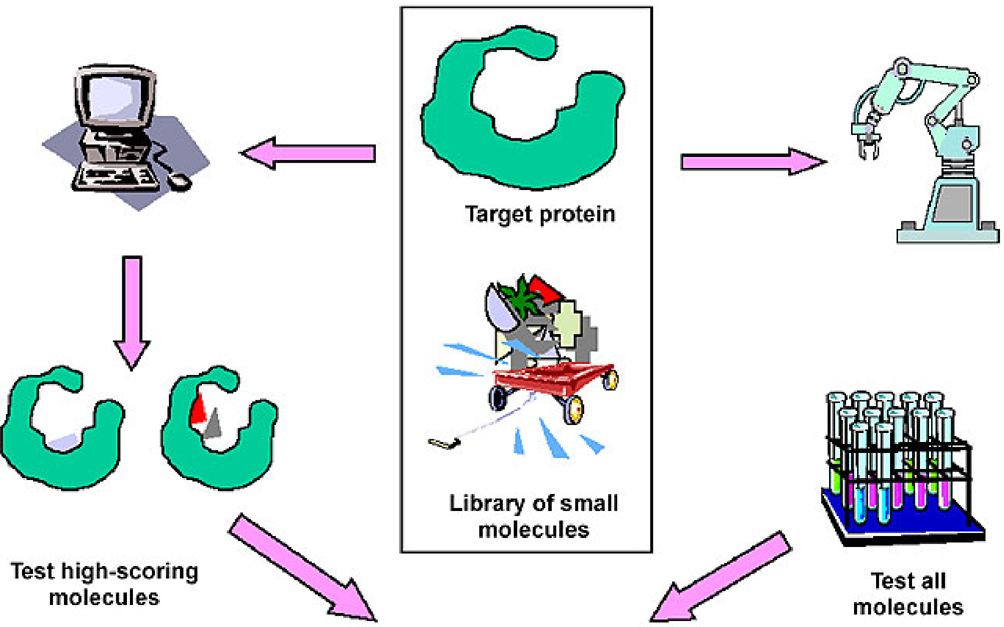

Figure 1. Docking (left) and high throughput screening (right) to discover new leads for drug discovery.

In docking a database of 10^5 to 10^6 small molecules, one cannot afford the time devoted to FEP nor can one afford the assumption that one will be able to compare similar molecules—the databases are purposefully diverse, often maddeningly so. Thus we must make breathtaking assumptions to calculate docking energies or, as they are often (and more honestly) called, docking scores. Our force-fields are inaccurate, the role of solvent is difficult to model (2), we do not relax our systems and therefore do irreversible work, charges are poorly modeled and don't polarize, and we massively under-sample. Getting absolute binding energies from docking calculations is currently well beyond the field. Even monotonic rankings are untrustworthy. Database docking is best considered a screening process, that in favorable circumstances can enrich possible true ligands and filter out unlikely ligands. Like experimental screens, docking screens are plagued by false positives and false negatives.

An appropriate question is why go through the bother of docking at all? Why not just use high throughput methods to *experimentally* screen a database of molecules? Surely this would avoid all the ambiguities of docking and discover more compounds to boot?

Here we consider three related projects ongoing in our laboratories at Northwestern University that consider several of these problems. To investigate how well docking might do at predicting new compounds and their geometries, we first consider a very simple binding site, one that

avoids many of the problems that one usually faces in docking. This cavity site in T4 lysozyme is in some senses a "perfect" docking site, since it is so simple. We then consider how well docking does when compared to a HTS project against the same target. These were studies performed in collaboration with Doman and colleagues at Pharmacia, and consider hit rates and quality of hits using both docking and HTS against the enzyme Protein Tyrosine Phosphatase 1B, a diabetes target (3). Finally, we turn to consider a class of promiscuous inhibitors that appear as "hits" from both virtual and high throughput screens. Through a series of biophysical experiments we seek to define a common mechanism of action for a broad range of small molecule non-specific "inhibitors" that have turned up over the years from screens. These nuisance compounds are among the biggest practical problems in using screening for drug discovery research.

