3DPL Technology and Validation

Dr. Tad Hurst
CTO and Vice President,
ChemNavigator
San Diego, California

Background

Drug discovery typically involves two steps: 1) screen development; and 2) lead identification and optimization. In the first step, a method of screening potential drug compounds must be developed and validated. This generally involves detailed knowledge of the biochemistry of the disease state. The second step, lead identification and optimization, is very expensive and time consuming. It is not uncommon for researchers to screen millions of small molecules for those likely to be good drug leads.

Once one or more drug leads to a target have been identified, it is still necessary to optimize those leads to produce a drug candidate. In this optimization process, synthetic chemists synthesize variants of the lead compound to increase its efficacy, improve its toxicity profile, modify its susceptibility to degradative pathways, or modify its pharmacokinetics. The chemist makes a set of small changes to the structure, and determines if those changes had a beneficial or detrimental effect on the efficacy. This process, called analog synthesis, is effective, although time-consuming and expensive. Analog synthesis is the basis for medicinal chemistry, and remains an important part of drug discovery today.

Much effort has been directed toward the use of computers and computational methods to improve the efficiency of drug discovery. This offers the possibility of reducing drug discovery time and expenses by reducing the number of compounds screened to discover a lead.

Early attempts at using computers to make the lead optimization process more efficient involved determining chemical similarity between a potential lead and a known lead using graph-theoretical treatments. These methods mimic the actions of the chemist – finding compounds that are only slightly different than the known lead. Methods of this type include searching a database of compounds for those that contain the same core structure as the lead compound – called substructure searching or two-dimensional (2-D) searching – and searching for compounds that are generally similar based on the presence of a large number of common fragments between the potential lead expansion compound and the lead compound – called 2-D similarity searching. These techniques are effective, but are limited to finding compounds that are obviously similar to the lead, thus affording the medicinal chemist few new insights for directing the synthesis project.

Most small molecule drugs affect the biological system by binding to a large molecule – usually a protein or an enzyme. Thus, the more advance computational techniques involve prediction or determination of the potential for the compound to bind to a biological receptor.

If the 3 dimensional (3-D) structure of the receptor is not known, the 3-D arrangement of the chemical groups responsible for binding can often be inferred from a set of known binding agents. These chemical groups – called pharmacophore groups - are responsible for most of the stabilization energy of the complex of the small molecule and the receptor, and the 3-D arrangement of the groups that is responsible for biological activity is called a "pharmacophore model", or "3-D query".

Three-dimensional (3-D) searching techniques search through large database of potential lead compounds to find those that have essentially the same geometrical arrangement of pharmacophore groups as the lead compound. These 3-D hits are candidates for screening. Hits from 3-D searching differ from the substructure or 2-D similarity hits in that the backbone of the structure may be quite different from that of the original lead compound, and often represents an important, new area of chemistry to be explored.

The simplest of the 3-D searching techniques compares the position and arrangement of the pharmacophore groups of 3-D structures as stored in the database. This is referred to as static 3-D searching, and does not consider the flexibility of the structures in the database. Most drug-like molecules have a large number of accessible conformations formed from the modification of the dihedral angles of the bonds that are freely rotatable. Small molecules in pharmaceutical databases typically contain an average of six to eight rotatable bonds per molecule. This can easily afford a set of accessible conformations that number in the millions. Searching just one static conformation from among the millions that are possible will cause many compounds that could be good leads to be missed.

In order to consider energetically accessible conformations, many 3-D searching systems require the storage of a small subset of the accessible conformations of each small molecule in the database, or they produce the small set of conformations on the fly. This technique is called multi-conformational 3-D searching, and is sometimes erroneously called conformationally flexible 3-D searching. Trying to sample the conformational space of a small molecule with a handful of conformations is impractical, as it often requires millions of conformations to adequately represent the entire accessible space. Multi-conformational 3-D techniques therefore only partially address the flexibility problem.

A further extension of 3-D search technology involves investigation of the accessible conformational space of the potential hits as part of the searching process. These techniques – the true "conformationally flexible 3-D searching techniques" - adjust the conformation of the potential hit according to the requirements of the 3-D query. The most effective of these methods is called Directed Tweak[1]. This method is very effective for finding molecules of interest when the geometry of the binding site of the large molecule is not known. Directed Tweak adjusts the conformation of the small molecule by changing the angle values of the rotatable bonds. This method therefore ignores changes in conformation because of bond stretching and bond bending. Bond stretching vibrations for molecules near room temperature typically change the length of a bond by about 0.05 Angstroms (Å). Bond bending between three connected atoms typically moves one of the atoms by about 0.1 Å. Rotation about rotatable bonds often moves atoms by several Ås or tens of Ås. Thus, adjusting only the rotatable bond values includes essentially all of the accessible conformational flexibility of a small molecule.

When the binding site of the target protein is known, the potential for a small molecule to dock into the binding site can be determined computationally. Docking approaches can be classified based on how they characterize the ligand-binding site of the protein. Grid-search techniques fill the space around the binding site with a 3-D grid, precompute the potentials (van de Waals, electrostatic, etc.) at each grid point, and then sample different ligand conformations and orientations on the grid to compute the resulting binding energy.

