formed by heating the mixture at 110 °C for 30 min.

Materials. Bacteriological peptone and yeast extract powder were purchased from OXOID Limited, Basingstoke, Hampshire, England. Sterile donor horse serum was obtained from Flow Laboratories. Benzylpenicillin G was a generous gift from Gist-brocades N.V., Delft, The Netherlands. All chemicals used were of the highest obtainable quality.

Apparatus. Optical density of growing cultures were determined at 660 nm with a Zeiss PMQ3 spectrophotometer. pH measurements were performed with a saturated calomel electrode. Test tubes were incubated in a water bath at 37 °C.

Test **Organism.** *M. gallisepticum* K514, kindly supplied by the research management of Gist-brocades N.V., was used as the test organism. *M. gallisepticum* strains can be stored at −20 °C for several months.[23] After being thawed at room temperature, the culture was transferred to a bottle with fresh Adler medium in such a way that the original culture was diluted 10 times. The culture was incubated overnight at 37 °C. When the pH of the culture had dropped to 6.8 and the density (determined as A_{660nm}) had reached a value of 0.22, the culture was used for inoculation purposes. The remaining part was stored at −20 °C.

Determination of Antimycoplasmal Activity. The antimycoplasmal activity of all compounds was determined in the presence or the absence of copper and expressed as the minimal inhibitory concentration (MIC). In the former case, the final concentration in the test tube was 40 μM $CuSO_4$. Tylosin and compound 1 were included as controls in every test. All compounds were dissolved in dimethyl sulfoxide whereas Tylosin was dissolved in water. It was established that DMSO in the final concentration in the Adler medium (1.25%) has no effect on mycoplasmal growth. Serial twofold dilutions (in duplicate) of test compounds were made in Adler medium. Each tube, containing 3 mL of medium, was inoculated with 1 mL of a fresh culture of *M. gallisepticum* K514, and these mixtures were incubated at 37 °C for 24 h. Mycoplasmal growth was indicated by a change in color of the indicator present in the medium. The minimal inhibitory concentration was determined as the lowest concentration that did not cause a change in color.

Data Processing. Statistical correlations were performed by using a commercial multiple linear regression program (Statworks, Cricket Software Inc., Philadelphia, PA). The figures in parentheses are the standard errors of regression coefficients. The parameters included in each equation are significant on a 1% level. For a given equation, n is the number of compounds, r is the multiple correlation coefficient, s is the standard error of estimate and F represents the value of the F test.

Acknowledgment. These investigations were supported by the Netherlands Technology Foundation (STW). We thank John Ch. Eriks and Greetje Bijloo for their valuable advice concerning synthesis and regression analysis, respectively.

Registry No. 1, 37989-04-1; **2a**, 112575-42-5; **2b**, 112575-43-6; **2c**, 112575-44-7; **2d**, 112575-45-8; **2e**, 112575-46-9; **2f**, 112575-47-0; **3a**, 112575-48-1; **3b**, 112575-49-2; **3c**, 112575-50-5; **3d**, 112575-51-6; **3e**, 112575-52-7; **3f**, 112575-53-8; **3g**, 112575-54-9; **3h**, 112575-55-0; **3i**, 112575-56-1; **3j**, 112575-57-2; **3k**, 112575-58-3; **3l**, 112575-59-4; **3m**, 112575-60-7; H_3CCOCl, 75-36-5; H_5C_2COCl, 79-03-8; $(CH_3)_2CHCOCl$, 79-30-1; $(CH_3)_3CCOCl$, 3282-30-2; $C_4H_9CH(C_2H_5)COCl$, 760-67-8; $C_9H_{19}COCl$, 112-13-0; C_6H_5COCl, 98-88-4; $4\text{-}H_3CC_6H_4COCl$, 874-60-2; $4\text{-}H_3COC_6H_4COCl$, 100-07-2; $4\text{-}ClC_6H_4COCl$, 122-01-0; $3,4\text{-}Cl_2C_6H_3COCl$, 3024-72-4; $3\text{-}H_3CC_6H_4COCl$, 1711-06-4; $3\text{-}H_3COC_6H_4COCl$, 1711-05-3; $3\text{-}ClC_6H_4COCl$, 618-46-2; $2,4\text{-}(CH_3)_2C_6H_3COCl$, 21900-42-5; $2,5\text{-}(CH_3)_2C_6H_3COCl$, 22328-43-4; $2,6\text{-}(CH_3)_2C_6H_3COCl$, 21900-37-8; $3,4\text{-}(CH_3)_2C_6H_3COCl$, 21900-23-2; $3,5\text{-}(CH_3)_2C_6H_3COCl$, 6613-44-1; $2\text{-}H_3CC_6H_4CN$, 529-19-1; 2-cyanopyridine, 100-70-9.

Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure

Renee L. DesJarlais,[†] Robert P. Sheridan,[‡] George L. Seibel,[†] J. Scott Dixon,[‡§] Irwin D. Kuntz,*[†] and R. Venkataraghavan[‡]

Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143, and Medical Research Division, Lederle Laboratories, Pearl River, New York 10965. *Received April 27, 1987*

Finding novel leads from which to design drug molecules has traditionally been a matter of screening and serendipity. We present a method for finding a wide assortment of chemical structures that are complementary to the shape of a macromoleculer receptor site whose X-ray crystallographic structure is known. Each of a set of small molecules from the Cambridge Crystallographic Database (Allen; et al. *J. Chem. Doc.* **1973**, *13*, 119) is individually docked to the receptor in a number of geometrically permissible orientations with use of the docking algorithm developed by Kuntz et al. (*J. Mol. Biol.* **1982**, *161*, 269). The orientations are evaluated for goodness-of-fit, and the best are kept for further examination using the molecular mechanics program AMBER (Weiner; Kollman *J. Comput. Chem.* **1981**, *106*, 765). The shape-search algorithm finds known ligands as well as novel molecules that fit the binding site being studied. The highest scoring orientations of known ligands resemble binding modes generated by interactive modeling or determined crystallographically. We describe the application of this procedure to the binding sites of papain and carbonic anhydrase. While the compounds recovered from the Cambridge Crystallographic Database are not, themselves, likely to be inhibitors or substrates of these enzymes, we expect that the structures from such searches will be useful in the design of active compounds.

The process of developing a new drug is long and complicated. The first step in this process is to find a "lead": a compound active in a particular therapeutic area. Analogues of the lead compound are made in order to study the way that its properties vary with structural changes. The resulting structure–activity data can be analyzed to optimize the overall profile of the potential drug. Lead compounds are usually found by screening many compounds or are discovered accidentally. There exists a good deal of expertise in making analogues and optimizing the lead compound's properties, but finding such compounds is still largely a matter of chance. In this paper, we attempt to assist drug design by finding structural precursors from which lead compounds can be designed when the leads are to be ligands for a receptor of

[†] University of California at San Francisco.
[‡] Lederle Laboratories.
[§] Present address: Smith Kline French Laboratories, King of Prussia, Pennsylvania 19406.

0022-2623/88/1831-0722$01.50/0 © 1988 American Chemical Society

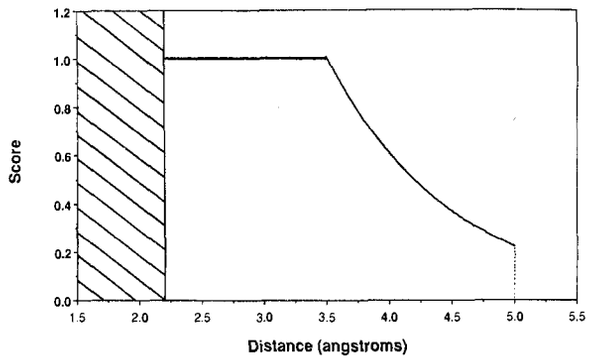


Figure 1. The score assigned to a receptor atom–candidate atom pair versus the distance between them. Pairs with a distance greater than 5.0 Å the score assigned is zero. The total score for an orientation is the sum over the scores for all receptor atom–candidate atom pairs. If any pair is shorter than 2.3 Å, then the orientation is dropped from consideration. All the distance parameters are set by the user.

known three-dimensional structure.

Recent advances in biochemistry allow us to understand better how drugs interact with their macromolecular receptors. For example, the structures of several protein receptors of medicinal interest are known at the atomic level from X-ray crystallography.[1-3] In other cases, the structure of a related protein is known and can often be used as a model for the actual target receptor.[4] The development of molecular modeling with computer graphics has made it possible to easily visualize these receptors and their interactions with ligands, which might be substrates, inhibitors, agonists, antagonists, or allosteric effectors.[5,6] From such studies, it is clear that a tightly binding ligand possesses both chemical and shape complementarity to the receptor binding site. In this paper we make two assumptions. The first assumption is that shape complementarity is a useful starting point for designing novel ligands. The second assumption is that a large collection of molecular structures such as the Cambridge Crystallographic Database[7] will contain a number of useful molecular shapes. Given a macromolecular receptor for which a structure is known at the atomic level, and a "candidate" ligand from the small molecule database, we use the automatic procedure developed by Kuntz et al.[8] to dock the candidate ligand into the receptor binding site. After being docked, the candidates are sorted by a simple scoring routine that measures their fit to the receptor. The method is able to find ligands known to bind to a particular site and can dock them correctly. It also provides reasonable dockings of novel molecules that are complementary in shape to the binding site. Then these molecules are examined by using interactive computer graphics.[5,9] We observe that those candidates with the highest score fit the bumps and grooves of the site best. We anticipate that these molecules could serve as the beginning of a design effort; other methods must be combined with the shape analysis to suggest molecular structures that are complementary to the site in chemical interactions as well as shape. We stress that we are not aiming at a comprehensive search for all potential ligands, all plausible conformers of a given ligand, or all possible docking geometries of a particular conformer. Our goal is to retrieve a set of interesting and novel molecular fragments that can lead, ultimately, to the development of new pharmaceutical agents.

Scheme I

A. Read candidate coordinates
B. Calculate atom/atom distances for candidate
C. Match atom/atom distances from the candidate to sphere/sphere distances in the receptor
D. Calculate for each match:
   (1) Rotation/translation matrix to orient atom centers onto paired spheres
   (2) Orientation of candidate based using rotation/translation matrix
   (3) Score for each orientation
E. Save best orientation for the candidate

Methods

Our basic approach to explore the docking of two rigid objects uses a set of algorithms that develop a molecular surface for the receptor,[10-13] produce a negative image of the receptor site,[8] and match the molecular structure of a potential ligand to the negative image of the receptor.[8] The method explores the full six degrees of freedom required to position two rigid objects. This work has been described in earlier publications.[8,14] The major new feature in this paper is the ability to search a large structural database for interesting molecular shapes.

