Journal of Computer-Aided Molecular Design, 8 (1994) 231–242
ESCOM

J-CAMD 238

A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors

Renee L. DesJarlais* and J. Scott Dixon

Department of Physical and Structural Chemistry, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406-0939, U.S.A.

Received 25 June 1993
Accepted 22 September 1993

Key words: Ligand design; Molecular shape; DOCK; Database searching

SUMMARY

The program DOCK [1,2] has been used successfully to identify molecules which will bind to a specified receptor [3]. The original method ranks molecules based on their shape complementarity to the receptor site and relies on the chemist to bring the appropriate electrostatic or hydrogen bond properties into the molecular skeletons obtained in the search. This is useful when screening a small database of compounds, where it is not likely that molecules with both the correct shape and electrostatic properties will be found. As large databases are more likely to have redundant molecular shapes with a variety of functionality (e.g., members of a congeneric series), it would be useful to have a method which identifies molecules with both the correct shape and functionality. To this end we have modified the DOCK 1.0 method to target user-specified atom types to selected positions in the receptor site. The target sites can be chosen based on structural evidence, calculation or inspection. Targeted-DOCK improves the ability of the DOCK method to find the crystallographically determined binding mode of a ligand. Additionally, targeted-DOCK searches a database of small molecules at 100–1000 times the rate of DOCK 1.0, allowing more molecules to be screened and more sophisticated scoring schemes to be employed. Targeted-DOCK has been used successfully in the design of a novel non-peptide inhibitor of HIV-1 protease.

INTRODUCTION

Identification of molecules with a particular biological activity is a first step in drug development. As the availability of structures at atomic resolution for known or potential drug receptors has increased, the development of methods for exploiting this structural information in the design of ligands has become an active area of research. Software enabling the display of the 3D structure of molecules has allowed detailed examination of macromolecular structure and the interactive design of molecules to complement the shape and functionality present in a potential binding

*To whom correspondence should be addressed.

site. Recent advances in nongraphical computational techniques have allowed the exploration of many more potential ligands than can be examined interactively.

Approaches to finding ligands which bind to a given receptor fall into two categories: de novo design systems, which attempt to build a ligand to fit a binding site [4–6], and search systems, which screen a database of known structures for those complementary to a given receptor [2]. The program DOCK, developed by Kuntz and co-workers [1,2], has been extensively used to screen databases for molecules likely to bind to a macromolecular receptor and has proven successful in identifying molecules with the expected binding properties [3]. Initial versions of DOCK focused on shape exclusively. While the appropriate shape is necessary for achieving good van der Waals interactions, a tightly binding ligand will complement its receptor in electrostatic charge and hydrogen bonding characteristics as well. For a small database, where one might not expect to find ligands with both appropriate shape and appropriate hydrogen bonding groups, it has been useful to look at shape only and subsequently design the appropriate hydrogen bonding groups into the molecular frameworks identified in the shape search. For large databases, there may be significant redundancy in shape and a wide variety of functionality. If samples of the molecules in the database are available, one might be interested in identifying molecules which could be tested directly, without requiring new synthesis. In such cases, it would be desirable for a computer-based screen to return molecules which have electrostatic and hydrogen bond complementarity as well as the appropriate shape. Kuntz and co-workers have addressed this problem by developing a scoring function which incorporates both van der Waals and electrostatic energies [7].

Alternatively, one is often able to propose important interaction sites based on structural data, calculation, or interactive examination of a binding site. Where this knowledge is available, it would be useful to target such positions at the earliest possible phase of a DOCK calculation, as opposed to evaluating interactions at the end of the calculation via a scoring function. To this end, we have developed a method for including such specific interactions very early in the DOCK procedure. Our method, 'targeted-DOCK', allows for the specification of two or more positions within the binding site and the specification of atom types to occupy these sites. The DOCK 1.0 matching algorithm is then used with the initial pair of receptor centers and a pair of ligand atoms chosen from those specified. The results yield molecules with desired chemical interaction at two sites. This work commenced prior to the release of DOCK 2.0 and is currently only implemented with version 1.0. Our approach is similar to that developed by Leach and Kuntz as part of an effort to analyze the interactions of flexible ligands with macromolecular receptors [8].

This paper describes targeted-DOCK and its application to three test systems. Improved ability to locate the crystallographically determined binding mode of a ligand is illustrated for the proteins influenza hemagglutinin and thermolysin. The utility of targeted-DOCK in providing ideas for novel inhibitors is demonstrated with the design of a non-peptide inhibitor of HIV-1 protease which has been synthesized and is active [9].