Some docking methods use molecular mechanics minimization techniques. These methods calculate long and a short range contributions to the interaction energy including such terms as electrostatic and van der Walls energies. The position of the small molecule

[1] Hurst, T. "Flexible 3D Searching: The Directed Tweak Technique", *J. Chem. Inf. Comput. Sci*. **34** (1994), 190.

is then adjusted iteratively so as to give lower and lower energies. This continues until a low energy arrangement is found.

Another well-known docking tool is DOCK[2]. This program generates an inverse image of the protein's binding site that consists of up to 100 spheres. During the search, subsets of ligand atoms are matched to spheres, based on the distances between ligand atoms.

Another docking program, FlexX[3] uses a template of 400 to 800 points when docking small molecules (up to 40 atoms, not including hydrogen atoms) to define positions for favorable interactions of groups such as hydrogen-bond donors and acceptors, metal ions, aromatic rings, and methyl groups. The ligand is fragmented and incrementally reconstructed in the binding site to provide good overlap of the groups and the receptor interaction points.

Hammerhead[4] uses up to 300 hydrogen-bonding and steric interaction points to define the template, and the ligand is incrementally constructed, as in FlexX. A fragment is docked based on matching ligand atoms and template points with compatible internal distances. If a new fragment is positioned closely enough to the partially constructed ligand, the two parts are merged, and the most promising placements kept.

Other successful docking approaches, such as GOLD[5], AutoDock[6], and the method of Oshiro et al.[7], use genetic algorithms to sample over possible matches of conformationally flexible ligands to the template. GOLD uses a template based on hydrogen-bond donors and acceptors of the protein and applies a genetic algorithm to sample over all possible combinations of intermolecular hydrogen-bonds and ligand conformations. These methods are computationally intense, and do not lend themselves to searching of databases of millions of compounds in an efficient manner.

Another current docking method, SPECITOPE[8], combines grid methods with distance geometry techniques in order to model protein side chain flexibility. The speed gained by distance geometry methods allows the modeling of protein side chain flexibility during docking.

The UNITY 3-D Searching System[9] has been extended to provide what is essentially a docking tool. In this approach, six parameters corresponding to the six rotational/translational degrees of freedom are added to the rotatable bond list, and these parameters are adjusted to place pharmacophoric groups at the positions giving favorable interactions with the receptor. This method produces acceptable accuracy, but is time consuming because the derivatives needed for the minimization are calculated numerically.

---

[2] Schoichet, B.K., Bodian, D.L., and Kuntz, I.D., *J. Comput. Chem.*, **13** (1992) 380.

[3] FlexX is licensed by Tripos Inc. - 1699 South Hanley Road, St. Louis, MO 63144-2913, phone: 314-647-1099.

[4] Welch W, Ruppert J, Jain AN. 1996. "Hammerhead: Fast, fully automated docking of flexible ligands into protein binding sites", *Chem Biol* **3**(1996), 449-462.

[5] G. Jones, P. Willett, R. C. Glen, A. R. Leach & R. Taylor, *J. Mol. Biol,* **267** (1997) 727.

[6] Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. "Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function", *J. Computational Chemistry*, **19** (1998), 1639-1662.

[7] Oshiro C, Kuntz I, Dixon J. "Flexible ligand docking using a genetic algorithm", *J Compu.t Aided Mol. Des.*, **9** (1995)113--130.

[8] Volker Schnecke, Craig A. Swanson, Elizabeth D. Getzoff, John A. Tainer, and Leslie A. Kuhn, "Screening a Peptidyl Database for Potential Ligands to Proteins with Side-chain Flexibility Proteins: Structure, Function, and Genetics", **33**, (1998), 74-87

[9] UNITY is licensed by Tripos Inc. - 1699 South Hanley Road, St. Louis, MO 63144-2913, phone: 314-647-1099.

In most docking approaches, the ligand binding site on the receptor must be known. When the active site of a protein or other target molecule is not known, an estimate of the binding site must be made. Even when the active site is known, it may still be useful to determine possible allosteric binding sites.

The DockIt[10] program uses a negative image of the receptor site based on filling it with spheres, as in Dock. DockIt rates the generated spheres based on their burial score – the most buried spheres are considered more likely to be part of a binding site.

GOLD calculates, for each point potentially in a binding site, the number of times lines through the point intersect the solvent accessible surface of the protein. Points that are deep in pockets, and are thus potential member points of a binding site, will have lines with larger number of receptor intersections than points on the exposed surface of the protein.

The SiteID program[11] displays various properties of a target molecule surface that may relate to the likelihood of the area being an active site. The user can then visualize the structure looking for potential binding sites. Connolly[12] has reviewed various procedures and methods for visualizing the surface topology of target molecules. These procedures may assist in the identification of potential binding sites.

The Insight[13] program flood-fills a cavity with solvent spheres to investigate the depth of a cavity. Points that are several layers deep represent potential binding sites.

