11

Docking ligands to proteins

BRIAN K. SHOICHET

1. Introduction

A docking method seeks to find ways of fitting two molecules together in favourable configurations. This simple goal is actively pursued because of its central role in two very practical areas of research: structure-based drug discovery and the molecular modelling of biological function. Sadly, our understanding of how molecules recognize one another is still rather poor. Both the importance of the field, and our limited grasp of its intricacies, have lead to a great number of proposals for solving what is often called 'the docking problem.' None of them is complete. For clarity's sake, I will discuss ways to approach docking problems as if there was a right way, and we knew what we were doing. Sometimes there isn't and we don't. Buyer beware.

Having begun with a caveat, I hasten to add that we have learned a great deal in the last ten years. We can automatically locate regions on proteins that are candidates for ligand binding ('hot-spots'). The components of a ligand–protein complex can be reliably and accurately reassembled, if they are in the proper (bound) conformations. Mounting evidence suggests that docking can discover novel lead inhibitors that bind in the micromolar concentration range, though many difficulties remain (1,2). Structure-based elaboration of novel leads can produce inhibitors that bind in the nanonmolar and picomolar concentration ranges (3,4).

Ambiguities attend like surly waiters on most docking calculations. To overcome this difficulty, I will emphasize problems for which docking can suggest experimentally testable hypotheses. Wherever possible, I shall outline controls for docking calculations. These are often possible and always worthwhile.

I will familiarize you with approaches to three common goals of docking programs:

- fitting a small molecule into a protein
- docking two proteins together
- novel inhibitor discovery using molecule databases

For any of these you will need:

- an atomic resolution structure of your protein (*Table 1*)
- an atomic resolution structure of your ligand or ligands (*Table 1*)
- a scientific-style computer workstation with high resolution graphics
- a docking computer program (*Table 2*)

2. Docking programs

There are a number of different docking computer programs available to the reader. I will illustrate most examples using just one of these, the Dock program, introduced by Kuntz and colleagues (5–7). It is appropriate to begin, however, with a survey of the field.

All docking programs must solve three fundamental problems:

- where, in a relatively large protein site, to fit the ligand
- what conformations of ligand and protein best complement each other
- how to evaluate the energies of the various complexes

Table 1. Molecular databases

Type of molecule	Source	Comments
Protein or DNA	Protein Data Bank (PDB) (47) Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, USA e-mail: pdb@bnl.gov	Atomic resolution protein and DNA structures, mostlyfrom X-ray and NMR experiments
Small molecule	Cambridge Structural Database (CSD) (38) Dr Olga Kennard FRS, Cambridge Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK	X-ray determined structures 10^5 compounds
Small molecule	Available Chemical Directory (ACD), MDL Information Systems, Inc, San Leandro, CA 94577, USA	Computer-generated structures commercially available, 1.5×10^5 compounds
Small molecule	Marketing Department, Chemical Abstracts Service, 2540 Olentaey River Road, PO Box 3012, Columbus, OH 43210–0012, USA	Computer-generated structures of literature compounds, approx. 6×10^6 compounds
Small molecule	Triad and Iliad Dr Paul A. Bartlett, Department of Chemistry, University of California, Berkeley, CA 94720, USA; e-mail: paul@fire.cchem.berkeley.edu	Computer-generated structures hydrocarbon scaffolds, tricyclic and acyclic, respectively, 4×10^5 and 1×10^5 compounds

Table 2a. Some docking programs

Program	Algorithm	Small ligand docking?	Protein–protein docking?	Database search?	Ligand design	Conformation search	Scoring
Rendezvous	Grid search	Yes	Yes	No	No	Yes	Force field
Soft docking	Grid search	Yes	Yes	No	No	No	Polar/apolar contact
Grow	Fragment	Yes	No	No	Yes	Yes	Force field
Concepts	Fragment	No	No	No	Yes	Yes	Force field
Ludi	Fragment	Yes	No	Limited	Yes	Possible	Force field
Hook	Fragment	Possilbe	No	No	Yes	Possible	Force field
AutoDock	Kinetic	Yes	Yes	No	No	Yes	Force field
Dock	Descriptor	Yes	Yes	Yes	No	No	Force field
Caveat	Descriptor	Yes	No	Yes	No	No	Steric fit
Clix	Descriptor	Yes	ND	Yes	No	No	Force field

Table 2b. Corresponding authors for docking programs

Program name	Corresponding author
Rendezvous Batiment	Joel Janin, Universite Paris-Sud, Laboratoire de Biologie Physicochimique, 433, 91405 Orsay Cedex, France Shoshana Wodak, Universite Libre de Bruxelles, 67 Rue des Chevaux, Rhode-St-Genese, 1640, Belgium
Soft docking	Sung-Hou Kim, Department of Chemistry, University of California, Berkeley, CA 94720, USA
Grow	Joeseph B. Moon or Jeffrey Howe, The Upjohn Company, Computational Chemistry, 301 Henrietta Street, Kalamazoo, MI 49001, USA
Concepts	David Pearlman, Vertex Pharmaceuticals Inc., Cambridge, MA 02139–4211, USA
Ludi	Hans-Joachim Bohm, BASF AG, Central Research, D-6700 Ludwigshafen, Germany Distributed by Biosym Technologies Inc., 1515 Rt 10, Suite 1000, Parsippany, NJ 07054, USA
Hook	Martin Karplus, Department of Chemistry, Harvard University, Cambridge, MA 02138, USA
AutoDock	Arthur J. Olson, The Scripps Research Institute, Department of Molecular Biology, MB5, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA
Dock	Irwin D. Kuntz, Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143, USA
Caveat	Paul Bartlett, Department of Chemistry, University of California, Berkeley, CA 94720, USA; e-mail: `paul@fire.cchem.berkeley.edu`
Clix	M. C. Lawrence, CSIRO, Division of Biomolecular Engineering, 343 Royal Parade, Parvill, Victoria, 3052, Australia

Solving these problems rigorously requires elaborate energy calculations (8) and consideration of many more conformational and configurational degrees of freedom than is now computationally feasible for a docking method. To get reasonable answers in reasonable amounts of time, docking algorithms simplify the problem. Common approximations include modelling explicit waters by a dielectric continuum, using enthalpy as a proxy for free energy, and treating intrisincally flexible ligands and proteins as rigid objects.

Docking programs may be grouped into four families, depending on how

they address these problems, and what simplifications they use (2). These families are: descriptor methods, grid methods, fragment methods, and kinetic methods (*Table 2*). All begin with the structure of a protein and that of a ligand, or ligands, and ask: where does the ligand bind, and, for inhibitor discovery, which ligand binds best?

2.1 Docking with descriptors

The protein is first analysed for regions of likely complementarity. These 'hot-spots' are places on the protein surface where a ligand atom might fit well; they describe the binding region. The idea is to match ligand atoms to receptor hot-spots, and by so doing generate orientations of the ligand in the protein. For any given ligand many orientations are sampled. Dock, for instance, might try several thousand orientations for a small ligand while a protein ligand might sample tens or even hundreds of thousands of orientations. Orientations are evaluated for goodness of fit, typically through use of a molecular mechanics-type energy function. Though not exhaustive, descriptor methods are fast and can often sample densely in a particular region of the protein. They rely on being able to identify the hot-spots well, a point to which I shall return. Most descriptor methods treat ligands and proteins as rigid objects, though this is not inherent to the algorithms (see for instance ref. 9). Descriptor programs have been used extensively for single ligand docking and for inhibitor discovery (Dock, Caveat, and Clix programs).