Our modification improves the ability of DOCK to locate the experimentally determined orientation of a ligand in its binding site, has proven useful for pointing out substructures which have been incorporated into molecules with the desired activity, and provides a 100–1000-fold decrease in computation time required for a given search. This streamlined approach allows for more molecules to be searched and more sophisticated scoring schemes to be used.

METHODS

Program development

The basic outline of targeted-DOCK is similar to that of DOCK 1.0 [1,2]. The differences are restricted to the matching step and the determination of the rotation/translation matrix. In the matching step, targeted-DOCK forces the match to begin with two receptor spheres paired with two ligand atoms, all of which come from the user-specified list of target receptor spheres and atom types. When determining the rotation/translation matrix, targeted-DOCK allows the two target spheres and their paired atoms to be weighted more heavily than the other sphere/ligand atom pairs. The DOCK procedure is reviewed below with emphasis on these changes and the selection of the target receptor points.

First, a molecular surface [10] is generated around the binding site using the program MS [11]. The DOCK 1.0 module SPHGEN calculates a set of spheres to describe the complement of the binding-site shape (i.e., the volume available for ligand binding). Orientations of each potential ligand are generated by finding sets of atoms in a potential ligand and sets of receptor spheres, such that the distances between ligand atoms are within a specified tolerance of the distance from the receptor spheres paired with them. Targeted-DOCK differs from DOCK 1.0 in that the first pair of receptor spheres and ligand atoms must be from a user-specified list of spheres and atom types.

The selection of target spheres is critical to the success of targeted-DOCK. Sphere sites of interest need not be limited to the SPHGEN spheres. New spheres may be added based on the position of tightly bound waters seen in the crystal structure, the position of hydrogen bonding groups from a known ligand/macromolecular complex, an interactive examination of the binding site, or the results of a GRID [12] or similar calculation [5,13] which determines favorable interaction sites for a given functional group.

The DOCK 1.0 matching process begins by creating a list of all pairs of receptor spheres and ligand atoms. In targeted-DOCK, this list is created by pairing each of the target spheres to each of the ligand atoms whose type matches those specified by the user. The second sphere/atom pair is also selected from this special list, such that the distance between the two target receptor spheres is equal to the distance between the ligand atoms within a user-specified tolerance. If a second pair cannot be found, the molecule is skipped. From this point on, the matching proceeds as in DOCK 1.0 with any sphere or atom being eligible for addition to the set, provided the inter-sphere and inter-ligand distances are within the allowed tolerance. In a typical DOCK 1.0 run, inter-receptor sphere and inter-ligand atomic distances must be within 1.5–2.0 Å of each other to be paired. The typical number of sphere/atom pairs required for each match is 4–6. Orienting the small molecule by a least-squares fit of the atom centers onto their paired spheres results in general in a poor fit between any given atom and its paired sphere (i.e., an atom may be 1.0 Å or more away from its matched sphere). This is acceptable when the spheres are used to characterize a more global property, such as the overall shape of the active site, but when certain spheres are used to target specific interactions, it is desirable to have the target atoms very close to the target sphere centers when the ligand is oriented in the binding site. To achieve this, we have allowed for the tolerance between the target sphere/atom pairs to be different from the tolerance allowed for other sphere/atom pairs, and allowed the least-squares fit to be weighted in favor of a close match of the target spheres to their paired atoms.

As in DOCK 1.0, once a match is generated, the rotation/translation matrix that best fits the atoms of the small molecule to their paired sphere centers is used to orient the potential ligand into the macromolecular site. This orientation is scored using the DOCK 1.0 score. The score sums over all small molecule/macromolecule atom pairs, adding one to the score for each pair separated by a distance between the user-specified limits *concut* and *dmin*, an exponentially decreasing amount for pairs between the limits *dmin* and *discut* and none for pairs further than *discut* apart. An orientation with a small molecule/macromolecule contact of less than *concut* is eliminated from further consideration. Keeping the list of receptor atoms to a minimum reduces the time required to score each orientation. Since receptor atoms further than *discut* from the potential ligand do not contribute to the score, a list of all atoms which contribute to the molecular surface plus any within *discut* of these is sufficient. The highest scoring orientation of each molecule is checked against the set of high-scoring molecules examined so far and the best are saved.