Briefly, the docking algorithm represents the negative image of the shape of the receptor site with a set of spheres. The spheres are constructed from the Connolly rendering[10-12] of the molecular surface described by Richards.[13] Each sphere touches the surface at two points and has its center along the direction of the surface normal to one of the points. Spheres are generated between all pairs of surface points whose normals are directed toward each other. The program keeps the smallest sphere of those generated for each surface point. This eliminates spheres that intersect the protein. Each receptor atom has many surface points, each point with its smallest sphere. We keep only the largest sphere associated with each atom. This selection assures that the set of spheres span the site. We have found that keeping only a single sphere for each atom that borders the receptor site is sufficient to characterize the shape of the site. More could be used for complex sites if desired. In general there are many grooves or invaginations in a macromolecular surface, each such site characterized by a set of overlapping spheres of various sizes. Definition of the possible binding sites is made by the program automatically. The user must then specify

(1) Liljas, A.; Kannan, K. K.; Bergsten, P.-C.; Waara, I.; Fridborg, K.; Strandberg, B.; Carlbom, U.; Jarup, L.; Lovgren, S.; Petef, M. *Nat. New. Biol.* **1972**, *235*, 131.
(2) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650.
(3) Hardy, L. W.; Finer-Moore, J. S.; Montfort, W. R.; Jones, M. O.; Santi, D. V.; Stroud, R. M. *Science (Washington, D.C.)* **1987**, *235*, 448.
(4) Cushman, D. W.; Cheung, H. S.; Sabo, E. F.; Ondetti, M. A. *Biochemistry* **1977**, *25*, 5484.
(5) Langridge, R.; Ferrin, T. E.; Kuntz, I. D.; Connolly, M. L. *Science (Washington, D.C.)* **1981**, *211*, 661.
(6) Blaney, J. M.; Jorgensen, E. C.; Connolly, M. L.; Ferrin, T. E.; Langridge, R.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. *J. Med. Chem.* **1982**, *25*, 785.
(7) Allen, F. H.; Kennard, O.; Motherwell, W. D. S.; Town, W. G.; Watson, D. G. *J. Chem. Doc.* **1973**, *13*, 119.
(8) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
(9) Huang, C.; Jarvis, L.; Ferrin, T. E.; Langridge, R. UCSF MIDAS: Molecular Interactive Display and Simulation, 1983.
(10) Connolly, M. L. Thesis, University of California, 1981.
(11) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548.
(12) Connolly, M. L. *Science (Washington, D.C.)* **1983**, *221*, 709.
(13) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(14) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. *J. Med. Chem.* **1986**, *29*, 2149.

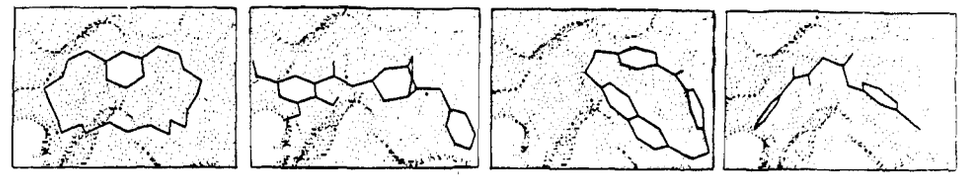


Figure 2. The orientations of three of the papain candidate ligands found by our program and the known substrate, 3-iodophenyl hippurate, are shown in the papain binding site. The dots represent the molecular surface of the papain site. The molecules illustrated are (A) molecule **1**, (B) molecule **3**, (C) molecule **4**, and (D) molecule **5**. An adequate two-dimensional projection of molecule **2** could not be found.

which of these sites is of interest. Typically the largest set is the crystallographically determined binding site of a receptor, but in some cases the smaller sites may be of interest (e.g. for allosteric effectors). For the enzymes discussed in this paper, the receptor sites are large, well-defined invaginations that extend considerably beyond the catalytic residues. Generating the spheres for the active site of a typical protein receptor takes approximately 1 h of CPU time on a VAX 8600 with the UNIX operating system. This time is dependent on the size of the binding site.

The small molecule database should ideally span a wide range of molecular shapes while minimizing redundancies. Our shape data base was derived from the July and October updates in 1985 to the Cambridge Crystallographic Database.[7] Using the Cambridge Crystallographic Database is attractive for a number of reasons. First, the three-dimensional coordinates are already present and need not be generated by the user. Further, the molecules in the Cambridge Database represent a wide variety of chemical structures. Finally, the conformation present in the crystal is likely to be one of the low energy structures for each molecule. A minimum of processing was required. For each structure, hydrogen atoms and counterions were removed, and the reference codes, atom names, and Cartesian coordinates were stored in the shape database for use by the shape search program. The particular subset of the Cambridge Crystallographic Database[7] that we used (about 2700 structures) was chosen arbitrarily for demonstration purposes.

The matching program docks the candidate ligands from the shape database into the receptor binding site in the following way (steps C and D in Scheme I): The distances between sphere centers of the site are matched to the distances between atom centers of a candidate ligand. If a particular distance (d_{ij}) between spheres i and j is within a cutoff C (typically 1.5 Å) of the distance ($d_{a,b}$) between candidate atoms a and b, then a sphere center k and atom center c are sought such that $d_{i,k} = d_{a,c} \pm C$ and $d_{j,k} = d_{b,c} \pm C$. For any group of n atoms to be matched to a group of n spheres, all $(n)(n-1)/2$ interatomic distances and corresponding intersphere distances must meet the cutoff criterion. Atoms and spheres are added to the list until a violation occurs. The value of the cutoff, C, can be set by the user. A rotation/translation matrix is calculated by using the Ferro and Hermans least squares fitting algorithm[15] to find the best superposition of each atom onto its paired sphere. There must be at least four sphere centers matched to four atom centers (corresponding to six matched distances) to determine a unique rotation/translation matrix. The program systematically examines the matching of every atom to every sphere. While this is not an exhaustive search, several hundred matches are found for each candidate in the database.