2.2 Grid search

Grid searches fit the ligand into the protein by rotating and translating the ligand in discrete steps while holding the protein rigid. Grid searches have the advantage of always getting to the neighbourhood of the correct solution, which cannot be said of the descriptor matching methods. They often take a long time. The accuracy of grid methods is limited to the resolution of the step size used in the search, but the higher the accuracy the longer the search. Most grid methods treat the ligand and protein as rigid objects, though some allow for conformational relaxation (10). These programs have been used extensively for single ligand docking problems (10,11,49).

2.3 Fragment

Fragment methods identify regions of high complementarity on a protein surface by docking functional groups independently into proteins. By breaking ligands into fragments, many of the configurational and conformational issues in docking disappear. This is done at the expense of connectivity information, which fragment methods can in principle gain back by reconnection algorithms at the end of the calculation. Fragment methods can be used for molecular elaboration of existing inhibitors. They are useful for novel

inhibitor design when working with molecules whose synthetic chemistry is as modular as the computer-generated fragments, such as peptides and oligonucleotides. The Concepts (12), Ludi (13), and Grow programs (14) are good examples of this idea, with the latter having been experimentally tested.

2.4 Kinetic

Kinetic docking techniques sample potential surface, using molecular dynamics or simulated annealing, to fit ligands to proteins. Like the grid methods, kinetic approaches are conceptually appealing because they mimic how we imagine ligands encounter proteins in solution. An advantage of kinetic methods is that they merge the configurational and conformational aspects of the docking problem smoothly. A disadvantage is that the complex topography and multiple minima of molecular potential surfaces often lead to long run times and minima traps. They have been used extensively in single ligand docking, see especially the method of Goodsell and Olson (15) and its application by Stoddard (16), and the recent method of Abagyan (50).

3. Preparing your input

All docking projects begin with the structure of the ligand and the protein. One has important decisions to make regarding the reliability of these structures and which regions on them to target most heavily.

3.1 Using protein structures

Atomic resolution protein structures have three sources: X-ray crystallography, nuclear magnetic resonance (NMR), and homology model building. The last of these is described in Chapters 6 and 7; my only caution here is that model-built structures are the riskiest to employ, but see Ring *et al.* (17) for their use in inhibitor design, and Bax *et al.* on protein–protein docking (18). I will focus on what regions of experimental structures require consideration before docking.

While experimental structures give you coordinates for most or all atoms in a protein, they are not equally trustworthy or useful. For example, most X-ray structures include bound waters and ions along with protein atoms. Many of these waters and ions, unlike the protein atoms, are displaceable by a ligand. Some, however, are tightly bound and are best considered constitutive parts of the protein. Most docking programs have no way of knowing which is which; you must decide which of these displaceable groups to include and which to exclude before the docking program sees your protein. Another common problem is that some protein atoms are ill-determined. Some appear to be poorly constrained in the structure (high thermal factors), others are ill-resolved (poor electron density, few NOE peaks). You must decide which atoms are reliable before beginning a docking calculation. In happy

circumstances the structures of several different protein conformers will be available to you; which should you use?

Eliminate as many waters and ions as possible to have the largest protein or protein ligand surface possible to dock to. In most docking calculations performed in the Kuntz group in the last several years, *all* waters and ions were eliminated. This is the simplest approach, but it has several times lead to problems (7,19). Nevertheless, I recommend deleting all waters and ions in a potential binding site except those that fulfil at least one of the following criteria:

- they occur in three or more different states of the protein structure (e.g. different conformers, different crystal forms, different ligands bound)
- they have unusually low thermal factors
- they have three or more polar interactions with protein atoms
- they have been implicated in the mechanism of the protein

Poorly determined protein atoms can lead to the same sorts of difficulties as waters and ions; they preclude ligand binding to the volume that they occupy. Nevertheless, we usually include all the atoms present in the published structure. It is occasionally useful to remove the following atoms from the binding site of a protein (20):

- atoms with occupancies of zero
- atoms with thermal factors greater than 60 $Å^2$
- poorly constrained regions of NMR structures showing large deviations from model to model

Zero occupancy atoms can be removed without much angst, but the thermal factor and NMR constraint criteria are riskier. Deleting protein atoms should be resorted to only when one suspects that they are interfering with a docking calculation (20).

There are occasionally several structures of the same protein. These may have been solved in complexes with different ligands, or in different crystal forms. Use the experimental form that most closely resembles the state that you expect the protein to be in when it binds the ligand. For novel inhibitor discovery consider using *all the* protein conformations, sequentially. This will not seriously increase the computation time and will improve one's chance of getting a highly complementary ligand–protein fit. Any well-determined protein structure is a good target for inhibitor discovery.

3.2 Targeting binding sites

Any docking project will benefit from knowing, in advance, what parts of the protein surface are most likely to fit the ligand, and what parts can be ignored. For a descriptor method, such as Dock, such information is critical.

3.2.1 Sphere descriptors

There are two computational ways to identify these hot-spots: geometrically and energetically. An experimentally-based method uses the location of a bound ligand. Dock has traditionally relied on a geometric method that maps clefts or ridges on the surface of a protein (5) (*Protocol 1*), though ligand coordinates have also been used to map the site (7). Spheres are calculated to fit, but not intersect, local pockets in the molecular surface of a protein (*Figure 1*). Such pockets are good places to try to fit a ligand atom in docking;

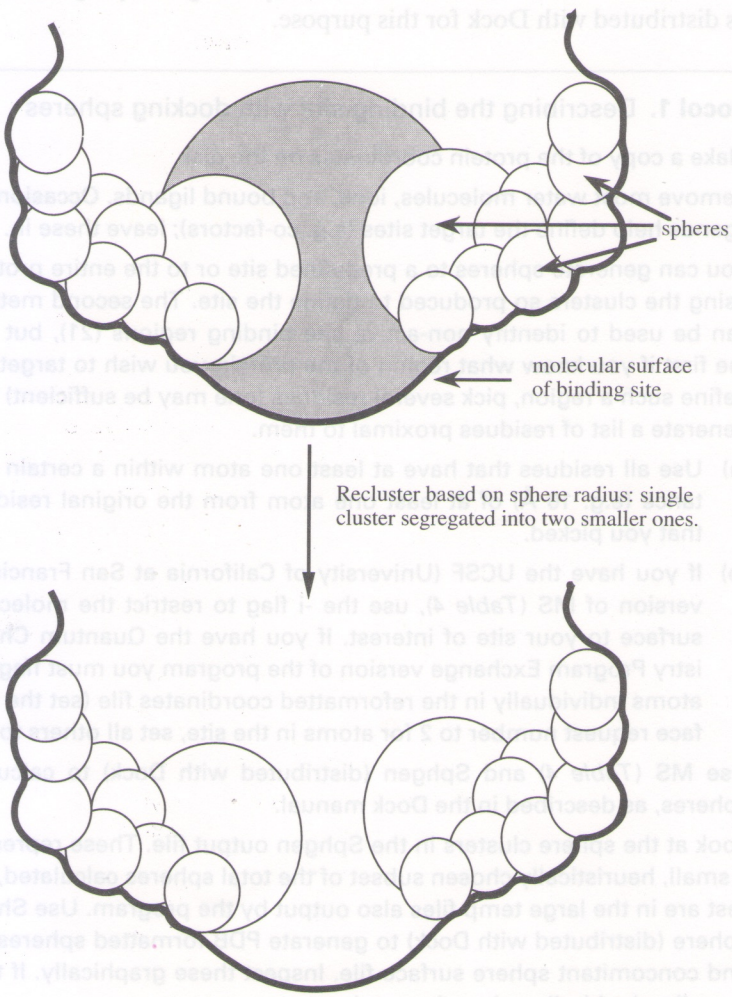

Figure 1. Dock spheres and the subclustering algorithm. Subclustering splits up large binding sites by segregating groups of spheres based on their radii. Here, a large radius sphere (shaded) is removed from the sphere set, creating two subsites.

they are a positive image of the negative mould presented by a protein surface to a ligand. The spheres are produced as intersecting clusters—there will usually be several clusters on the surface of a protein, each defines a putative binding site in that surface. A typical size for a binding site cluster used in docking is 30–60 spheres. The more spheres in a cluster the larger the site that Dock will explore and the longer it will take to do so. Use the smallest number of spheres, or any type of descriptor, as possible. If there are more than 100 spheres in your putative binding site, divide the clusters into smaller groups of spheres and dock to each one separately. Reclustering into smaller sphere sets can be done with an editor, or by using the program Cluster, which is distributed with Dock for this purpose.