A Cavity Binding Site in T4 Lysozyme.

In 1991, Matthews and colleagues introduced a cavity into the hydrophobic core of T4 lysozyme by the substitution Leu99→Ala (L99A) (4). This left a completely hydrophobic cavity of about 150 $Å^2$ in size. As it happened, this site was able to bind small, typically aryl, hydrocarbons in sizes that ranged from benzene, towards the lower end, to naphthalene towards the upper end (Figure 2). Through the work of Morton and Baase (5, 6), over 50 ligands were found that bound to this site, and nine of them were characterized crystallographically.

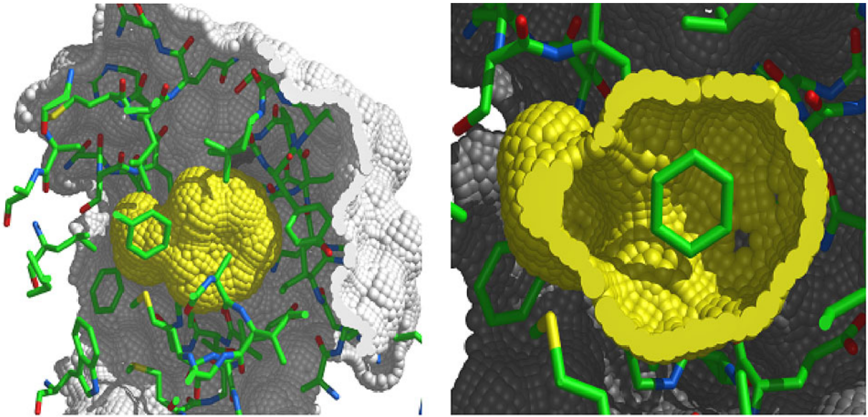

Figure 2. Two views of cavity site in the mutant T4 lysozyme L99A. Outer protein surface in gray, inner cavity surface in yellow. The right panel shows a cutaway of the site, revealing benzene bound in its crystallographic orientation.

We first asked how well docking the Available Chemicals Directory (ACD), which contained most of the characterized ligands for this site, would do at predicting known ligands, using the Northwestern University version of DOCK [Kuntz, 1982 #35; Ewing, 1997 #1107] (NWU

DOCK) (7, 9). As we moved from simple, steric-based scoring to more sophisticated energy and solvation-corrected methods, molecular docking was better and better at enriching known ligands from among the ~170,000 decoys in the database (Figure 3). The best enrichment came when we moved to calculating partial atomic charges and solvation energies for the database molecules using semi-empirical quantum-mechanics through the program AMSOL (10).

Having found that we could retrospectively reproduce known ligands for L99A, we turned to prospective prediction.

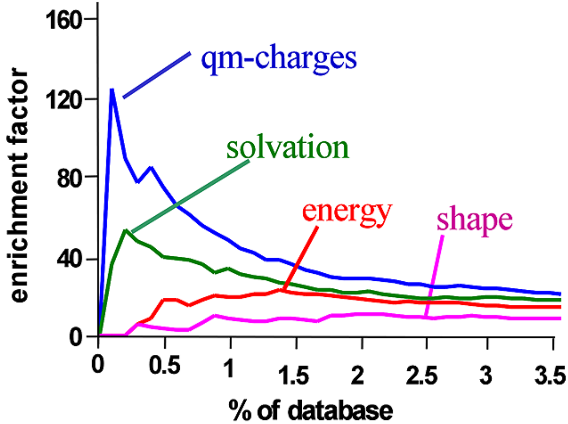

Figure 3. Enrichment plots for docking against the L99A hydrophobic cavity using different scoring functions (Wei et al., submitted for publication).