Applications

Influenza hemagglutinin

The hemagglutinin protein of the influenza virus mediates the interaction of the virus with cell-surface receptors containing sialic acid. The X-ray crystallographic structure has been solved for a bromelin-cleaved soluble portion of hemagglutinin, complexed with sialic acid (Brookhaven Protein Data Bank [14,15], structure 4hmg [16]). In order to investigate the ability of targeted-DOCK to identify the experimentally determined binding mode of a ligand and to find other potential ligands in a database search, two experiments were performed on the hemagglutinin/sialic acid system. First, targeted-DOCK was used to position sialic acid in its binding site. In the second experiment, a database of molecules was searched to find those complementary to the hemagglutinin binding site.

A molecular surface [10,11] was calculated for all hemagglutinin residues within 8.0 Å of any atom of the crystallographically determined location of sialic acid. This surface was used as input for the SPHGEN module of the DOCK package to calculate spheres with radii between 1.4 and 5.0 Å. The positions of the sphere centers were examined graphically to find spheres, located such that an atom placed at the sphere center could make a favorable interaction with the protein.

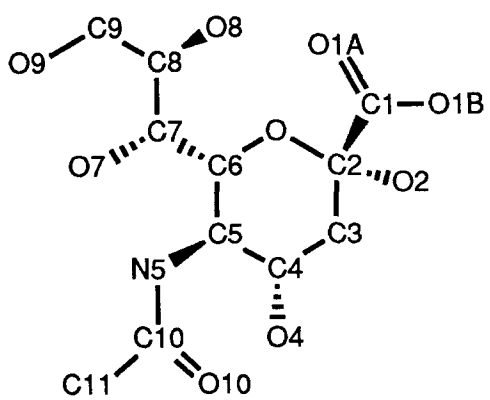


Fig. 1. The structure of sialic acid with atom numbering used in the Brookhaven PDB [13,14], structure 4hmg [15].

TABLE 1
PARAMETERS USED IN DOCK RUNS OF HEMAGGLUTININ, THERMOLYSIN AND HIV-1 PROTEASE

DOCK parameter	Hemagglutinin	Thermolysin	HIV-1 protease
Standard match			
Dislim[a] (Å)	1.5	1.5	1.5
Nodlim[b]	4	4	4
Special match			
Dsllim[c] (Å)	1.0	1.0	0.5
Limwt[d]	5.0	5.0	5.0
Scoring			
Concut (Å)	2.2	2.2	2.2
Dmin (Å)	3.5	3.5	3.5
Discut (Å)	5.0	5.0	5.0

[a] The tolerance between sphere/sphere and atom/atom distances for a standard match.
[b] The number of sphere/atom pairs required.
[c] The tolerance between target sphere/target sphere and special ligand atom/special ligand atom distances.
[d] The weight given to the target sphere/special ligand atom pairs in determining the rotation/translation matrix.

One sphere was identified whose position would allow hydrogen bonds with Glu^{190} and Ser^{228}, analogously to O9 of sialic acid. A second sphere was identified, positioned such that hydrogen bonds to Ser^{136} and Asn^{137} could be made, analogously to O1A of sialic acid (see Fig. 1 for the atom numbering of sialic acid). These two spheres were designated as the target spheres. Oxygen atoms were required to be matched to these spheres. For comparison, a DOCK 1.0 run with the same input was carried out. In both the targeted-DOCK run and the DOCK 1.0 run the same sphere set was used. This set included the SPHGEN spheres plus the two target spheres.

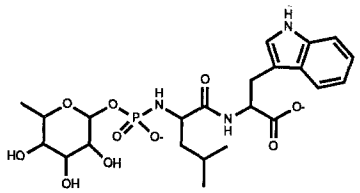
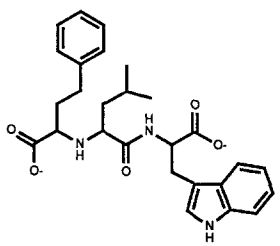
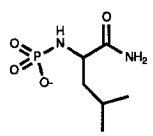
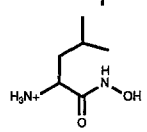
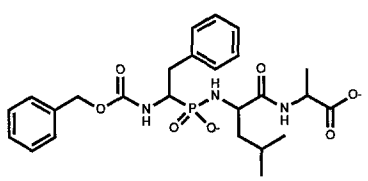
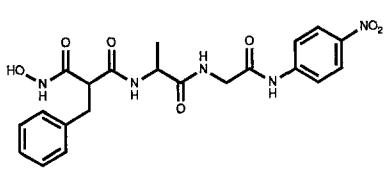
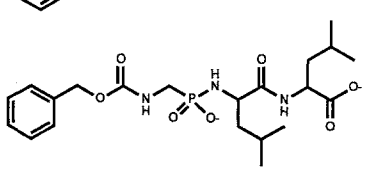
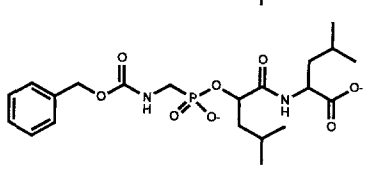
The parameters used for the DOCK run are listed in Table 1. All protein atoms that contributed to the molecular surface of the binding site as described above, plus any atom within 5.0 Å of an atom contributing to the surface, were used to determine the DOCK score.