Most of the docking methods described were created with the goal of either reproducing the binding configuration of known ligands, or accurately estimating the binding energy of those interactions. They are generally not well suited to searching databases of millions of compounds because of their computationally intense nature. In addition, most require specific indication of the binding site. This is often known for receptors that have 3-D structures that have been determined by X-ray crystallography, but is not known for 3-D structures of many proteins produced in other methods as part of ongoing proteomics work.

Many research laboratories have assembled large farms of computers, sometimes numbering in the thousands, in order to dock large numbers of potential lead compounds using standard docking approaches such as those discussed above. Whereas this may be effective at increasing the throughput of docking systems, the advent of an ultra-fast database docking system, combined with the server-farm approach, will allow the entire proteome to be investigated computationally.

3DPL Technology

The 3DPL system is designed to provide database docking for millions of compounds. In addition, it is designed such that the binding site need not be known in advance. This allows 3DPL to be applied to proteins whose 3-D structures do not include a co-crystallized ligand or any other indication of the binding site. It also allows the system to identify compounds that might bind in other allosteric binding sites.

There are two proprietary aspects of the 3DPL system technology: the method of derivative calculation, and the method for identifying potential binding sites on the receptor. The former is an integral part of a search method that is a hybrid of 3-D searching and

---

[10] Metaphorics (27401 Los Altos #360, Mission Viejo, CA 92691, Phone: (949) 367-9091.

[11] SiteID is licensed by Tripos Inc. - 1699 South Hanley Road, St. Louis, MO 63144-2913, phone: 314-647-1099.

[12] Connolly, M, *Molecular Surfaces*: *A Review*, http://www.biochem.usyd.edu.au/~bchurch/netsci.html

[13] Insight is distributed by Accelrys, 9685 Scranton Road, San Diego, CA 92121, 858-799-5509.

traditional docking. This method is very fast, and can investigate potential lead compounds from a database at a rate of up to 10 per second.

The site-finding methodology is called the Concave-Volume method. This method identifies potential site points at locations all around the receptor, thus including primary and allosteric binding sites.

Derivative Calculations

The 3DPL methodology employs a torsional space minimizer to explore ligand conformations that can dock into the receptor. In torsional space, the position and conformation of the putative ligand structure are specified as a function of parameters that reflect the 6 translational/rotational degrees and the torsion angles of the rotatable bonds. This treatment ignores the conformational changes that result from bond stretching and bond bending. As discussed earlier, bond stretching and bond bending result in very little conformational change compared to the changes that result from rotation about rotatable bonds.

In general, any minimization method may be used. Suitable minimization methods include BFGS, Steepest Descent, and the Conjugate Gradient[14] methods. The Steepest Descent method uses the first derivative of the pseudo-energy as a function of the geometric parameters. Some methods, like the Conjugate Gradient and BFGS methods, consider both the first and second derivatives of the pseudo-energy. These have been shown in some cases to have better convergence behavior than the methods that only calculate first derivatives of the pseudo-energy. Typically in 3DPL, the steepest descent method is used.

Any pseudo-energy function may be used. Typical functions include terms that reflect the electrostatic forces, the van der Waals (VDW) energies, hydrogen-bonding interactions, and hydrophobic terms. The 3DPL system normally uses hydrogen-bonding and steric (VDW) terms. Both the VDW and H-bonding terms have both an attractive and a repulsive component. We typically soften the exponent of the attractive term to a value of 2 to get better convergence behavior.

The determination of the energy terms and their derivatives are often computationally intense. The interaction of each atom of the receptor with each atom of the putative ligand must be considered. If the ligand contains a few hundred atoms and the receptor has several thousand atoms, the task becomes computationally expensive. 3DPL addresses this by dividing the derivative calculations into two parts. The desired quantity is the partial derivative of the pseudo-energy with respect to one of the geometric parameters Q_j. These parameters Q_j are either one of the translation or rotation values, or one of the rotatable bond torsion angle values. This partial derivative is a scalar quantity, and is represented as the dot product or two vectors (Eq 1). Here $\delta E/\delta P_i$ is the derivative vector or the energy with respect the position of the i'th atom, and $\delta P_i/Q_j$ is the derivative vector of the position of the I'th atom with respect to the j'th geometric parameter.

Eq 1

$$\frac{\partial E}{\partial Q_j} = \frac{\partial E}{\partial P_i} \cdot \frac{\partial P_i}{\partial Q_j}$$

For translation parameters, consider translation along the x-axis (Eq 2)

[14] Shanno, D.F. "Conjugate gradient methods with inexact searches". Mathematics of Operations Research 3-3, 244-256.

Eq 2

$$\frac{\partial E}{\partial T_x} = \sum_i \left(\frac{\partial E}{\partial P_i} \cdot \frac{\partial P_i}{\partial T_x} \right)$$

The latter term $\frac{\partial P_i}{\partial T_x}$ is the unit vector (1,0,0). Likewise, $\frac{\partial P_i}{\partial T_y}$ is the unit vector (0,1,0) and $\frac{\partial P_i}{\partial T_z}$ is the unit vector (0,0,1). Thus the overall derivative with respect to translation is given by Eq 3

Eq 3

$$\frac{\partial E}{\partial T} = \left(\frac{\partial E}{\partial T_x}, \frac{\partial E}{\partial T_y}, \frac{\partial E}{\partial T_z} \right) = \sum_i \frac{\partial E}{\partial P_i}$$

For rotation parameters, consider first the rotation about the x-axis (Eq 4), where $\frac{\partial P_i}{\partial R_x}$ represents the partial derivative of the position of the i'th atom with respect to rotation about the axis parallel to the x-axis.