We substituted one of the hydrophobic residues that line the cavity, Met102, with a more polar glutamine (L99A/M102Q). X-ray crystallography suggested that this substitution introduced a single polar atom, the Oe1 of now Gln102, into the cavity surface. We re-docked the ACD against this slightly polar site, and looked for molecules that: a. scored better against L99A/ M102Q than they did against L99A; b. ranked better in the L99A/M102Q screen than they did in L99A screen; and c. were not observed to bind to L99A site experimentally. Seven molecules were picked and tested for binding; all seven were observed to bind to L99A/M102Q. Five of these were tested in detail using isothermal titration calorimetry (ITC), and were found to have dissociation constants in the 100 µM range (Table 1).

To investigate how well the predicted docked structure of these new compounds corresponded to experiment, the structure of the complexes of five of these compounds was determined by x-ray crystallography, to between 2.0 and 1.85 Å resolution. Before structure determination, predictions were sent to our collaborators in the Matthews lab (Larry Weaver & Walt Baase) to make it a fair test. For all structures, the docking predictions corresponded to the experimental result to with 0.4 Å rms (Figure 4).

Table 1. Binding data for L99A/M102Q

ligand	ΔTm (K)[a]	K_d (µM)[b]
3-methylpyrrole	2.1	160
3-chlorophenol	2.7	56
2-fluoroaniline	1.8	100
2,4-difluoroaniline	1.69	
phenole	2.25	91
2,4-difluorophenol	1.9	
3,5-difluoraniline	1.75	
toluene	0.5	160

[a] Binding measured from Tm upshift.
[b] ITC binding data.

In this simple site, molecular docking can predict novel ligands and do so with high geometric accuracy. Perhaps more importantly, the cavity sites L99A, L99A/M102Q, and other derivatives, provide good model systems for testing future developments in docking programs. Docking has advanced to a point where there is a need for model systems that allow both retrospective and prospective testing.

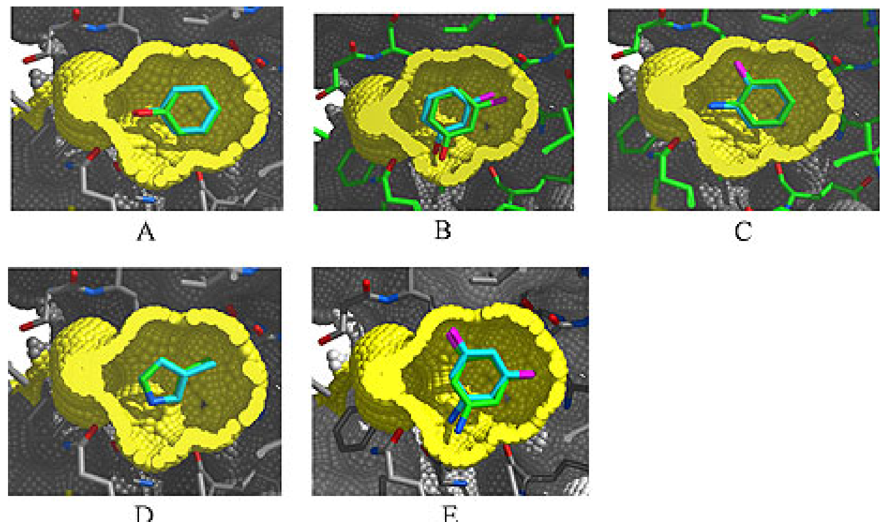

Figure 4. Correspondence between docked (carbons in cyan) and crystallographic configurations of novel ligands in the L99A/M102Q binding site. A: Phenol, B: 3-chlorophenol, C: 2-fluoroaniline, D: 3-methylpyrrole, E: 3,5-difluoroaniline.

Docking vs. High Throughput Screening

It's one thing to find that docking can make predictions in what amounts to a "toy" site, but how does it do against a real drug target, and how does it compare to the dominant tool used in the pharmaceutical industry for discovery research, high throughput screening?