The small molecule database consisted of all organic molecules from the Cambridge Structural Database (CSD) [17] not belonging to Cambridge classes 65–86. This was done to exclude any metal complexes. Because the CSD contains only the β-anomer of sialic acid and hemagglutinin binds the α-anomer, the sialic acid structure from 4hmg was added to the database.

Thermolysin

The X-ray crystallographic structure of the enzyme thermolysin has been solved both unliganded and complexed to a variety of inhibitors (see Table 2 for structures and references). These inhibitors vary both in their chemical structure and their inhibition constants, presenting an interesting test of our methodology. Table 2 lists the thermolysin structures for which data are available in the Brookhaven Protein Data Bank (PDB) [14,15]. To test the sensitivity of the method to differences in the protein structure and to the position of the special spheres, and to test the ability of targeted-DOCK to reproduce the crystallographically determined binding mode of the inhibitors, calculations were run to search for molecules with shape complementarity to the native thermolysin (3tln) and thermolysin from a complex with Cbz-Gly-ψ(PO_2NH)-Leu-Leu (5tmn). For each protein a molecular surface [10,11] was calculated for all residues within 12.0 Å of the active site zinc atom. This surface was used to create a set of spheres with radii between

TABLE 2
CRYSTAL STRUCTURES OF THERMOLYSIN COMPLEXES

Ligand	Structure	PDB file name
Phosphoramidon		1tlp [17]
N-(1-carboxy-3-phenylpropyl)-Leu-Trp		1tmn [18]
N-phosphorinyl-Leu-amide		2tmn [17]
Leu-hydroxylamine		4tln [19]
Cbz-Phe-ψ(PO_2NH)-Leu-Ala		4tmn [20]
HONH-benzylmalonyl-Ala-Gly-*p*NO_2-analide		5tln [19]
Cbz-Gly-ψ(PO_2NH)-Leu-Leu		5tmn [20]
Cbz-Gly-ψ(PO_2O)-Leu-Leu		6tmn [21]

1.4 and 5.0 Å. The search was directed to find molecules which could provide one oxygen to ligate the zinc and a second oxygen to accept a hydrogen bond from Arg^{203}. Neither structure provided SPHGEN-generated spheres in the appropriate positions to make one of these interactions. For the 3tln structure, new sphere positions were taken from the positions of water molecules 363 and 392 and used as targets. For the 5tmn structure, the positions of the two phosphonamidate oxygens and the oxygen of Leu^3 from the inhibitor Cbz-Gly-ψ(PO_2NH)-Leu-Leu were added as the target spheres. Oxygen atoms were required to match the target spheres in each of these cases. DOCK 1.0 runs were performed for comparison. The DOCK 1.0 and targeted-DOCK runs used the same sphere set. All protein atoms that contributed to the molecular surface of the binding site as described above, plus any atom within 5.0 Å of an atom contributing to the surface, were used to determine the DOCK score. The zinc itself was specifically excluded, because the optimal distance between zinc and a good ligand is smaller than the optimal nonbonded distance between other atoms in the system. The DOCK parameters used in these runs are listed in Table 1. The database derived from the CSD, with the thermolysin inhibitors from the PDB added, was searched.