The rotation/translation matrix generated from a match is applied to all atoms of the candidate to orient it in the receptor site. Each match corresponds to a different orientation of the candidate in the active site. The program generates a large number of orientations and they must be evaluated efficiently. We use a simple scoring routine based on candidate/receptor interatomic distances to screen the orientations. Distances between all the receptor atoms and the atoms of the oriented candidate are evaluated according to eq 1. Very short distances (typically less than 2.3 Å) represent excessive atomic overlap and a poor score is returned for the orientation. If there are no such distances, then for each receptor atom-candidate atom distance between 2.3 and 3.5 Å the score is increased by one. For distances between 3.5 and 5.0 Å, the amount added to the score decreases with increasing distance. For distances that are greater than 5.0 Å, nothing is added to the score (Figure 1). There is nothing unique about this function form. It was chosen to identify orientations that have favorable receptor contact in a quick and convenient way and to eliminate orientations that make bad van der Waals contact with the receptor. For each candidate, the orientation with the best score is kept for further calculation. The user may select the number of top scoring candidates to be saved. The scoring function tends to give higher scores to the candidates with more atoms.

if $d_{i,j} < 2.3$ Å for any i,j then
  score = −999.0
else

$$\text{score} = \sum_{i}^{\text{receptor atoms}} \sum_{j}^{\text{candidate atoms}} F_{i,j}$$

$$F_{i,j} = \begin{cases} 1.0 & \text{if } 2.3 \text{ Å} \le d_{i,j} \le 3.5 \text{ Å} \\ \exp[-(d_{i,j} - 3.5 \text{ Å})^2] & \text{if } 3.5 \text{ Å} < d_{i,j} \le 5.0 \text{ Å} \\ 0.0 & \text{if } 5.0 \text{ Å} < d_{i,j} \end{cases} \quad (1)$$

The matching, least-squares fitting, and scoring are performed by a single program. It is written in FORTRAN and runs on a number of machines. The work reported here was done on a Floating Point Systems 264 array processor with a VAX host. It takes approximately 2.5 h of CPU time on this machine to go through the 2700 candidates in our database. The same run would take about 40 h on a VAX 780 without an array processor. The time required is dependent on the size of the binding site, the matching parameters used, and the number of molecules in the database searched.

Finally, we optimize the steric fit between the high-scoring candidates and the receptor. Our aim is to assess the difficulty of relieving van der Waals contacts that may be present after the docking step. These candidates are minimized in the receptor size by using the molecular mechanics program AMBER,[16] with the united atom force field developed by Weiner et al.[17] Since we are concerned

(15) Ferro, D. R.; Hermans, J. *Acta Crystallogr., Sect. A* **1977**, *A33*, 345.