This question cannot be answered definitively by any single project, on which caveats will always hang like scabby mendicants. In the spirit of comparing virtual to high throughput screening in as head-to-head manner as possible, we were pleased to collaborate with Doman and colleagues at Pharmacia in their effort to discover novel inhibitors of the Type II Diabetes target PTP1B. At Pharmacia, an in-house library of about 400,000 compounds was screened by HTS.

At Northwestern, about 250,000 commercially available molecules (most from the ACD) were screened using NWU DOCK against the structure of PTP1B (11). About 1000 high scoring compounds were selected by our group at Northwestern, and of these the Pharmacia group chose 365 to actually purchase and test. The results from these 365 compounds were compared to the results from the 400,000 compounds tested experimentally by HTS. All compounds were tested at Pharmacia by Pharmacia biochemists.

The hit rate resulting from docking was 1,700-fold better than the hit rate from HTS (Table 2) (3). More absolute inhibitors were found by testing 365 dock-derived molecules than were found from testing 400,000 compounds from HTS. Surprisingly, the dock-derived inhibitors were more drug-like than the HTS hits (Figure 5). Intriguingly, there was no overlap between the docking and the HTS hits, even at the chemical similarity level, when the two groups of hits were clustered. This last observation suggests that virtual and high throughput screening are complementary techniques; the high hit rate enhancement from docking, should it turn out to be general, suggests that virtual screening is not uncompetitive with HTS.

The thoughtful reader might ask themselves why so many HTS hits were non-drug like? There are several answers to this question, but among them is that many screening hits are artifactual. This is a horrible problem for early drug discovery, because these nuisance compounds can overwhelm true ligands that might exist in one's hit lists. The mechanistic bases of one class of these artifacts is the subject of our last section.

Table 2. Hit rates from docking and high throughput screening against PTP1B.

Technique	Compounds tested	Hits with IC_{50} < 100 μM	Hits with IC_{50} < 10 μM	Hit Rate
HTS	400,000	85	6	0.021%
Docking	365	127	18	34.8%

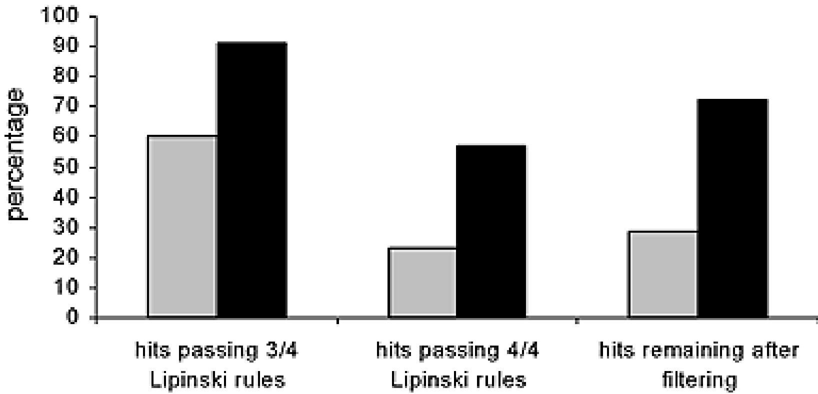

Figure 5. Drug like qualities of PTP1B HTS (diagonal lines) and docking (solid bars) hits, inhibiting at the 100 µM level. Filtering was performed at Pharmacia using internal rules (3).

PROMISCUOUS INHIBITORS FOR VIRTUAL AND HIGH THROUGHPUT SCREENING

We backed our way into this problem, not meaning to. We had undertaken a docking screen against AmpC b-lactamase, an enzyme with which we have a great deal of experience as an experimental system-enzymology, stability, and crystallography are all well in hand. We found tens of novel micromolar inhibitors for this enzyme, which was at first gratifying. To test specificity, we did counter screens against other enzymes including chymotrypsin, trypsin, dihydrofolate reductase (DHFR), malate dehydrogenase (MDH) and b-galactosidase. All of the b-lactamase inhibitors we had discovered turned out to be inhibitors, to varying degrees, of these other, unrelated enzymes (Table 3) (12).