HIV-protease

The protease produced by HIV-1 is essential for production of infectious virus and thus is an attractive target for the design of an anti-HIV agent [23]. Structures of this enzyme both unliganded [24] and complexed to a variety of inhibitors [25–31] are available. For our calculation, the complex of HIV-1 protease with the inhibitor MVT-101 [25] was used. A molecular surface [10,11] was calculated for all protein atoms within 8 Å of the bound MVT-101. SPHGEN was used to calculate spheres between 1.4 and 5.0 Å, which were used to characterize the shape of the active site. Two spheres were added to this list as the target spheres. Coordinates for the centers of these spheres were taken from the position of WAT^{511} in the MVT-101 structure and the coordinates of the hydroxyethylamine oxygen from the complex of HIV-1 protease with JG-365 [26], which is in the same reference frame as the MVT-101 structure. Oxygen atoms were targeted to these positions. Parameters for this calculation are listed in Table 1. All protein atoms that contributed to the molecular surface of the binding site as described above, plus any atom within 5.0 Å of an atom contributing to the surface, were used to determine the DOCK score. For the HIV-1 protease calculations, a database of 9561 SmithKline Beecham molecules, for which coordinates were generated with CONCORD version 2.9.3 [32], was searched. These molecules were selected to have at least one oxygen–oxygen distance of 5.87 ± 1.00 Å (the distance between the target spheres with ± 1.00 Å tolerance).

RESULTS

Hemagglutinin

The first experiment on the hemagglutinin system compares the sampling of orientation space for α-sialic acid with DOCK 1.0 and with targeted-DOCK. DOCK 1.0 finds 163 orientations of α-sialic acid. One of these orientations is 1.3 Å from the crystallographically determined position. The targeted-DOCK run finds just one orientation, 1.1 Å from that determined crystallographically. Although our modified algorithm explores many fewer orientations, it does focus the search on an area that agrees with the experimentally determined binding mode.

In the second hemagglutinin experiment, a database derived from the CSD [17], to which α-sialic acid was added, was searched for molecules with shape complementarity to the sialic acid binding site. We are able to search 33 070 molecules in 7.3 min with the targeted-DOCK procedure (All times stated are for runs on a single processor of an SGI 4D/380). This is substantially improved over a DOCK 1.0 run, which takes 66 h. In the DOCK 1.0 run, α-sialic acid is not among the top-200 scoring molecules. With targeted-DOCK, α-sialic acid ranks 73rd. It should be kept in mind that in both cases, the score used to rank the molecules takes into account only a van der Waals-like term which will favor large molecules over small, and that sialic acid itself is only a weak ligand (K_i = 2.8 mM) [33]. It is interesting to note that β-sialic acid ranks 18th in the targeted-DOCK run. Upon closer examination, however, it is clear that some of the polar groups of β-sialic acid are buried in hydrophobic areas. An energy minimization of the orientations of the α- and β-anomers with AMBER 3.0 Rev. A [34,35] reverses the order such that the α-anomer is preferred, in agreement with experiment. This result implies that more chemically reasonable ordering could be obtained with a scoring scheme that incorporates electrostatic interaction [7].

The hemagglutinin/sialic acid results may be compared to a similar method, CLIX [36]. This method matches three points in the receptor site to selected atoms, and uses a more sophisticated score which includes electrostatic interactions. Lawrence and Davis searched a database very similar to that used in our work. Both are derived from the CSD and contain ~30 000 molecules (31 042 for theirs and 33 071 for ours). They achieved a somewhat better orientation of α-sialic acid compared to the crystal structure (rms deviation 0.41Å), but their search requires 33 h on a machine which is comparable to ours. As expected for methods which rank molecules using different scoring schemes, there is very little overlap in the molecules discussed in the paper by Lawrence and Davis [36] and those found in our search.

TABLE 3
DOCK AND TARGETED-DOCK RESULTS FOR THERMOLYSIN

Ligand	Rms deviation (Å) from crystal structure[a]			
	3tln used as receptor		5tmn used as receptor	
	DOCK 1.0	Targeted-DOCK	DOCK 1.0	Targeted-DOCK
Phosphoramidon	11.17 (14 475)	–[b] (36)	0.82 (10 826)	–[b] (96)
N-(1-carboxy-3-phenylpropyl)-Leu-Trp	10.53 (13 545)	–[b] (12)	9.35 (10 762)	–[b] (96)
N-phosphorinyl-Leu-amide	7.77 (1888)	–[b] (17)	5.27 (1656)	1.52 (97)
Leu-hydroxylamine	8.60 (937)	–[c] (0)	8.54 (922)	0.60 (14)
Cbz-Phe-ψ(PO_2NH)-Leu-Ala	8.94 (13 286)	–[b] (45)	1.21 (9448)	1.83 (205)
HONH-benzylmalonyl-Ala-Gly-*p*-NO_2-analide	9.17 (6627)	8.94 (63)	7.83 (5777)	1.17 (84)
Cbz-Gly-ψ(PO_2NH)-Leu-Leu	10.44 (11 522)	1.31 (46)	5.86 (8430)	1.40 (184)
Cbz-Gly-ψ(PO_2O)-Leu-Leu	13.54 (11 541)	1.34 (78)	11.00 (8678)	1.37 (360)
CPU time to search 33 000 molecules (h)	437.25	0.79	331.13	2.75

[a] The number of orientations explored is given between brackets.
[b] All orientations resulted in close contacts between ligand and receptor atoms.
[c] No oxygen–oxygen distances were within tolerance of the distance between target spheres.