Eq 4

$$\frac{\partial E}{\partial R_x} = \sum_i \frac{\partial E}{\partial P_i} \cdot \frac{\partial P_i}{\partial R_x}$$

The derivative of the position of atom i with respect to the rotation of a candidate molecule about this axis may be determined by the cross product in Eq 5, where D_i is the difference vector of between position of atom i and the center of rotation.

Eq 5

$$\frac{\partial P_i}{\partial R_x} = D_i \times (1,0,0)$$

Where R = (R_x,R_y,R_z), this gives Eq 6.

Eq 6

$$\frac{\partial E}{\partial R} = \sum_i \frac{\partial E}{\partial P_i} \times D_i$$

Now for each of the *b* rotatable bond parameters, the derivative can be represented by Eq 7, where, $\frac{\partial P_i}{\partial \theta_b}$ is the partial derivative of P_i, with respect to the rotation angle θ_b of the b'th rotatable bond.

Eq 7

$$\frac{\partial E}{\partial \theta_b} = \sum_i \frac{\partial E}{\partial P_i} \cdot \frac{\partial P_i}{\partial \theta_b}$$

All three types of parameters depend on the terms $\frac{\partial E}{\partial P_i}$, which is in turn a summation of the contributions from each of the receptor atoms. The 3DPL technology pre-

computes these terms for each possible type of atom and each energy term in the pseudo-energy equation, at every point in a regularly spaced grid. Thus the grid points include not only the energy of an atom of a particular type at each location, but also the derivative vector as well. This is computationally demanding, but must be done only once prior to the database search. One such grid is calculated and stored for each atom type to be encountered in the database search (e.g. Carbon atoms, Hydrogen atom that are not donors, Hydrogen atoms that are donors, *etc.*), millions of compounds can be searched simply by looking up the values of these derivatives. Typically the grids are calculated at 0.2 angstroms.

The use of these pre-computed derivatives to get the derivatives of translation and rotation are simple, as described above. The rotatable bond derivatives require calculation of the additional term $\frac{\partial P_i}{\partial \theta_b}$, which is the derivative of the atom position with respect to rotation about the b'th bond. This quantity is easily calculable from the cross product in Eq 8.

Eq 8

$$\frac{\partial P_i}{\partial \theta_x} = u_b \times d_{i,b}$$

where u_b is the unit vector along the b'th rotatable bond, and $d_{i,b}$ is the vector from one end of the rotatable bond to the atom position (Figure 1). Thus the overall derivative of Energy with respect to one rotatable bond is calculated by one cross product and one dot product. This is computationally simple, and affords ultra-fast searching times.

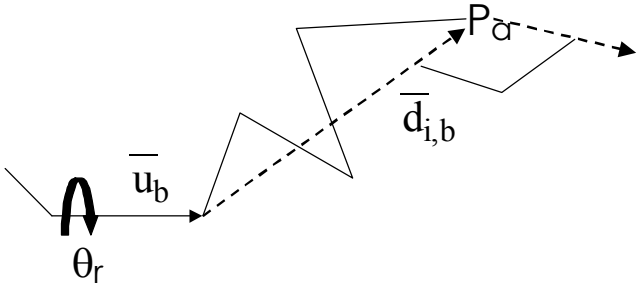

Figure 1. Calculating the derivative of an atom position in rotatable bond space

Binding Site Determination

The 3DPL methodology determines the potential binding site positions using a technique call "Concave Volumes". The general approach is to find the volumes large enough to bind a potential ligand in the concave portions of the receptor.

To find the concave portions of the receptor, the 3DPL system first defines the parts that are not concave. This is accomplished by determination of the convex hull of the protein. A convex hull is a mathematical construct that represents the smallest convex polyhedron that contains all of the defining points. For a receptor, the defining points are the coordinates of the atoms, and the convex hull represents the convex shape of the protein (Figure 2).