Protocol 1. Describing the binding site with docking spheres

1. Make a copy of the protein coordinates on the disk.
2. Remove most water molecules, ions, and bound ligands. Occasionally ligands help define the target sites (e.g. co-factors); leave these in.
3. You can generate spheres to a predefined site or to the entire protein, using the clusters so produced to define the site. The second method can be used to identify non-active site binding regions (21), but use the first if you know what region of the protein you wish to target. To define such a region, pick several residues (one may be sufficient) and generate a list of residues proximal to them.
   (a) Use all residues that have at least one atom within a certain distance (e.g. 10 Å) of at least one atom from the original residues that you picked.
   (b) If you have the UCSF (University of California at San Francisco) version of MS (*Table 4*), use the -i flag to restrict the molecular surface to your site of interest. If you have the Quantum Chemistry Program Exchange version of the program you must flag the atoms individually in the reformatted coordinates file (set the surface request number to 2 for atoms in the site, set all others to 1).
4. Use MS (*Table 4*) and Sphgen (distributed with Dock) to calculate spheres, as described in the Dock manual.
5. Look at the sphere clusters in the Sphgen output file. These represent a small, heuristically chosen subset of the total spheres calculated, the rest are in the large temp files also output by the program. Use Showsphere (distributed with Dock) to generate PDB formatted spheres file and concomitant sphere surface file. Inspect these graphically. If they describe the binding site adequately, stop.
6. If the sphere centres are unsatisfactory, consider the larger set also output by Sphgen.

(a) Convert the temp files to a spheres file with Tosph (distributed with Dock). There will typically be a few thousand spheres in this file.

(b) Using an editor and a graphics program, flag several of the spheres that are in hot-spots on the protein surface.

(c) Use Cluster to calculate a new sphere set that is biased towards the regions of the surface that you have flagged (see Dock manual).

7. For clusters containing more than 100 spheres, delete large radius spheres on the edge of the site. If this does not remove enough spheres, or you have difficulty choosing which spheres to delete, the Cluster program will do this automatically. Often, several subclusters are produced that can be docked to sequentially. Reclustering spheres is essential for protein ligands (20).

3.2.2 Alternative site description methods

High potential energy regions on the surface of a protein can also locate hot-spots. Such high potential regions will complement a ligand atom of opposite atomic properties. Thus an electropositive region of the protein will interact favourably with a negatively charged ligand atom, a hydrogen bond donor will do the same with a hydrogen bond acceptor, and so on. Protein surfaces may be mapped for high energy regions using methods such as those listed in *Table 3*. The Grid program has been used most extensively for this purpose (22). The same rules apply to energy-based descriptors as to spheres: one

Table 3. Programs to determine high potential energy regions on protein surfaces

Energy surface mapped	Program	Corresponding author or distributor
Molecular mechanics	Grid	Peter Goodford, The Laboratory of Molecular Biophysics, The Rex Richards Building, University of Oxford, Oxford OX1 3QU, UK
Molecular mechanics	Chemgrid	Irwin Kuntz, Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143, USA
Electrostatic	Delphi	Barry Honig, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 168th Street, New York, NY 10032, USA Also distributed by Biosym Technologies Inc., 9685 Scranton Road, San Diego, CA 92121, USA
Hydrogen bonding	Hsite	Philip Dean, Department of Phamacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK
Hydrophobicity	Hint	Donald Abraham, Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298–0540, USA

needs enough high potential points to map adequately the site, but having too many leads to excessive computation times.

If one knows the structure of a ligand–protein complex, and the ligand is bound in the site you wish to target, then it is a simple matter to use the ligand coordinates themselves as the hot-spots with which to dock other ligands. Assign types to ligand atom descriptors in the same manner as spheres to improve run time and selectivity. Ligand atom descriptors can be usefully combined with computed descriptors to expand the limits of the binding site.

We find it useful to combine the sphere and energy approaches to site description. Spheres are calculated as usual, but are 'labelled' according to chemical potential. A sphere in an electropositive region is 'labelled' positive, one in an electronegative region is 'labelled' negative, and so on. We only allow matches between spheres and atoms with complementary chemical labels. This speeds the docking calculation tenfold and improves selectivity (23).

3.2.3 Experimental constraints

A final site description issue is that of using external constraints in the docking calculations. Experimental data, such as mutagenesis experiments, often suggest that a certain residue is directly involved in ligand binding. Including this information in a docking calculation can dramatically improve selectivity (24) and run time (25). Dock allows one to flag critical protein spheres and ligand atoms and insist that these be involved in any docked complex.

3.3 Scoring docked complexes: energy calculations

Most docking methods calculate an interaction energy to evaluate and rank ligand–protein complexes. This energy usually includes some or all of the terms found in molecular mechanics force fields, such as dispersion, polar, hydrogen bonding, and occasionally hydrophobic interactions (Chapter 10). To calculate the interaction energy (E_{int}) of a given docked complex one needs to know the potential energy of the protein (P) and the atomic properties of the ligand (A) at all ligand atom positions (i):

$$E_{\text{int}} = \sum_{i=0}^{n} P_i A_i \qquad [1]$$

The potential P is usually calculated once, and stored on a lattice of points in memory during the docking calculation (see Dock manual for details). Every docked ligand orientation is fit on to this lattice and its energy calculated by summing the product of the lattice point energies and the atomic properties. Lattice energies can be calculated with programs such as Grid, Delphi, and Chemgrid (*Table 3*).

Ligand atomic properties, such as partial charges and van der Waals parameters, can be calculated using a molecular editor such as exist in pro-

Table 4. Some useful computer programs for docking projects

Purpose	Program name	Corresponding author or distributor
Macromolecular graphics	MidasPlus	Tom Ferrin, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143–0446, USA e-mail: tef@socrates.ucsf.edu
Macromolecular graphics	Grasp	Anthony Nichols, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 168th Street, New York, NY 10032, USA e-mail: nicholls@cuhhca.hhmi.columbia.edu
Molecular surface	MS	Quantum Chemistry Program Exchange, Department of Chemistry, University of Indiana, Bloomington, IN 47405, USA Also distributed with MidasPlus
Molecular mechanics	AMBER	Peter Kollman, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143–0446, USA e-mail: pak@socrates.ucsf.edu
Molecular mechanics	CHARMM	Martin Karplus, Department of Chemistry, Harvard University, Cambridge, MA 02138, USA
Molecular modelling, editing and graphics	Sybyl	Tripos Associates, Inc., 16299 S. Hanley Road, Suite 303, St. Louis, MO 63144, USA
Molecular modelling, editing and graphics	MacroModel	Office of Science and Technology Development, Columbia University, New York, NY 10027, USA
3D ligand structures from 2D representations database conversions	Concord	Tripos Associates, Inc., 16299 S. Hanley Road, Suite 303, St. Louis, MO 63144, USA
Multiple conformations for small molecules	Cobra	Oxford Molecular Ltd., Terrapin House, South Parks Road, Oxford OX1 3UB, UK
Ligand solvation energy	Hydren	Alexander Rashin, Biosym Technologies Inc., 1515 Rt 10, Suite 1000, Parsippany, NJ 07054, USA
Small molecule database, substructure searching	Isis	MDL Information Systems, Inc, San Leandro, CA 94577, USA
Substructure searching	Aladdin	Daylight Chemical Information Systems, 18500 Von Karman Avenue 450, Irvine, CA 92715, USA

grams like Sybyl and MacroModel (26) (*Table 4*). One typically needs at least partial atomic charges. If the ligand is a nucleotide or peptide such charges and parameters exist in parameter files distributed with programs such as AMBER, CHARMM, or GROMOS (*Table 4*) or can be found in the literature. Small molecule parameters, especially hydrophobicities, can often also be found in the literature (27), but one must usually calculate charges oneself.