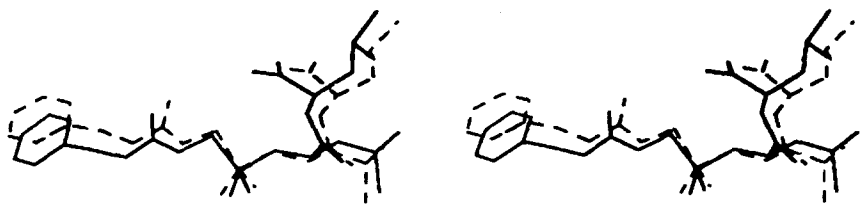


Fig. 2. A stereodiagram comparing the orientation of Cbz-Gly-ψ(PO_2NH)-Leu-Leu as found in the Brookhaven PDB [13,14], structure 5tmn [20] (solid lines) and as positioned by targeted-DOCK (dashed lines).

Thermolysin

The results of our calculations are summarized in Table 3. DOCK 1.0 is able to identify the correct orientation of only two thermolysin inhibitors and then only when 5tmn is used as the receptor. Targeted-DOCK is able to orient two of the eight inhibitors when 3tln is used for the receptor structure and six out of eight when 5tmn is used. These results show that targeted-DOCK is more successful than DOCK 1.0 in identifying orientations that are similar to the experimentally determined orientation, even though targeted-DOCK explores many fewer orientations and thus takes less CPU time than DOCK 1.0 to perform a search on a given database of molecules. In the case of thermolysin, targeted-DOCK obtains the best results when the positions of the target spheres are based on the positions of atoms in an inhibitor–protein complex. The positions of waters from the 3tln structure, which were used as target spheres, are different from the ligand atom positions from the 5tmn structure by about 0.6 Å. Because the thermolysin binding site is narrow, this shift causes more orientations to result in bad contacts with the receptor. Figure 2 shows the targeted-DOCK orientation and the crystallographically determined orientation of Cbz-Gly-ψ(PO_2NH)-Leu-Leu. Although the rms deviation of these two orientations is 1.40 Å, it is clear that the general features of the crystallographic orientation are duplicated. Only in the targeted-DOCK runs, known ligands occur within the top-200 scoring molecules. For the 3tln targeted run, 5tmn and 6tmn rank 104^{th} and 108^{th}, respectively, and for the 5tmn targeted run, 4tmn ranks 3^{rd}. Given that the scores are based on shape alone, it is not surprising that few of the known ligands rank highly, but targeted-DOCK does improve the ranking of known ligands. More improvement would be expected with a score based upon both shape and electrostatic properties (e.g. Meng et al. [7]).

HIV-1 protease

The HIV-1 protease provided an opportunity to apply the targeted-DOCK method in a system where results of the search could be tested experimentally for their ability to inhibit protease activity. Our intent was to select molecules for testing from our in-house database. After examination of the targeted-DOCK results it became obvious that, although the molecules found in the search do have some shape complementarity to the protease active site, most of them place hydrophilic groups in the hydrophobic S1 and S1' subsites, or fill these pockets inadequately. A common feature of these molecules, however, is a six-membered ring with *para*-hydroxyl groups overlaying the target spheres such that the hydroxyls can interact with the active-site aspartates (25 and 25') and the NH groups of Ile^{50} and $Ile^{50'}$ in the flaps. Molecule **I** was designed based on this central six-membered ring and knowledge of known inhibitors. Molecule **I** and its predicted interactions with HIV-1 protease are illustrated in Fig. 3. This molecule was found to inhibit