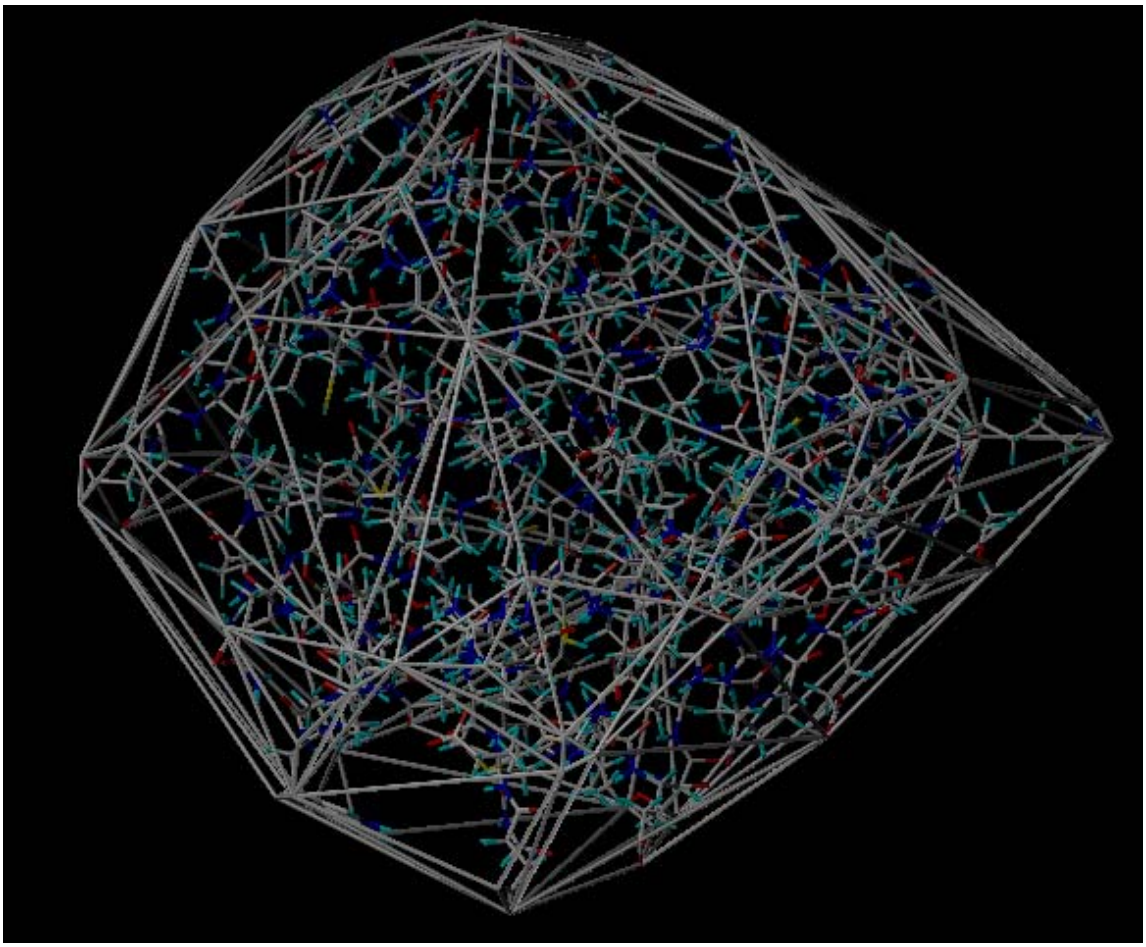

Figure 2. The Convex Hull of Flavodoxin

The volumes that lie inside the convex hull yet outside of the receptor are potential binding sites. The points that lie outside of the receptor are taken directly from the steric field points for Carbon atoms. Those grid points that have negative (attractive energies are considered outside of the protein. Note that this includes pockets that are completely contained inside the surface of the protein.

For a set of points to be a potential binding site, the set must be large enough to hold a ligand molecule. Points are removed from consideration if they have neighbors that are not also in the concave volume set of points. This process of point removal is done iteratively to remove layers of points. This process is called "onion-peeling". If a group of points is not thick enough, onion-peeling will remove the entire set. Thus only those sets of points that are large enough to contain a ligand molecule remain after the onion-peeling process. Typically, we apply onion-peeling to remove 1.5 angstroms from the outside of each set of points:

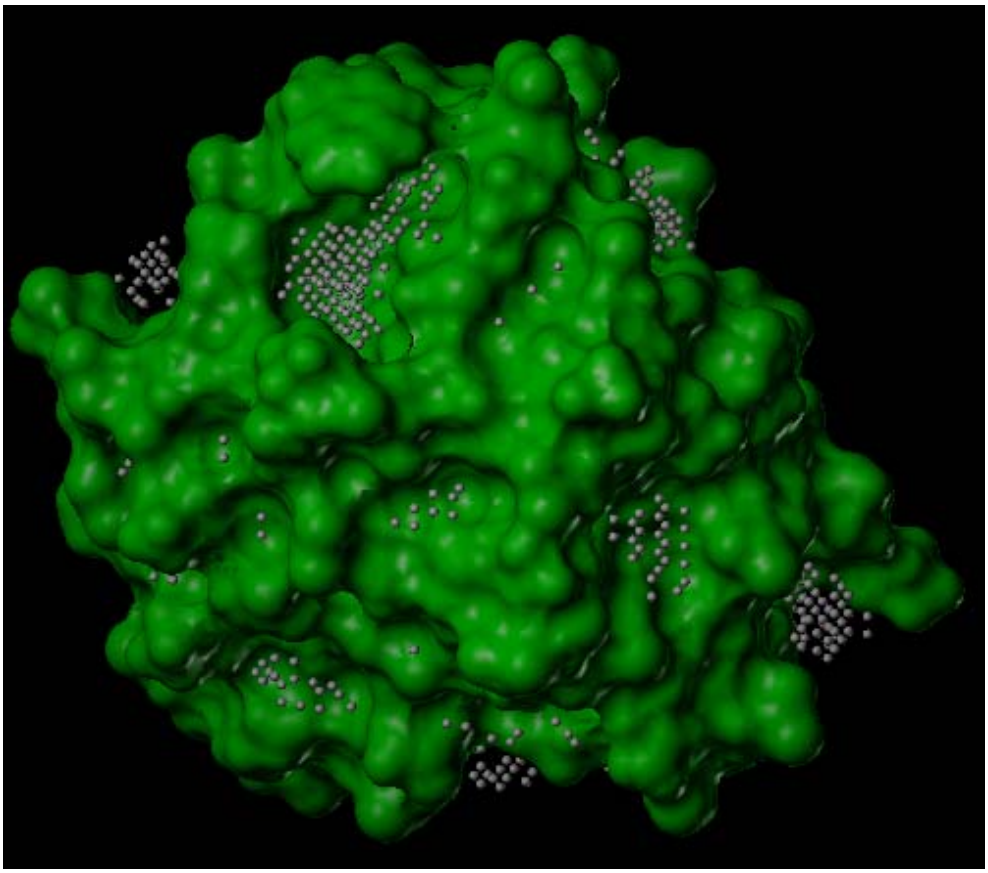

Figure 3. Contiguous points in deep clefts

The contiguous sets of points that remain (Figure 3) each represent a possible binding site. These sets are reduced to a representation that allows putative ligand to be placed in the site as an initial position for further minimization. We represent each site as a vector that starts at the centroid of the set of points, and extends along the direction of the first principal component (Figure 4). Potential ligands are thus aligned by placing their centroids at the site point centroid, and aligning their principal components with that of the site vector. Several rotations about the principal component vector may be tried as starting points for the minimization. It is also necessary to flip the ligand so that its first principal component is anti-aligned with that of the site vector, because the sense of the principal component vector is artificial.

If a particular binding site is very large, it is not possible for a single site vector to adequately represent the entire site, as it is possible for a ligand to bind in several places within the large cavity. When this happens, the 3DPL system creates the site vector at the center of the large site, and then removes the defining points within a certain radius of the centroid. Typically, the radius used is 8.0 angstroms. The remaining points – those outside the central sphere – are re-examined for additional site points. This allows large clefts to be filled with an appropriate number of site vectors. This method is specifically good at marking smaller cavities within the larger cavity.

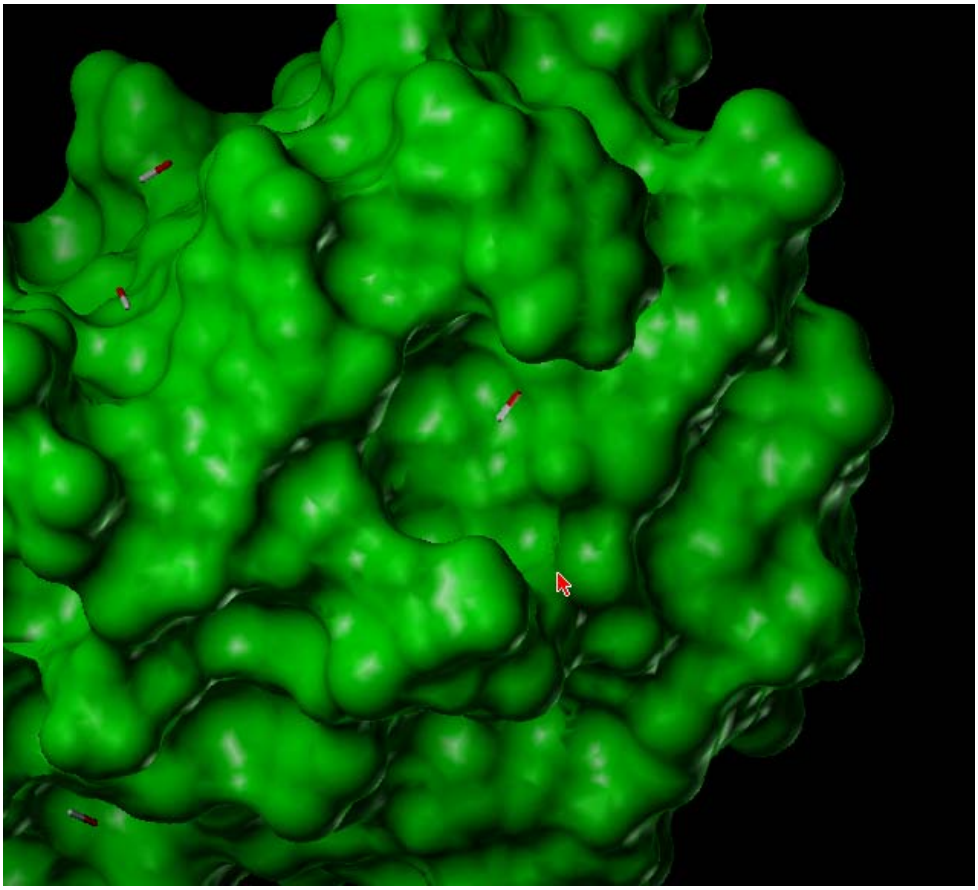

Figure 4. Site vectors for Flavodoxin

Occasionally, the shape of a cleft is itself concave, and the centroid of the cleft may not be within the cleft (Figure 5). When this occurs, the 3DPL system divides the cleft along the plane perpendicular to the first principal component of the concave site, and treats each half-site independently. If either of the half-sites so produced is also concave, it is treated in the same manner – it is divided in half again along the direction of its first principal component.