4. Docking a small molecule to a protein

We begin with the structure of a protein and a small molecule (generally less than 50 heavy atoms) and ask: where best does the ligand fit? *Protocol 2* summarizes the preparatory steps we have just considered.

Protocol 2. Preparatory steps before docking

1. Remove appropriate waters and ions.
2. Calculate a potential energy lattice for the protein.
3. Locate descriptor positions on the protein surface (spheres, ligand atoms, points of high potential).
4. Calculate or look up ligand parameters such as partial atomic charge.
5. Calculate alternate conformations for the ligand, if desired (Cobra (28), Sybyl, or MacroModel programs, *Table 4*).

4.1 Controls

Before beginning the docking calculation, perform the following controls, if possible:

(a) Calculate an interaction energy for a known complex of a ligand with your protein. Map the ligand atomic positions from the complex on to the lattice (subprograms to do this are often distributed with the energy potential program) to find the energy (see Equation 1). If the energy is positive (poor fit), you have probably made a mistake in your potential energy calculation or your ligand parameters (*Protocol 3*).

(b) Use Dock to reproduce the complex between a known ligand and the protein, using their bound conformations. This will allow you to see if your docking parameters (below) are sensible. One should be able to reproduce a known complex to better than 1 Å root mean square deviation (rmsd) (*Protocol 4*).

If a complex structure is unavailable, consider doing these controls on another protein–ligand system. At least you will learn whether the general steps you have taken are correct.

Protocol 3. Insuring that a known ligand–protein complex has a negative interaction energy

The purpose is to test your ligand parameterization and lattice energies. Begin with the coordinates of a known ligand–protein complex.

Method

1. Does the ligand stericaly fit the protein site? Use the program Score-opt2 (distributed with Dock) to test this. If there is a close contact between the ligand and the protein, consider reducing the close contact limits and recalculating the contact map. Alternatively, delete poorly determined protein or ligand atoms if these are responsible for

the contact. Insure that the close contact does not involve a water, ion, or ligand inadvertently left in the site.

2. Does the ligand electrostatically fit the protein site? Use Scoreopt2 again, this time with the Delphi phi map (see Dock manual). If the interaction energy is positive, examine the ligand partial atomic charges and the protein potentials at the ligand atom positions. Common problems include:
   (a) Lack of hydrogen atoms on ligand and protein hetero-atoms.
   (b) Mis-parameterized protein residues (check the Delphi output to insure that the overall charge on the protein is sensible and an integer).
   (c) Waters, ions, or ligands included in the site for the electrostatic calculation but not the Distmap calculation.
   (d) Mis-parameterized or non-parameterized, and therefore unrecognized, waters, ions, or ligands.
3. The same approach is appropriate when using force field scoring (Dock manual).

4.2 The docking calculation

Docking programs often involve choosing values for numerous adjustable variables. Dock has nine important ones (*Table 5*); it is unfortunately necessary to play with these numbers in many cases (see the Dock manual). The goal is to sample the site with the ligand densely enough to get the correct answer, but not so densely that the calculation takes forever. This is usually possible with a small molecule. For rigid body docking, expect a calculation time of less than one hour on a Silicon Graphics Indigo 2 or similar computer. Use external constraints, such as those from mutagenesis or kinetic experiments, if you can. Try to generate more configurations of the ligand in the site to see how robust the docking solution is energetically and geometrically. Rigid body energy minimization within Dock can improve the ability of the program to find favourable configurations (29).

Protocol 4. Single ligand docking control

Can you reproduce the experimental configuration of a known ligand–protein complex? Use *Protocol 3* to insure that you have parameterized the ligand and mapped the protein potential correctly.

Method

1. Begin with the parameters in *Table 5* to dock the ligand to the site. If Dock produces configurations within 1 Å rmsd of the bound structure

Protocol 4. *Continued*

the control is successful. If you wish to perform a database search calculation, pay attention to the number of orientations sampled (NMATCH, reported in the OUTDOCK file).

2. To reduce the number of matches, decrease the bin sizes. If you have trouble getting a good correspondence between search time and reproduction of the experimental structure, try lowering *dislim* or increasing *nodlim*. Using rigid body minimization with limited sampling is another sensible strategy (29).
3. If no configurations close to the experimental complex are produced, experiment with the following Dock input parameters:
   (a) Increase the number of matches sampled by increasing the bin sizes.
   (b) If you have had focusing turned off, turn it on.
   (c) Increase the number of allowed close contacts in focusing (*fctbmp*, *Table 4*). For protein–protein docking, an *fctbmp* of as high as 15 is not unwarranted. For small molecules *fctbmp* should not exceed 3.
   (d) Try increasing *dislim* or decreasing *nodlim*.
4. If the experimental configuration is still not found, look for errors in your site description:
   (a) Do the protein descriptor centres (e.g. sphere centres) overlap your ligand atoms? If fewer than six do so, consider recalculating your spheres.
   (b) If you have labelled your spheres and atoms by chemical type, insure that the types you have chosen are complementary.
   (c) Check the format of your sphere and atom input files. Are the correct numbers of spheres and atoms being read? Insure that the longest distances between sphere and atom points are sensible (OUTDOCK file).
5. Is the ligand, or the ligand sphere set for protein ligands, bigger than the protein sphere set? Consider using fewer ligand atoms or spheres to match with protein descriptors, while continuing to fit the entire set of ligand atoms into the protein site. Create a separate, sphere formatted file using the atom coordinates as descriptor centres. Redock with only a subset of these descriptors (see Dock manual). For protein–protein docking, consider reducing the size of your ligand sphere set. Alternately, consider increasing the number of spheres in the protein cluster.

It is impossible to know whether the docking is correct without experimental verification. When X-ray or NMR experiments are impractical, consider

Table 5. Dock version 3.0 database search: sample input

Variable	Value	Meaning
versn	2	Scoring scheme (2 = electrostatic and contact)
mode	search	Single ligand docking or database search
clufil	file	File containing cluster
nclus	3	Sphere cluster—can search multiple sites
mapnam	file	File containing contact grid
dislim nodlim ratiom **lownod**	**1.5 5 0. 5**	Distance tolerance on atom-sphere matches Number of matches to orient ligand Ignore Minimum matches to orient ligand
lbinsz lovlap sbinsz sovlap	**.1 .1 .2 .2**	Bins: sphere and atom internal distance histograms, used for matching
ligfil	file	Database filename
outfil	file	Output filename
ictbmp	0	Number of close contacts allowed between a ligand and the protein
natmin natmax nsav	5 100 200	Minimum number of atoms in a ligand to dock Maximum number of atoms in a ligand to dock Number of ligands to save
irestr moltot molsav	0 60 000 100	Restart run? Number of molecules to search Save results every molsav ligands
inchyd	Y	Include ligand hydrogens?
expmax (0 or 1)	**1**	Focus—amplifies the number of complementary dockings when value is 1
fctbmp (1)	**1**	Number of close contacts allowed to signal focusing algorithm
phifil (filename)	name	File containing electrostatic grid

Variables in **bold** face affect the amount of ligand orientational sampling.

other tests. For instance, your docking calculations will suggest roles for specific residues in recognizing your ligand, you may often test such roles with site-directed mutagenesis.