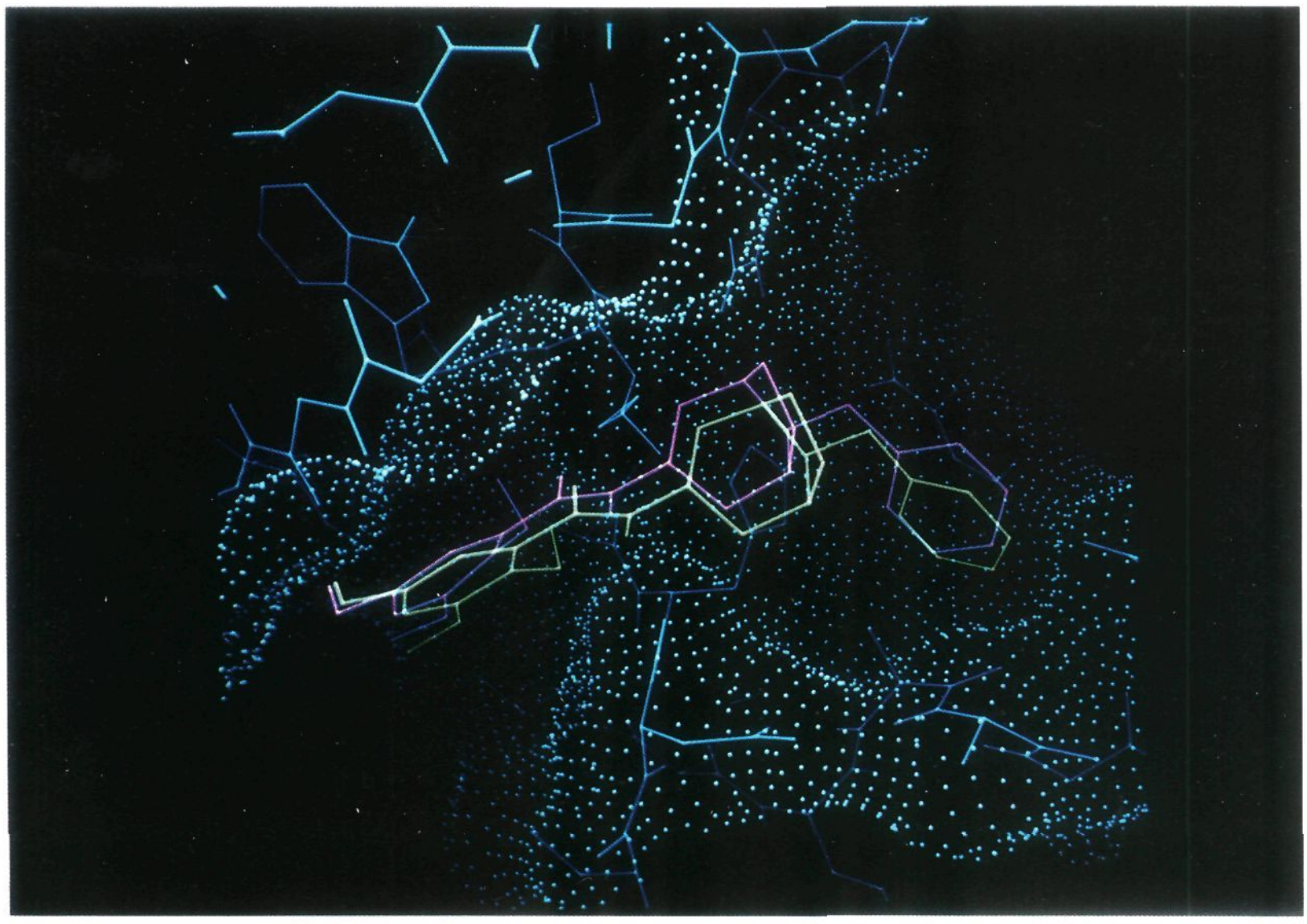


Figure 3. Molecule **3** is shown docked in the active site of papain. Two orientations are shown: the orientation that is found by the search procedure (green), and the orientation after 500 cycles of AMBER[16,17] minimization (purple). The protein and its molecular surface before minimization are shown. The structure is not altered much during the minimization.

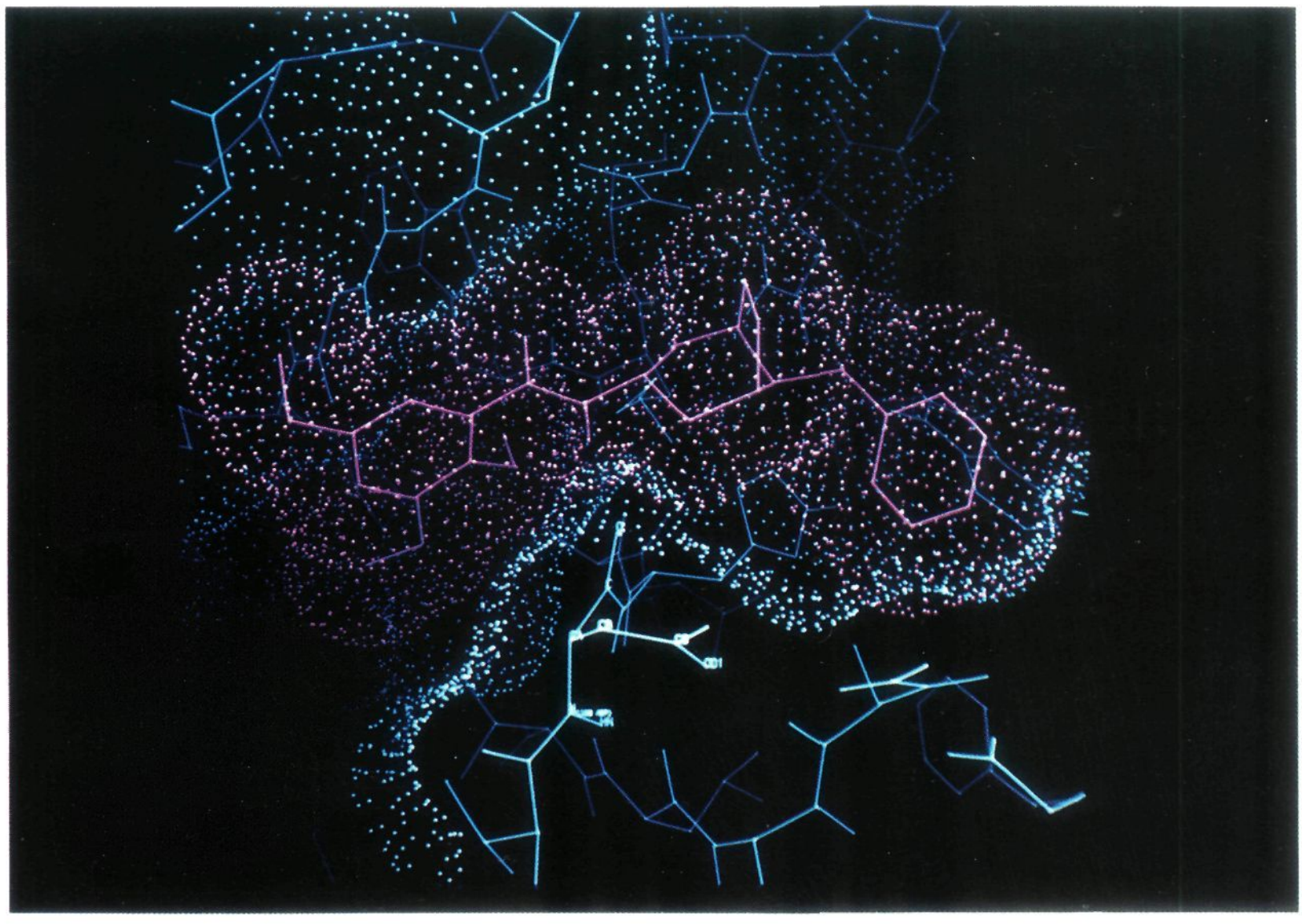


Figure 4. Molecule **3** docked in the papain active site. Papain is in blue and its molecular surface is displayed. Molecule **3** is in purple and is shown with its van der Waals surface. Molecule **3** is shown after 500 cycles of AMBER[16,17] minimization.