We wondered how widespread this phenomenon of promiscuous inhibition was. We looked through the literature for virtual or HTS hits that looked, vaguely, like the ones we had seen for AmpC. Those that were commercially available we tested against our panel of model, out-group enzymes: AmpC, chymotrypsin, DHFR (or MDH) and b-galactosidase. Many of these compounds inhibited these model enzymes (Table 3).

The inhibition properties were unusual. All of these molecules showed time dependent, but apparently reversible inhibition. Inhibition was very sensitive to ionic strength. Wondering if these compounds were acting as denaturants, we looked to see if urea or guanidinium improved inhibition. Just the opposite happened, inhibition got worse. Similarly, inhibition was very sensitive to the presence of albumin (BSA), which at the 1mg/ml level dramatically attenuated inhibition.

Table 3. Nonspecific inhibitors discovered by screening (12).

Structure	Original Target(s)	IC_{50} (µM)			
		β-lactamase	**Chymotrypsin**	**cDHFR**	**β-gal**
	0.5 β-lactamase[a]	0.5	2.5	5	15
	5 β-lactamase[a]	5	25	35	90
	5 β-lactamase[a]	5	15	N.D.	N.D.
	8 malarial protease	10	55	70	180
	7 pDHFR	10	50	60	300

Table 3, continued. Nonspecific inhibitors discovered by screening (12).

Structure	Original Target(s)	β-lactamase	Chymotrypsin	cDHFR	β-gal
		IC$_{50}$ (μM)			
	80 pDHFR	50	25	N.D.	600
	50 HIV Tar RNA	10	90	N.D.	600
	3 TS; 30 kinesin	3	11	20	200

Table 3, continued. Nonspecific inhibitors discovered by screening (12).

			IC50 (μM)			
Structure	**Original Target(s)**		**β-lactamase**	**Chymotrypsin**	**cDHFR**	**β-gal**
	20[b] insulin receptor	7.5 kinesin	16	50	N.D.	80
	5.2 VEGF	10.0 IGF-1	6	30	30	55
	25 farnesyltransferase		3	9	25	150

Table 3, continued. Nonspecific inhibitors discovered by screening (12).

Structure	Original Target(s)		IC$_{50}$ (μM) β-lactamase	Chymotrypsin	cDHFR	β-gal
	15^{c} gyrase		18	100	150	320
	1 prion	30.4 TIM	3.9	40	0.4	100
	17 eNOS	24 nNOS	7	60	N.D.	N.D.
	3.8 P13K	11.0 integrase	4	100	N.D.	220

[a]Our unpublished observations. [b]K_d. [c]maximal non-effective concentration. cDHFR, chicken DHFR; β-gal, β-galactosidase; pDHFR, *Pneumocystis carinii* DHFR; TS, thymidylate synthase; VEGF, vascular endothelial growth factor receptor tyrosine kinase; IGF-1, insulin-like growth factor receptor tyrosine kinase; TIM, triosephosphate isomerase; eNOS, endothelial nitric oxide synthase; nNOS, neuronal nitric oxide synthase; PI3K, phosphoinositide 3-kinase; N.D., not determined

The experiment that put us onto the right way of thinking (after months of befuddlement) was increasing the enzyme concentration ten-fold, while leaving the inhibitor concentration untouched. Since this involved raising β-lactamase from 1 nM to 10 nM, and left the average inhibitor at 10 μM, this should have had no effect on inhibition levels. But instead it attenuated them dramatically. We wondered if the inhibitory species was not a single small molecule, or even two or three, but an aggregate of thousands.

If an aggregate was responsible for inhibition, it should measurable by direct methods. Using dynamic light scattering (DLS) we found that in common buffers these "inhibitors" formed particles of 50 to 450 nm in diameter-almost two orders of magnitude larger than the enzymes that they inhibited. These aggregates were also observed by transmission electron microscopy (TEM). These results are consistent with the hypothesis that these promiscuous inhibitors are acting by forming an aggregate in solution, and that it is these aggregates that inhibit enzymes non-specifically.