4.3 Example: docking phenolphthalein to thymidylate synthase

Some of the possibilities and limitations of a single molecule docking project may be illustrated with an example from our own work (7). We had discovered a new class of micromolar inhibitors, the phenolphthalein family of compounds, of the cancer target thymidylate synthase (TS) and wished to predict an accurate binding mode before the crystallographic solution. We began with:

- the structure of the enzyme in its open conformation
- the location of a lead (but dissimilar) inhibitor in the enzyme site
- the computer-generated structures of a number of analogues that inhibited the enzyme

We used the location of the atoms of the lead inhibitor as descriptors, and calculated a molecular electrostatic potential using the Delphi program. We eliminated all waters from the structure but retained two phosphate ions that were well resolved crystallographically and which occurred, in one form or another, in every crystal structure of TS of which we knew (then over 30). Ligand partial atomic charges were calculated with the Gasteiger–Marsilli (30) algorithm using Sybyl. We assumed that all analogues bound in the same manner, and insisted that a docking of any single analogue must be able to accommodate them all. This reduced the number of possible dockings to two families of highly related orientations, of which one was favoured energetically.

The crystallographic complex, when solved, showed the ligand in the same binding region as the docked complex, but translated by 1.4 Å and rotated by 43° (*Figure 2*). In the crystallographic structure, the inhibitor formed hydrogen bonds with two bound water molecules. One of these waters bound in a

Figure 2. Superposition of experimental (green) and low energy Dock consensus configuration (cyan) positions of phenolphthalein in the target site of TS (graphics made with MIDAS+) (48). Descriptors used in the docking calculations are represented as magenta spheres, enzyme side chains are coloured red. Ligand binding waters are represented as red tetrahedrons, their interactions with TSs residues and phenolphthalein are represented as dashed yellow lines.

novel site in TS, while the second occurred in several other crystal structures of the enzyme. The novel water interacted with an arginine that appeared in a significantly different conformation in the liganded form of the protein than the form that we had used for our modelling. In the unliganded form this arginine was poorly determined crystallographically. There were smaller accommodations involving several other residues, but for the most part the protein structure was unchanged. While we were able to get close with our docking calculations, the precise ligand geometry eluded us because of the roles of specific waters, and that of a residue whose conformation changed, for which we did not account. This is a good example of the sort of problems one can run into from waters, ions, and mobile residues.

5. Docking a protein ligand to a protein receptor

The fundamental problem is that there are many possible interfaces between two proteins, and considering all of them would take a great deal of time (31). It is important, even more so than in the small molecule case, to target regions on the protein surfaces that are likely to interact. One can target more than one region, but consider each separately. This is especially important for descriptor methods.

5.1 Defining interaction regions

Divide both the ligand and the receptor into different possible interaction regions. Interaction regions can be defined using experimental information (16), for instance from mutagenesis or mechanistic considerations. An alternative is to section the surface using structural criteria, such as the sphere method of Kuntz (5,20) (*Figure 1*). When using Dock, use less than 100 descriptors in any given receptor or ligand site.

It is sometimes sensible to eliminate flexible, poorly determined atoms in the binding region from the initial fitting calculations (20). An alternative is to increase the tolerances for close receptor–ligand contacts for such residues, to increase the number of allowed close contacts, or to model these residues as multiple conformations (32). Removing these atoms will reduce the potential complementarity of some complexes but will allow other complexes to be explored (see preparation of input for a rough guide). Use the coordinates of all atoms in the structure to evaluate the complexes once docked, even highly flexible or poorly determined ones.

5.2 Choosing the best docked complexes

The energy of the docked complexes may be evaluated by a number of different criteria, including:

- energy minimization
- solvation free energy (33,34)

- electrostatic complementarity
- packing density (35)
- mechanistic criteria

Of these methods, energy minimization is probably the most reliable (20). Use it before the other methods to remove high energy contacts that can be present in the initial docked structures owing to conformational uncertainties.

The results with test systems in the last few years suggest that it will be possible to dock proteins together using their unbound conformations, generating complexes that resemble the bound forms. In most cases, the invariant side chains and backbone atoms contain enough information to specify to a good approximation, if not uniquely, the binding mode of complex. The catch is 'not uniquely'. We found, in docking protein protease inhibitors into proteases, that Dock would generate not only the native configuration, but non-native complexes as well (20). Such non-native complexes differed from the crystallographic complex by between 8 Å and 22 Å rms—completely different sets of inhibitor residues were docked into the active site of the protease. While some of these non-native complexes were easy to exclude by relatively simple criteria, some others were quite complementary. An example of such a non-native complex, compared to the native, is shown in *Figure 3*. Both complexes show extensive and complementary surface area burial. We could not reliably distinguish the near native complexes from several of the complexes that were far from the native mode using such standard techniques as surface area burial, steric packing, hydrophobicity, mechanistic criteria, electrostatic complementarity, or molecular mechanics interaction energy. The same result, and even some of the same binding modes, was found by Cherfils *et al.* (10) using a grid-based docking approach very different from Dock. An obvious conclusion is that our energy evaluation schemes are still too simple, as is our ability to sample conformation space. A more intriguing hypothesis is that some of these non-native, but complementary dockings, might be sampled in solution, though they will be less favourable than the native (10,20). See Cherfils and Janin (36) for a thoughtful discussion of these issues. Recently, after this chapter was submitted, several docking groups were able to predict a protein–protein complex structure *de novo* (51). This encouraging result suggests that docking programs are increasingly reliable for predicting the structures of protein complexes.

6. Novel inhibitor discovery

There are several strategies towards structure-based inhibitor discovery (1,4,37). Docking programs use a database search approach (*Figure 4*). Rather than try to design a novel inhibitor from scratch, the idea is to scan a database of known, diverse molecules to see which, if any, will complement

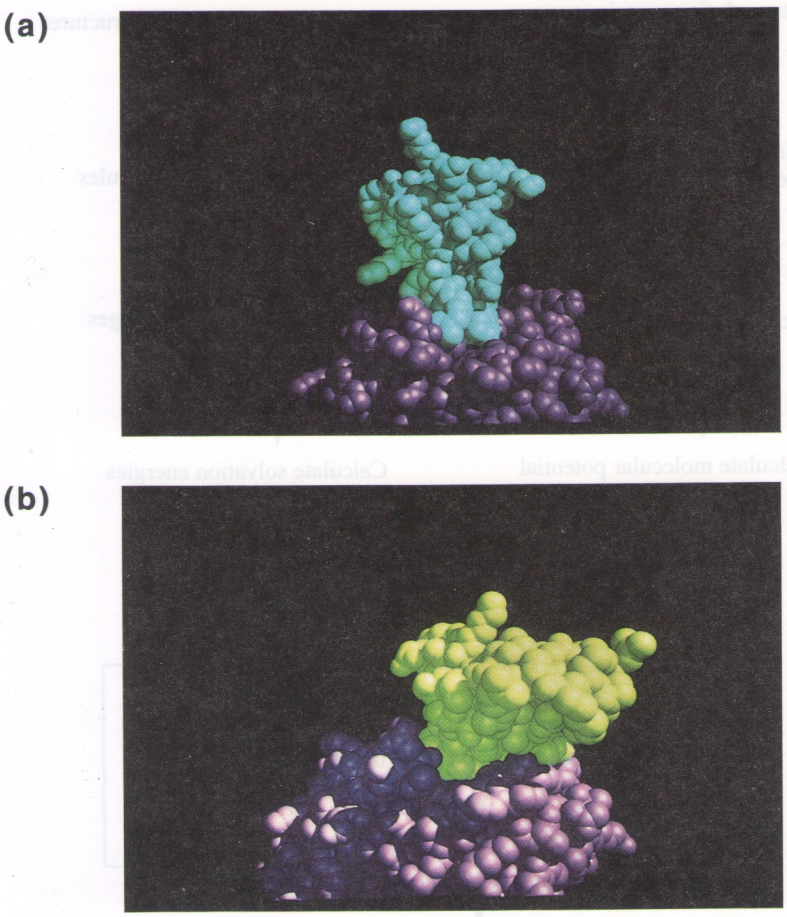

Figure 3. Energy minimized bovine pancreatic trypsin inhibitor (BPTI)/trypsin docking, free conformations, van der Waals representations of the interfaces (graphics made with MIDAS+) (48). (a) The lowest rms docked configuration, 0.52 Å from the crystallographic configuration, BPTI in cyan, trypsin in magenta. (b) A docked configuration of BPTI over 21 Å from the crystal structure of the complex, BPTI in green, trypsin in magenta. Both configurations have extensive, complementary interfaces.