These methods allow the specification of a set of site vectors that well represent the possible binding sites of a receptor. Many of the vectors so produced are not actually binding sites. This readily becomes apparent as the searching progresses. After a few thousand compounds have been docked, those site vectors that have not resulted in the lowest energies can be optionally discarded. This process is referred to as site-focusing. Typically, only a few site vectors are responsible for docking, and these always include the known active site.

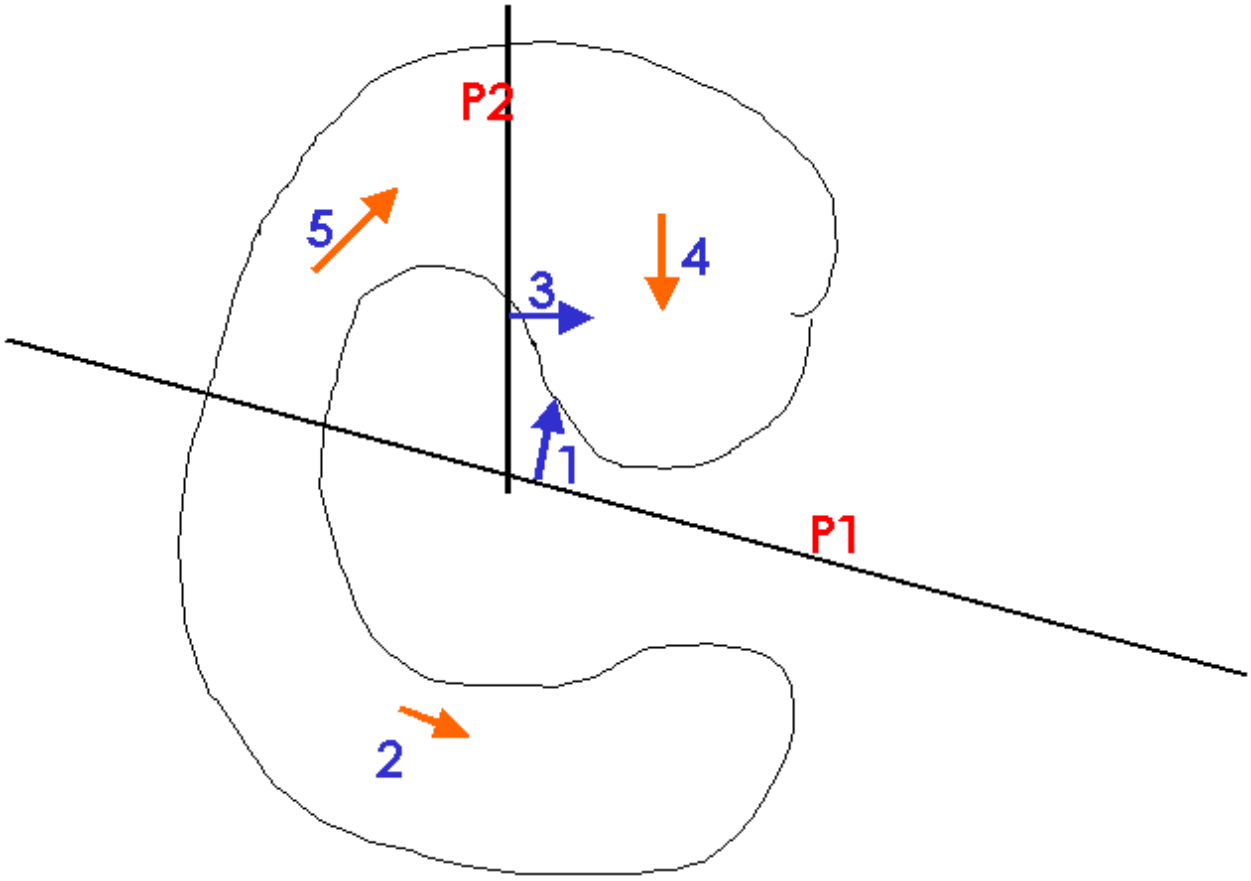

Figure 5. A concave site. Vector 1 is first produced, but is not in the site. Plane 1 (perpendicular to V1) is used to divide the site into two half sites, giving vectors V2 and V3. V3 is not within its half-site, so it is also used to divide the half-site at plane P2, producing vectors 4 and 5.

Validation

The 3DPL system has been tested on several test cases. In all cases, the 3-D structure of the protein or receptor was taken from the Protein Data Bank (PDB). These structures typically include a co-crystallized ligand that indicates the location of the known active site.