only with complementarity of shape, we remove the electrostatic interactions involving the candidate by giving its atoms zero charge. The atoms in the receptor have their usual charges. During the energy minimization, the allowed motions include all degrees of freedom of the candidate and all degrees of freedom (backbone and side chain) of the amino acid residues of the protein receptor that have any atom within 5.0 Å of any of the atoms of any of the candidates docked in the site. The positions of other protein residues are fixed. The same residues are mobile in each minimization of a given site. The minimization step relieves van der Waals contacts between the protein and candidate. To get a better assessment of the goodness-of-fit than can be inferred from eq 1, we have also minimized the energy of the free receptor and the free candidate. The final energies of the separate molecules are subtracted from the energy of the minimized complex. This result approximates the van der Waals energy of interaction between the protein and candidate and should groups separated by a less bulky chain. The other two are quite different. Each fits snugly into the site despite their structural differences. Molecules 2 and 3 fit into the site provide a reasonable basis of comparing the quality of the geometric match for the different ligand candidates.

Results

We have performed calculations on two protein receptor test cases: papain and carbonic anhydrase. Each is discussed separately below.

Papain. Papain is a sulfhydryl protease with broad substrate specificity. The binding site is an elongated groove in the protein's surface. The groove has a pocket at each end and separated by a ridge. We used the X-ray crystal structure of papain 8PAP[18] from the Brookhaven Protein Data Bank[19] for our calculation.

We searched the database of candidates for those with shape complementarity to the papain site. For the distance matching, we used a cutoff, C of 1.5 Å and required at least four candidate atom–receptor sphere pairs per match. The orientations of each candidate were scored using the parameter values discussed in the previous section. A large number of the small molecules fit neatly into the papain receptor site. We show four of the more diverse ones below.

For comparison, a known substrate, 3-iodophenylhippurate (**5**), is included. This molecule was also put in our database as a test for whether we were able to recover known ligands. The phenyl hippurate ranks low on the list of best shape fits because our scoring routine is biased toward larger molecules. It is recovered, however, in a reasonable orientation. Molecules **1**–**5** were minimized for 500 cycles by using the molecular mechanics program AMBER[16,17] as described above. The molecular mechanical energies and the molecules' initial scores are reported in Table I. The fact that each complex yields a negative interaction energy indicates that all repulsive atomic overlaps have been removed in the minimized structures. The orientations of molecules **1**, **3**, **4**,and **5** are illustrated in Figure 2. Molecules **2** and **3** have some resemblance to the known substrate, because they have two bulky groups separated by a less bulky chain. The other two are quite different. Each fits snugly into the site despite their structural differences. Molecules **2** and **3** fit into the site with a bulky group in each of the pockets at either end of the site and the connecting chain fitting over the ridge. This is reminiscent of the way the phenylhippurate substrate is thought to bind. In Figures 3 and 4, molecule **3** is shown in the papain binding site. Molecules **1** and **4** fit differently. Molecule **1** fits with the pyridinyl group

Table I. Scores and Energies of Candidates Bound to Papain

candidate	score[a]	energy[b] of free candidate	energy[b] of candidate/protein complex	interaction energy[b,c]
molecule 1	175.40	5.62	-696.00	-35.32
molecule 2	166.35	15.05	-678.17	-26.92
molecule 3	164.64	36.41	-663.78	-33.89
molecule 4	160.47	20.56	-674.00	-28.26
molecule 5	147.65	6.17	-685.31	-25.18

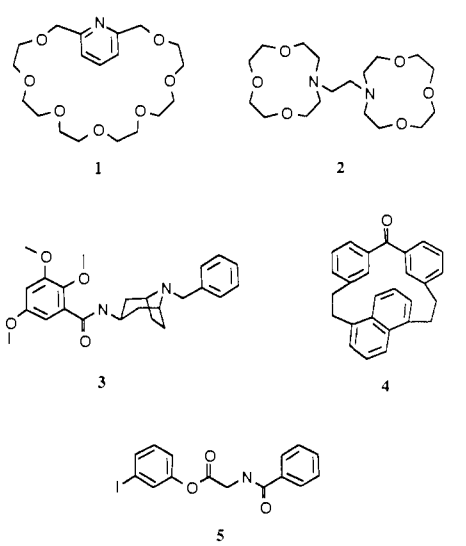
[a] Score reported here is from our simple scoring routine. The larger scores are more favorable. [b] All energies are reported in kilocalories after 500 cycles of AMBER[16,17] minimization with partial atomic charges set to zero on atoms of the candidates. [c] Interaction energy = (energy of the complex) − (energy of the free candidate + energy of the free protein). The energy of the free protein is −666.30 kcal.

Table II. Scores and Energies of Candidates Bound to Carbonic Anhydrase

candidate	score[a]	energy[b] of free candidate	energy[b] of candidate/protein complex	interaction energy[b,c]
molecule 6	178.38	6.73	-1130.73	-12.14
molecule 7	166.92	5.23	-1131.76	-11.67
molecule 8	148.65	3.13	-1156.12	-33.93
molecule 9	148.21	6.74	-1129.23	-10.65
molecule 10	70.32	NA	NA	NA

[a] Score reported here is from our simple scoring routine. The larger scores are more favorable. [b] All energies are reported in kilocalories after 500 cycles of AMBER[16,17] minimization with partial atomic charges set to zero on atoms of the candidates. [c] Interaction energy = (energy of the complex) − (energy of the free candidate + energy of the free protein). The energy of the free protein is −1125.32 kcal.