the protein. One typically finds unusual, previously unconsidered molecules as candidate inhibitors. Some of these candidates will not bind to the protein when tested experimentally (false positives). Other molecules, which would inhibit if tested, or are known to inhibit, might not score well in the database search (false negatives). Since the goal is usually to discover *a* novel inhibitor, but not *all* novel inhibitors, such false positives and false negatives are tolerable. Using present technologies, our experience is that one can reasonably hope to find molecules that inhibit in the 1–100 micromolar range.

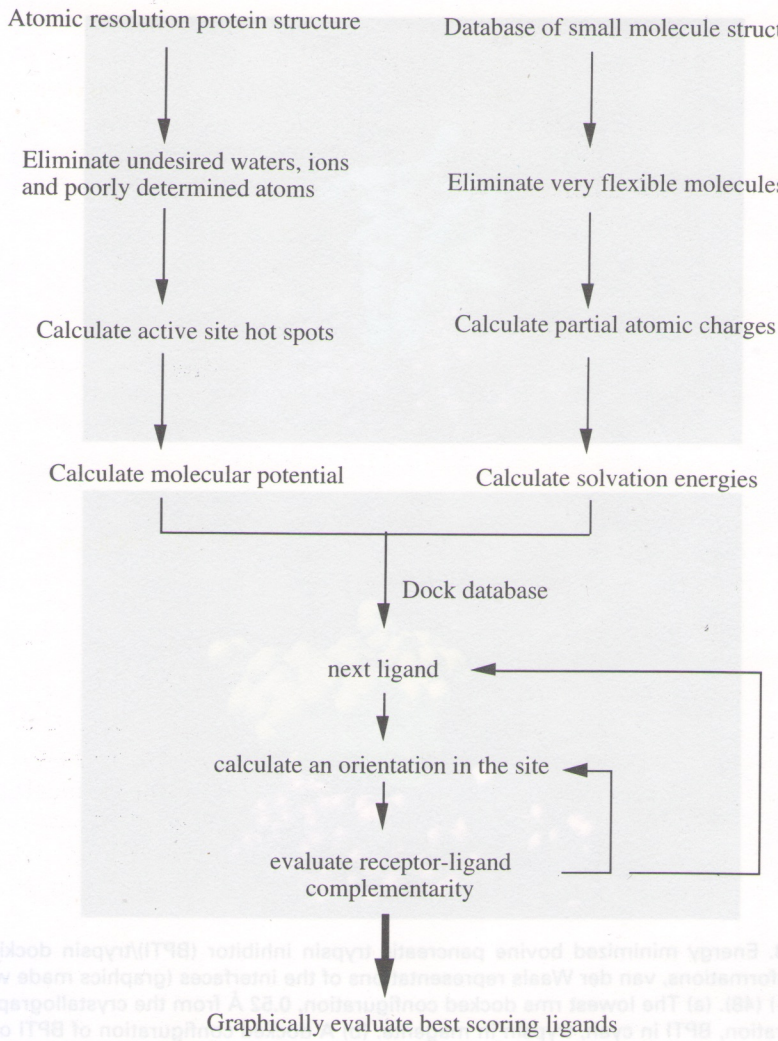

Figure 4. Flow chart for docking a molecular database to a protein.

6.1 Choosing a database

The first practical question is which molecular database to use (*Table 1*). One has two basic choices: using databases of experimentally-determined structures (e.g. the Cambridge Structural Database, CSD) (38), which are often not readily available for testing, or using databases of computer-generated structures (e.g. the Available Chemicals Directory, ACD) that are readily available for testing. I advise using the ACD or a similar database (scientists in the pharmaceutical industry will often have access to databases of pro-

prietary molecules), at least at the beginning of a project. Databases like the ACD allow for short cycle times between a computational hypothesis and an experimental test; you can simply buy any compound that scores well in the docking search. The quality of the structures in an experimentally-derived database, such as the CSD, are unquestionably more reliable than computer-generated. Such compounds are, however, typically difficult to acquire and often must be synthesized.

6.2 Preparing a database for docking

Database docking is single molecule docking magnified. The same input preparation is used (*Protocol 2*), save that one must assign parameters for not one but tens of thousands of molecules. Approximations are called for, as are programs to treat ligands sequentially and automatically. One must often assign the following characteristics to database molecules:

- hydrogen atom positions
- partial atomic charges
- atom types (for van der Waal parameters)
- solvation energies

6.2.1 Ligand atomic charges

Hydrogen atom positions and atomic charges are necessary to evaluate the electrostatic complementarity of the putative ligand–protein complexes. They can be calculated (using a program such as Sybyl or Macromodel) provided the database contains atomic number, hybridization, and connectivity information. Atom types are required for calculating dispersion interactions, for use with Dock they can be assigned with Mol2db (distributed with Dock).

6.2.2 Solvation corrections

Solvation energies are important to include when docking a diverse list of molecules to a protein. Charged molecules, for instance, will usually have a better calculated interaction energy with the protein than uncharged molecules, but they also will be better solvated by water. The proper way to balance the two effects is to calculate the differential desolvation energy of every orientation for every database molecule. This is not currently feasible. As a proxy for this calculation, we have instead calculated the full desolvation energy of every database molecule once, storing the energies in a look up table. This desolvation cost is subtracted from the interaction energy of every orientation of any given ligand on docking. This is a rough treatment a complex phenomenon, but it has significantly improved our ability to distinguish true inhibitors from false positives (23).

Balance each protein–ligand interaction term with a solvation term. Use Born equation methods, such as the Hydren program (39) or Delphi (40), to

calculate the electrostatic components of solvation for database molecules. This approach makes most sense when using a Poisson–Boltzmann-type method (41) (Delphi, for instance) to calculate the electrostatic potential of the protein. Set solvent and protein dielectrics to the same values in both the protein potential and ligand solvation calculations. Simple algorithms exist to calculate the hydrophobicity of most small molecules (e.g. CLOGP, available from the Medicinal Chemistry Project, Pomoma College, Claremont, CA 91711, USA) while programs such as Hint (42) (*Table 3*) evaluate hydrophobic contribution to surface area burial in a ligand–protein complex. Dispersion contributions to solvation are harder to explicitly calculate; the methods of Rashin (39) (Hydren program) and Still (43) that treat cavitation energies may be useful in this regard.

6.2.3 Ligand conformations

Small molecule databases for which structures have been calculated, such as the ACD, will often contain large, highly flexible molecules, such as long chain alkanes. Only one conformation of such molecules is typically represented, and this conformation is often unreasonable. Unless you wish to calculate multiple conformations for the molecules you should delete such molecules from the database. There are several programs to calculate multiple ligand conformations based on an initial two- or three-dimensional structure (e.g. Sybyl, MacroModel), we have found the Cobra program (28) useful in this regard (23). The Concord (44) program is widely used for calculating a single three-dimensional structure from a two-dimensional representation of a molecule (*Table 4*).