In a final experiment, we turned to compounds from the Pharmacia screening database, and asked whether promiscuous, aggregating inhibitors occurred among them. Of the thirty compounds we investigated, twenty were promiscuous, aggregate-forming inhibitors.

In summary, we propose that a single mechanism of action underlies the inhibition pattern of many non-specific inhibitors that have been, and still are being, discovered by virtual and high throughput screening. A burning question to many is how one might recognize such inhibitors in advance, using chemical similarity techniques. This is a question that we cannot at this time answer - the compounds that show this behavior are only very loosely similar, and there are exceptions to every rule we have considered. What is clear is that there are unambiguous experimental tests that can identify such aggregating inhibitors. Such diagnostic experiments should be routinely performed before carrying forward a discovery project.

REPRISE: HITS, LEADS AND ARTIFACTS FROM DOCKING AND HIGH THROUGHPUT SCREENING

We return to the question posed at the beginning of this essay: why do docking at all, why not just screen experimentally? In well-controlled cases, docking can propose sensible novel ligands and can do so with some accuracy.

The cavity sites in lysozyme provide model systems for testing developments in docking programs, our own and those of others. Although the right head-to-head comparison between

docking and HTS has yet to be performed (in PTP1B we used different databases), the experience with PTP1B (3) and with several other systems (13) suggests that structure based efforts in discovery may be considered as alternatives to HTS.

Among the largest challenges facing both docking and HTS is that of promiscuity through aggregation. Small molecules have the option not only of binding to a receptor, but also of aggregating together. Such aggregates inhibit many enzymes non-specifically. In addressing this problem, docking and HTS are allies. Both techniques will gain much from eliminating these promiscuous inhibitors from their hit-lists (14, 15). An encouraging aspect to emerge from these early studies is that there are clear diagnostic tests for these inhibitors. These will allow investigators to eliminate aggregating inhibitors early and thereafter to focus on the true ligands that emerge from structure-based methods, which hold such promise for lead discovery.

References

[1] Abagyan, R. & Totrov, M. (2001). *Curr. Opin. Chem. Biol.* **5**:375-82.

[2] van Gunsteren, W. F. & Berendsen, H. J. C. (1990). *Angew. Chem. Int. Ed. Engl.*, **29**:992-1023.

[3] Doman, T. N. et al. (2002). *J. Med. Chem.*,in press.

[4] Eriksson, A. E., Baase, W. A., Wosniak, J. A., Matthews, B. W. (1992). *Nature* **355**:371-373.

[5] Morton, A. & Matthews, B. W. (1995). *Biochemistry* **34**:8576-8588.

[6] Su, A. I. et al. (2001). *Proteins* **42**:279-293.

[7] Lorber, D. M. & Shoichet, B. K. (1998). *Protein Sci*. **7**:938-950.

[8] Shoichet, B. K., Leach, A. R., Kuntz, I. D. (1999). *Proteins* **34**:4-16.

[9] Lorber, D. M., Udo, M. K., Shoichet, B. K. (2002). *Protein Science*, in press.

[10] Li, J. B., Zhu, T. H., Cramer, C. J., Truhlar, D. G. (1998). *Journal of Physical Chemistry A* **102**:1820-1831.

[11] Puius, Y. A. et al. (1997). *Proc. Natl. Acad. Sci. USA*. **94**:13420-5.

[12] McGovern, S. L., Caselli, E., Grigorieff, N., Shoichet, B. K. (2002). *J. Med. Chem.* **45**:1712-1722.

[13] Paiva, A. M. et al. (2001). *Biochim. Biophys. Acta* **1545**:67-77.

[14] Rishton, G. M. (1997). *Drug Discov. Today* **2**:382-384.

[15] Roche, O. et al. (2002). *J. Med. Chem*. **45**:137-142.