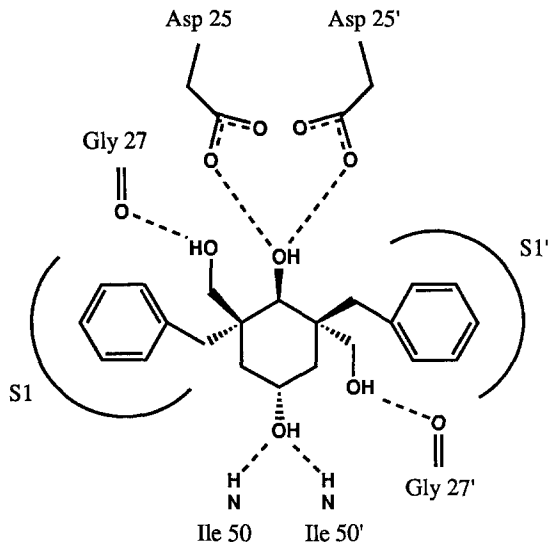


Fig. 3. Molecule **I**. Predicted interactions with HIV-1 protease are shown schematically.

HIV-1 protease with a K_i of 48 μM [9]. While this inhibition is weak compared with the best peptide-based protease inhibitors, molecule **I** is similar in size and activity to other dipeptide analogs. A number of derivatives of molecule **I** have been synthesized and their structure–activity relationships are consistent with the predicted binding mode [9].

DISCUSSION

Targeted-DOCK provides the ability to specify limited chemical complementarity in a shape-based search. The additional constraints make the method more efficient and better able to identify known binding modes than DOCK 1.0. Restricting the starting point of the match to particular sphere/atom pairs is a severe limit on the number of orientations that will be explored. The nonexhaustive and order-dependent nature of the DOCK matching algorithm does not guarantee that an interesting match from a DOCK 1.0 run will be present in a targeted-DOCK run, even though it may contain a pair of the selected spheres matched to a pair of the selected ligand atoms (D. Bodian, personal communication). In our experience, we have not missed an interesting orientation found by DOCK 1.0. As would be expected, targeted-DOCK is sensitive to the choice of target spheres. The results for thermolysin could likely be improved by the inclusion of more than one sphere as the target for a desired interaction or by describing a target by a volume instead of a point.

Currently, a limitation of targeted-DOCK is its use of a simple scoring routine that evaluates only the steric fit of a potential ligand with the receptor site. Improved scoring methods are now available [7] which include an electrostatic interaction energy term as well as a van der Waals energy term. Use of such a scoring scheme would be expected to order the hits in a more chemically reasonable way. Additionally, a new matching algorithm [37] has been developed which allows the user to control the number of matches explored and removes the order dependence

of the DOCK 1.0 matching. For DOCK generally, the use of SPHGEN spheres alone to characterize a binding site may be inadequate. The spheres as calculated with SPHGEN succeed reasonably well in characterizing the overall shape and size of a binding site, but can miss details. The addition of extra spheres to better characterize the shape details of a site could be accomplished by examination of experimental data, as we have done to target hydrogen bonding interactions, or automatically by allowing more than one sphere per atom bordering a site. Finally, the use of both a rigid ligand and a rigid receptor is a limitation shared by many design methods. Methods for including flexibility are being developed in a number of laboratories.

CONCLUSIONS

Targeted-DOCK provides an efficient means of searching a database of molecules for those with shape complementarity as well as limited chemical complementarity. Although relatively few orientations are explored, these appear to be sufficient to identify the experimentally determined orientation in most cases. This methodology has been used successfully in the design of a new class of HIV-1 protease inhibitors.

NOTE ADDED IN PROOF

After this paper had been accepted for publication, Lam et al. [38] published the design, synthesis and crystal structure of cyclic HIV-1 protease inhibitors resembling our molecule **I**. Their design focused on interactions similar to ours and they found that their molecules bound to the enzyme in the predicted orientation.

ACKNOWLEDGEMENTS

We wish to thank Geoffrey Dreyer and Balan Chenera for the synthesis of molecule **I** and their contributions to its design, Thaddeus Tomaszek and Thomas Meek for the enzyme inhibition assay of molecule **I**, and George Seibel for valuable discussions. We are grateful to Alexander Wlodawer who provided us with the coordinates for both the HIV-1 protease/MVT-101 complex and the HIV-1 protease/JG-365 complex prior to their release in the PDB.