6.3 Controls

Database searches take a fairly long time. Expect anywhere from 2–14 days on a workstation such as a Silicon Graphics Indigo 2 for about 120 000 ACD molecules, docked in one conformation. There are no formal upper limits on such a calculation. A week of your time might be spent evaluating the results of a search, considerably more testing them. Consider the following controls:

(a) Calculate the interaction energy for a known complex of a ligand with your protein—the energy should be negative (*Protocol 3*).

(b) Reproduce the experimentally-determined structure of a known ligand (*Protocol 4*).

(c) Establish that known ligands of the protein that exist in the database score well when docked into the protein (*Protocol 5*).

It is often possible to find known ligands of your protein in the molecular database. Extract such molecules from the database along with congeners that do not bind to your protein. Dock this mini-database into the protein. Begin with input variables that worked well for your single ligand docking

control. You may have to vary several of these to achieve a balance between the length of the calculation and the quality of your results. Molecules known to bind to the protein should score well (negative interaction energies), while the non-binding congeners should score relatively poorly (more positive interaction energies).

Protocol 5. Control for Dock database search

1. Dock the control database. The known ligands should score well (negative scores) and the non-ligand molecules should score less well. If this is true the control has worked.
2. High positive (poor) scores for known ligands often arise because Dock was unable to find the proper configuration for the ligands. Visually inspect the docked complexes to insure that the ligands are bound in the expected geometries. If they are not, the following problems may have arisen:
   (a) Configurational sampling may have been inadequate, check the NMATCH number reported in OUTDOCK. If this is low compared to the single ligand control, increase sampling by varying the Dock search parameters such as bin size, focusing, and so on.
   (b) Ligand parameterization may be incorrect. Inspect partial atomic charges, assigned atom types, atom labels (for matching), and so forth. If you have labelled your protein spheres you must label your ligand atoms in a complementary fashion. If you are correcting interaction energies with solvation energies, insure that you are using the same terms for each. For electrostatic solvation corrections insure that the dielectric of the desolvated phase is the same as the dielectric of the solvent excluded volume of the protein.
   (c) Ligand conformations may be wrong. Focus on ligands with as few 'rotatable' bonds as possible. If you have the resources, consider docking multiple configurations of the ligands.
3. Low negative (good) scores for your negative control molecules often arise because you have not corrected for solvation, because your close contact limits are too loose, or you are expecting the docking program to make too subtle a distinction.

6.3.1 Example database search control

Table 6 shows the results of a database search control calculation, docking against the enzyme thymidylate synthase (TS). Pyrimidine monophosphates should score well, while diphosphates, triphosphates, and higher charged species should score poorly. We chose a subset of 696 ACD molecules that

Table 6. Dock database search control example: sensitivity of results to variable choices

Variable changed	Values	No. of nucleotide monophosphates in top 25 molecules	No. of pyrimidine monophosphates in top 25 molecules	Average orientations per ligand	No. of ligands found	Run time (sec)[a]
None	Below[b]	16	12	390	321	990
dislim	0.75	14	4	11	33	443
dislim	2.0	6 (24/50)	5 (12/50)	1688	380	1984
nodlim	7	24	1	46	48	536
Bin sizes	0.3, 0, 0.8, 0	12	6	400	322	970
Focusing	0	6	1	221	245	800

[a] All calculations performed on a Silicon Graphics PI 25.
[b] Beginning Dock input variables were as in *Table 5*, except that bins were set to 0.15, 0., 1.0, and 0. for lbinsz, lovlap, sbinsz, and sovlap.

contained either a phosphate, a sulfate, a phosphonate, or a sulfonate and fitted them into the dUMP binding site of TS (23). Orientations were scored for electrostatic interaction and steric fit using Delphi and Distmap potential lattices. Ligand electrostatic solvation was corrected for using energies calculated with Hydren (*Table 4*). Generally, the pyrimidine monophosphates docked with large, negative binding energies that were typically better than most other molecules in the database and considerably better than pyrimidine di- and triphosphates, or any molecule bearing more than two negative charges.

We varied several of the Dock input parameters searching for conditions that would:

- maximize the number of high scoring pyrimidine monophosphates
- maximize the number of high scoring nucleotide monophosphates in general

Even though the latter are not recognized by TS, we felt that good scores for these molecules were appropriate given the simplicity of the scoring scheme. Note that different input parameters (*Table 5*) gave different results. Increasing the stringency of the matching criteria, by increasing *nodlim* or decreasing *dislim*, eliminated most non-nucleotide monophosphates. This in turn increased the proportion of nucleotide monophosphates in the final docked list, but at the cost of removing several pyrimidine monophosphates that are known ligands of TS. Decreasing the stringency by increasing *dislim* results in more ligands being found, including more high scoring non-nucleotide monophosphates. More nucleotide and pyrimidine monophosphates are, of course, also found. Using the focusing option (45) considerably improves the results. I favour using this option, but opinion in the Kuntz group is divided as to its utility.

6.4 Choosing molecules to assay for binding

At the end of a Dock database search you will have:

(a) An interaction score for every database molecule docked. One may choose several scoring criteria; Dock reports separate scores for each method.

(b) For each scoring method, a list of the best molecules in the database, ranked by decreasing interaction energy. The atomic coordinates of these molecules are produced in their best scoring configuration in the binding site.

Visually inspect the docked molecules using a molecular graphics program (*Table 4*). Some of the molecules that the docking program suggests may seem inappropriate on graphical inspection. Reasons for discarding a database molecule include:

(a) Complexes dominated by one favourable interaction, with much of the ligand extending out of the binding region into solvent.

(b) Molecules that will be insoluble under the conditions of your assay.

(c) Molecules that will react in a non-specific manner with your protein or components of your assay.

6.5 Strategies for testing the molecules

Many of your candidate inhibitors may be difficult to dissolve in water. Consider DMSO as a general solvent to deliver candidate inhibitors to the assay. You must control for the effect of the DMSO on the reaction, it is usually possible to limit the amount of DMSO delivered to one per cent or less of the total reaction volume.

The absolute interaction energy that a program such as Dock reports for any given database molecule is untrustworthy (1). At best we might hope for a monotonic relationship between relative rank and relative binding free energy, but even here our controls advise considerable caution (23). We are still at a stage where intuition, as well as computed scores, must guide our choices. Individuals will weight the computed scores and their own insight differently; there is no 'right' way to do this. I prefer a two track strategy:

- take several weeks to test the molecules in order of their docked scores
- take several weeks to test the high scoring molecules that you like the best

6.5.1 Specificity controls

When, and if, you find a molecule that binds to the target protein, control for non-specific binding. Generally, a molecule that will bind to most proteins is uninteresting.

(a) Establish that inhibition is reversible, unless one has specifically called for chemical modification of the protein.

(b) Determine whether the molecule binds competitively, uncompetitively, or non-competitively compared to a ligand known to bind in the target site (e.g. the substrate).

(c) Measure the inhibitory potential of the molecule against other proteins.

6.5.2 Finding analogues

Once you have established specific binding, look for analogues of the lead that will also bind to the protein. Such analogues often exist in a database and can be found using similarity searches using programs such as Isis (*Table 4*) or Aladdin (46). To improve the chances of getting useful predictions from the docking program, individually redock the analogues, using greater sampling and multiple conformations.

6.5.3 Establishing binding modes

It is important to know the bound geometry of a lead, especially if one wants to optimize inhibition through chemical elaboration. I have outlined a computational strategy towards geometry elucidation in Section 4.3, ultimately this will depend on the solution of atomic resolution structures. Molecules may bind in multiple configurations and conformations with similar energies, especially in the micromolar inhibition range. The first suggestion of multiple binding modes is often a series of analogues that inhibit but are difficult to fit into a single binding geometry. To test whether a ligand or series of ligands are binding in dissimilar geometries:

(a) Measure the binding of an analogue(s) that dramatically does not fit your favoured geometry (e.g. the analogue contains a bulky group that would intersect a rigid region of the protein). Binding of such an analogue suggests either that your model is wrong or there are multiple binding modes.