PAPAIN CANDIDATES



1 2,6-pyrido-24-crown-8[20]
2 10,10'-ethylene-bis(1,4,7-trioxa-10-azacyclododecane)[21]
3 N-(8-benzyl-1αH,5αH-nortropan-3β-yl)-2,3,5-trimethoxy-benzamide[22]
4 [2.2](4,4')benzophenono(2,6)naphthalenophane[23]
5 (3-iodo-phenyl) hippurate

(16) Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287.
(17) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(18) Drenth, J.; Jansonius, J. N.; Keokeok, R.; Swen, H. M.; Wolthers, B. G. *Nature (London)* **1968**, *218*, 929.
(19) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.

against the wall of the ridge. The ether chain bends across the ridge and into the pockets of either side. Molecule **4** fits into one of the pockets and near the ridge area, but it does not cross over the ridge into the other pocket. Even in this small sample of molecules, one sees the program's ability to locate a variety of molecular shapes and a variety of ways to fit molecules into a given site.

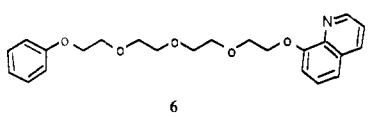
Carbonic Anhydrase. Carbonic anhydrase catalyzes the conversion of bicarbonate and hydrogen ions to carbon dioxide and water. The enzyme has a large, deep active site. The substrate binds to a zinc atom located in the bottom of the site. We used the X-ray crystallographic structure for human carbonic anhydrase C bound in the Brookhaven Protein Data Bank[19] set 1CAC.[2]

The large size of carbonic anhydrase site presented a problem because many proposed orientations were found at the lip of the site far from the zinc. We were looking for candidates from which we could design inhibitors, and so we needed to find orientations that would block the binding of bicarbonate ion. Therefore, we accepted only those orientations where at least one atom of the candidate was within 5.0 Å of the zinc atom. We used a smaller cutoff (2.0 Å as compared to the usual 2.3 Å) for the atomic overlap penalty. This allowed for more orientations that fit deeply into the site.

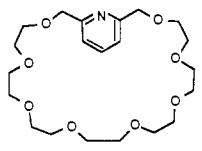
Four of the candidates found in our database search are shown below.

The known inhibitors of carbonic anhydrase are arenesulfonamides (**10**). We did not find any arenesulfonamides in our search because (1) they are small molecules and do not fill the site as well as the larger structures found here, and (2) they bind mainly on the basis of electrostatic interactions, which are ignored in the shape calculation. The scores and energies for four high-scoring molecules are listed in Table II. Once again the molecules that fit well are quite different from the known inhibitors and from each other. They also yield attractive van der Waals energies of interaction. The orientations of molecules **6**, **7**, **9**, and **10** are illustrated in Figure 5. Molecules **7**–**9** are

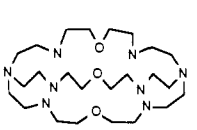
CARBONIC ANHYDRASE CANDIDATES



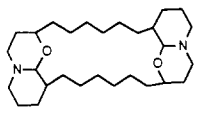
6



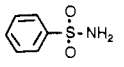
7



8



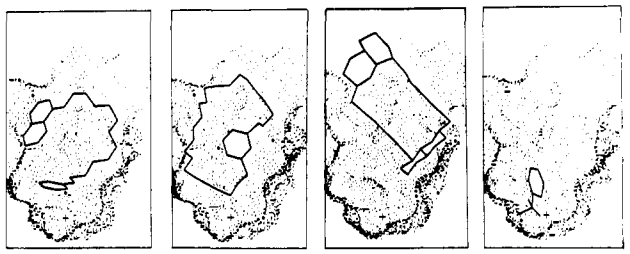
9



10

6 1-phenoxy-11-(8-quinolyoxy)-3,6,9-trioxaundecane[24]
7 2,6-pyrido-27-crown-9[20]
8 7,19,30-trioxa-1,13-diaza-4,10,16,22,27,33-hexa-azoniabicyclo (11.11.11)pentacontane[25]
9 xestospogin C[26]
10 arylsulfonamide

macrocycles, and molecule **6** bends around into a cyclic-like conformation. Molecule **6** fits most deeply into the site. Even molecules of this size fill only about one-half of the site and must choose one side of the site with which to



A B C D

Figure 5. The orientations of three of the carbonic anhydrase candidate ligands found by our program and the known inhibitor, arenesulfonamide, are shown in the carbonic anhydrase binding site. The dots represent the molecular surface of the carbonic anhydrase site. The position of the zinc atom is indicated with a plus sign (+). The molecules illustrated are (A) molecule **6**, (B) molecule **7**, (C) molecule **9**, and (D) molecule **10**. An adequate two-dimensional projection of molecule **8** could not be found.

interact. Molecule **8** has a better interaction energy than the other molecules, because, being cylindrical as opposed to being relatively flat, it can interact with a larger amount of the receptor surface. In Figures 6 and 7, molecule **8** is shown as it is oriented in the site after AMBER[16,17] minimization.