REFERENCES

1 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
2 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
3 Kuntz, I.D., Science, 257 (1992) 1078.
4 Moon, J.B. and Howe, W.J., Protein Struct. Funct. Genet., 11 (1991) 314.
5 Böhm, H.-J., J. Comput.-Aided Mol. Design, 6 (1992) 61.
6 Lewis, R.A., Roe, D.C., Huang, C., Ferrin, T.E., Langridge, R. and Kuntz, I.D., J. Mol. Graphics, 10 (1992) 66.
7 Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
8 Leach, A.R. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 730.
9 Chenera, B., DesJarlais, R.L., Finkelstein, J.P., Eggleston, D.S., Meek, T.D., Tomaszek Jr., T.A. and Dreyer, G.B., Bioorg. Med. Chem. Lett., 3 (1993) 2717.
10 Richards, F.M., Annu. Rev. Biophys. Bioeng., 6 (1977) 151.

11 Connolly, M.L., Science, 221 (1983) 709.
12 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
13 Miranker, A. and Karplus, M., Protein Struct. Funct. Genet., 11 (1991) 29.
14 Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
15 Abola, E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J., In Allen, F.H., Bergerhoff, G. and Sievers, R. (Eds.) Crystallographic Databases – Information Content, Software Systems, Scientific Applications, Data Commission of the International Union of Crystallography, Cambridge, 1987, p. 107.
16 Weis, W.I., Bruenger, A.T., Skehel, J.J. and Wiley, D.C., J. Mol. Biol., 212 (1990) 373.
17 Allen, F.H., Bellard, S.A., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R. and Watson, D.G., Acta Crystallogr., B35 (1979) 2331.
18 Tronrud, D.E., Monzingo, A.F. and Matthews, B.W., Eur. J. Biochem., 157 (1986) 261.
19 Monzingo, A.F. and Matthews, B.W., Biochemistry, 23 (1984) 5724.
20 Holmes, M.A. and Matthews, B.W., Biochemistry, 20 (1981) 6912.
21 Holden, H.M., Tronrud, D.E., Monzingo, A.F., Weaver, L.H. and Matthews, B.W., Biochemistry, 26 (1987) 8542.
22 Tronrud, D.E., Holden, H.M. and Matthews, B.W., Science, 235 (1987) 571.
23 Debouck, C., AIDS Res. Hum. Retroviruses, 8 (1992) 153.
24 Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, I.M., Clawson, L., Schneider, J. and Kent, S.B.H., Science, 245 (1989) 616.
25 Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, A., Science, 246 (1989) 1149.
26 Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Kent, S.B.H. and Wlodawer, A., Proc. Natl. Acad. Sci. USA, 87 (1990) 8805.
27 Erickson, J., Neidhart, D.J., Van Drie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M., Science, 249 (1990) 527.
28 Jaskolski, M., Tomasselli, A.G., Sawyer, T.K., Staples, D.G., Heinrikson, R.L., Schneider, J., Kent, S.B.H. and Wlodawer, A., Biochemistry, 30 (1991) 1600.
29 Bone, R., Vacca, J.P., Anderson, P.S. and Holloway, M.K., J. Am. Chem. Soc., 113 (1991) 9382.
30 Dreyer, G.B., Lambert, D.M., Meek, T.D., Carr, T.J., Tomaszek Jr., T.A., Fernandez, A.V., Bartus, H., Cacciavillani, E., Hassell, A.M., Mimmich, M., Petteway Jr., S.R., Metcalf, B.W. and Lewis, M., Biochemistry, 31 (1992) 6646.
31 Dreyer, G.B., Boehm, J.C., Chenera, B., DesJarlais, R.L., Hassell, A.M., Meek, T.D., Tomaszek Jr., T.A. and Lewis, M., Biochemistry, 32 (1993) 937.
32 Rusinko III, A., Skell, J.M., Balducci, R., McGarity, C.M. and Pearlman, R.S., CONCORD, University of Texas, Austin, TX and Tripos Associates, St. Louis, MO, 1988.
33 Sauter, N.K., Bednarski, M.D., Wurzburg, B.A., Hanson, J.E., Whitesides, G.M., Skehel, J.J. and Wiley, D.C., Biochemistry, 28 (1989) 8388.
34 Weiner, P.K. and Kollman, P.A., J. Comput. Chem., 2 (1981) 287.
35 Seibel, G.L., AMBER, Version 3.0A, University of California, San Francisco, CA, 1989.
36 Lawrence, M.C. and Davis, P.C., Protein Struct. Funct. Genet., 12 (1992) 31.
37 Shoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 380.
38 Lam, P.Y.S., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., Chang, C.-H., Weber, P.C., Jackson, D.A., Sharpe, T.R. and Erickson-Viitanen, S., Science, 263 (1994) 380.