(b) Make a residue substitution in the protein that intersects the putative binding region. If all analogues still bind there are either multiple binding modes or the model is wrong. If some analogues still bind but others do not then there are probably multiple binding modes.

Acknowledgements

I thank the UCSF docking group, past and present, especially Tack Kuntz, Elaine Meng, Dale Bodian, Renee DesJarlais, George Seibel, Andrew Leach, Richard Lewis, Dan Gschwend, Cindy Corwin, Diana Roe, Chris Ring, and Kathy Perry. Thanks also to John Irwin, Enoch Baldwin, and Paul Charifson, and to the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation Fellowship, DRG-1228, for financial support. The Available Chemical Directory and the Isis program were kindly provided by MDL Information Systems, Inc, San Leandro, CA 94577, USA.

References

1. Kuntz, I. D. (1992). *Science*, **257,** 1078.
2. Kuntz, I. D., Meng, E. C., and Shoichet, B. K. (1994). *Acc. Chem. Res.*, **27**, 117.
3. Varney, M. D., Marzoni, G. P., Palmer, C. L., Deal, J. G., Webber, S., Welsh, K. M., *et al.* (1992). *J. Med. Chem.*, **35,** 663.
4. Itzstein, M. V., Wu, W.-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., *et al.* (1993). *Nature*, **363,** 418.
5. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. (1982). *J. Mol. Biol.*, **161,** 269.
6. DesJarlais, R., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D., and Venkataraghavan, R. (1988). *J. Med. Chem.*, **31,** 722.
7. Shoichet, B. K., Perry, K. M., Santi, D. V., Stroud, R. M., and Kuntz, I. D. (1993). *Science*, **259,** 1445.
8. van Gunsteren, W. F. and Berendsen, H. J. C. (1990). *Angew. Chem. Int. Ed. Engl.*, **29,** 992.
9. Smellie, A. S., Crippen, G. M., and Richards, W. G. (1991). *J. Chem. Inf. Comput. Sci.*, **31,** 386.
10. Cherfils, J., Duquerroy, S., and Janin, J. (1991). *Proteins*, **11,** 271.
11. Jiang, F. and Kim, S. H. (1991). *J. Mol. Biol.*, **201,** 79.
12. Pearlman, D. A. and Murko, M. A. (1993). *J. Comput. Chem.*, **10,** 1184.
13. Bohm, H.-J. (1993). *J. Comput. Aid. Mol. Des.*, **6,** 61.
14. Moon, J. B. and Howe, W. J. (1991). *Proteins: Struct. Funct. Genet.*, **11,** 314.
15. Goodsell, D. S. and Olson, A. J. (1990). *Proteins*, **8,** 195.
16. Stoddard, B. L. and Koshland, D. E. (1992). *Nature*, **358,** 774.
17. Ring, C. S., Sun, E., McKerrow, J. H., Lee, G. K., Rosenthal, P. J., Kuntz, I. D., *et al.* (1993). *Proc. Natl. Acad. Sci. USA*, **90,** 3583.
18. Bax, B., Blaber, M., Ferguson, G., Sternberg, M. J. E., and Walls, P. H. (1993). *Protein Sci.*, **2,** 1229.
19. Rutenber, E., Fauman, E. B., Keenan, R. J., Fong, S., Furth, P. S., Ortiz-de-Montellano, P. R., *et al.* (1993). *J. Biol. Chem.*, **268,** 15343.
20. Shoichet, B. and Kuntz, I. D. (1991). *J. Mol. Biol.*, **221,** 327.
21. Bodian, D. L., Yamasaki, R. B., Buswell, R. L., Stearns, J. F., White, J. M., and Kuntz, I. D. (1993). *Biochemistry*, **32,** 2967.
22. Lawrence, M. C. and Davis, P. C. (1992). *Proteins*, **12,** 31.
23. Shoichet, B. K. and Kuntz, I. D. (1993). *Protein Eng.*, **6,** 723.
24. Wodak, S. J. and Janin, J. (1978). *J. Mol. Biol.*, **124,** 323.
25. DesJarlais, R. L. and Dixon, J. S. (1994). *J. Comput. Aid. Mol. Des.*, **8,** 231.
26. Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., *et al.* (1990). *J. Comput. Chem.*, **11,** 440.
27. Hansch, C. and Leo, A. J. (1979). *Substituent constants for correlation analysis in chemistry and biology*. John Wiley and Sons, New York.
28. Leach, A. R. and Prout, K. (1990). *J. Comput. Chem.*, **11,** 1193.
29. Meng, E. C., Gschwend, D. C., Blaney, J. M., and Kuntz, I. D. (1993). *Proteins*, **17,** 266.
30. Gasteiger, J. and Marsili, M. (1980). *Tetrahedron*, **36,** 3219.
31. Tilton, R. F. Jr., Singh, U. C., Weiner, S. J., Connolly, M. L., Kuntz, I. D. Jr., Kollman, P. A., *et al.* (1986). *J. Mol. Biol.*, **192,** 443.

32. Wilson, C., Mace, J. E., and Agard, D. A. (1991). *J. Mol. Biol.*, **220,** 495.
33. Eisenberg, D. and McLachlan, A. D. (1986). *Nature*, **319,** 199.
34. Horton, N. and Lewis, M. (1992). *Protein Sci.*, **1,** 169.
35. Gregoret, L. M. and Cohen, F. E. (1990). *J. Mol. Biol.*, **211,** 959.
36. Cherfils, J. and Janin, J. (1993). *Curr. Opin. Struct. Biol.*, **3,** 265.
37. Appelt, K., Bacquet, R. J., Bartlett, C. A., Booth, C. L., Freer, S. T., Fuhry, M. A. *et al.* (1991). *J. Med. Chem.*, **34,** 1925.
38. Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., *et al.* (1979). *Acta Crystallogr. Sect. B*, **B35,** 2331.
39. Rashin, A. A. (1990). *J. Phys. Chem.*, **94,** 1725.
40. Gilson, M. K. and Honig, B. H. (1987). *Nature*, **330,** 84.
41. Warwicker, J. and Watson, H. C. (1982). *J. Mol. Biol.*, **157,** 671.
42. Wireko, F. C., Kellogg, G. E., and Abraham, D. J. (1991). *J. Med. Chem.*, **34,** 758.
43. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990). *J. Am. Chem. Soc.*, **112,** 6127.
44. Rusinko, A., Sheridan, R. P., Nilakatan, R., Haraki, K. S., Bauman, N., and Venkataghavan, R. (1989). *J. Chem. Inf. Comput. Sci.*, **29,** 251.
45. Shoichet, B., Bodian, D. L., and Kuntz, I. D. (1992). *J. Comput. Chem.*, **13,** 380.
46. Van Drie, J. H., Weininger, D., and Martin, Y. C. (1989). *J. Comput. Aid. Mol. Des.*, **3,** 225.
47. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr., Brice, M. D., Rodgers, J. R., *et al.* (1977). *J. Mol. Biol.*, **112,** 535.
48. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988). *J. Mol. Graph.*, **6,** 13.
49. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. and Vakser, J. A. (1992). *Proc. Natl. Acad. Sci.* (USA), **89**, 2195.
50. Totrov, M., and Abagyan, R. (1994) *Nature Structural Biology*, **1**, 259.
51. Strynadka, N. C., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B. K., Kuntz, I. D., Abagyan, R. *et al.* (1996). *Nature Structural Biology*, **3**, 233–39.