The validation studies fall into two classes:

1) Known ligand docking
2) Blind screening data docking

In the first case, the known ligands are extracted from the PDB files and seeded into a set of random structures from our library. The atomic coordinates of the ligand are modified from those found in the X-ray structure to remove any bias towards the experimentally determined coordinates. The docking system is then used to extract a subset of the test structures. Successful validation occurs when the known ligands are found among a small number of structures that are found to dock.

The second type of validation uses sets of structures for which the biological activity or actual binding data are already known or can be determined by testing. The 3DPL system is used to select a small set of structures predicted to bind. When these include a good number of those compounds actually found to be active, the system is validated.

In both types of validation, the performance of the system is measured using an enhancement ratio (ER). The ER is the ratio of the number of active/good hits found to the

number of hits that would be expected by random selection. ER is given by the following formula:

$ER = (L_{found}/L_{total}) / (C_{selected}/C_{total})$

where

L_{found} is the number of known ligands found by 3DPL

L_{total} is the total numver of known ligands in the dataset

$C_{selected}$ is the count of compounds, active and inactive, selected by 3DPL

C_{total} is the total count of compounds in the dataset

We have performed a number of known-ligand validation studies (Table 1). In all cases, the co-crystallized ligand is retrieved from among a random set of structures. In addition, the ligands are always found to dock in the original known binding site (Figure 6). The ER values shown indicate very successful docking.

Table 1. Results of docking co-crystallized ligands

PDF	Name	SitePoint Count	Ligand Count	Total Count	Selected Count	Ligands Expected (Random)	Ligands Found	ER
Flav	Flavodoxin	6	6	962	40	0.25	6	24.0
4phv	HIV Protease	5	3	962	40	0.12	2	16.0
1dwd	Thrombin	7	3	962	40	0.12	2	16.0
1c88	PTP-1b	31	7	982	40	0.29	7	24.6

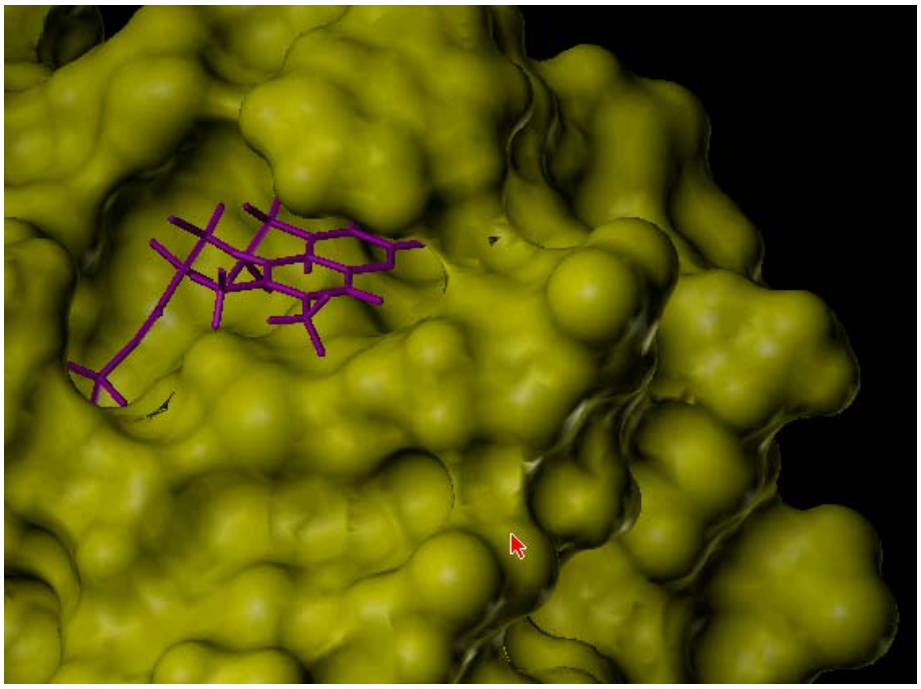

Figure 6. The co-crystallized ligand docked back into Flavodoxin

We have concluded several studies the uses 3DPL to predict biological activity of database compounds. The studies involved either known active molecules culled from the

literature or screened by collaborative partners. In all cases, 3DPL was able to select many more active molecules that would have been expected by random selection (Table 2). These tests span a wide variety of therapeutic areas. In one such test, 3.4 million compounds were screening to select 25 compounds. Samples of these 25 compounds were acquired, and tested in a standard Calcineurin screen. 4 of the compounds caused inhibition of the system.

Table 2. Results of Screening-Data Validation

PDF	Name	Site Point Count	Ligand Count	Total Count	Selected Count	Ligands Expected (Random)	Ligands (Found)	ER
1c88	PTP-1b	31	20	976	20	0.4	10	24.4
1a9u	Map Kinase P-38	5	21	3833	60	0.3	8	24.3
3std	Scytalone dehydratase	1	32	1006	50	1.6	8	5
6cha	alpha-chymotrypsin	5	97	1053	30	2,7	18	6.5
1fkj	FKBP-12	11	?	3.4 million	25	~0.25	4	>16

Conclusion

The 3DPL system is remarkably efficient in finding libraries to screen for biological activity when the structure of the receptor is known. The speed of the search system is such that it can be used for hundreds of proteins and receptors and millions of potential ligands.