Discussion

There are many different approaches to any selecting molecules for drug design. We have chosen to approach the design of lead compounds by starting from the general perspective of shape complementarity. We do not expect that the molecules found in the shape search will necessarily be leads themselves, since no evaluation of chemical interactions is made. We see the molecules found by shape search as frameworks: molecular skeletons to which appropriate atomic replacements must be made. Of course, the chemical complementarity of these molecules can be evaluated, but we expect that atom types will be changed to maximize the electrostatic, hydrogen bonding, and hydrophobic interactions with the receptor.

At this preliminary state, we can say that our shape analysis method is successful in finding a variety of molecules that are complementary in shape to a given site. It has several attractive features. First, it retrieves a remarkable diversity of molecular architectures. Second, the best structures show impressive shape complementarity over an extended surface area. Third, the overall approach appears to be quite robust with respect to small uncertainties in positioning of the candidate atoms.

We are intrigued that the molecules identified here are larger than most known ligands and interact with more of the receptor site than just the catalytic residues. Interacting with a larger part of the site should increase the specificity of these molecules compared to molecules that are designed on mechanism-based ideas. We speculate that a sensible combination of mechanism-based and steric-based design efforts could lead to a tightly binding ligands with a high level of specificity.

We recognize that this approach is at a very early stage and its limitations must be kept in mind. In the case of the receptor, it is necessary to know its three-dimensional structure, or the structure of a molecule that can be used as a model. Further, the candidate ligands and the receptor site are held rigid during the docking procedure. Thus, the candidates themselves are only considered in the conformation found in the crystal. These limitations can be ameliorated in several ways. First, the use of AMBER[16,17] permits some relaxation of the rigid candidate and receptor geometry. Second, if distinct conformations of a candidate are desired, they can be put into the database separately.

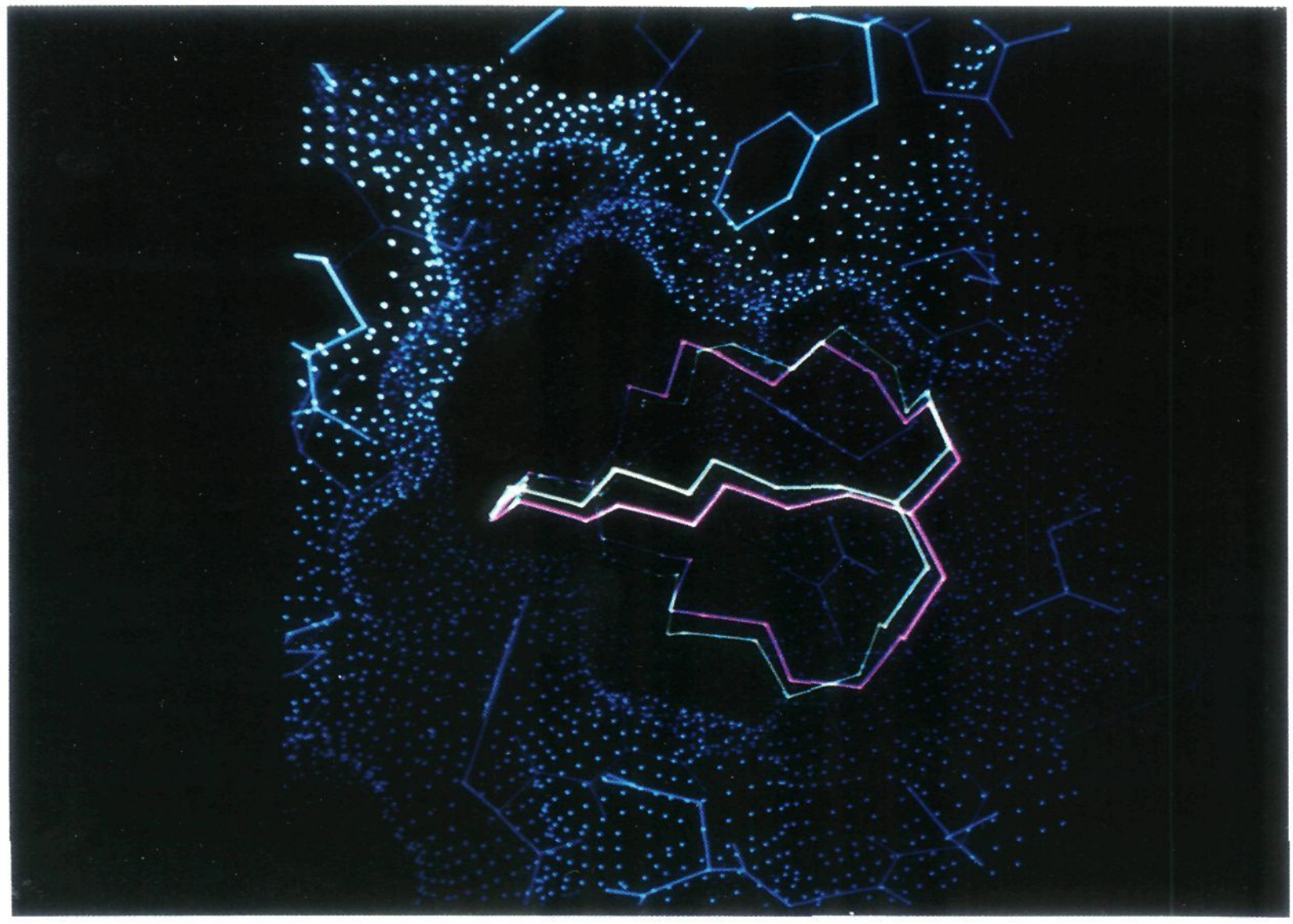


Figure 6. Molecule 8 is shown docked in the active site of carbonic anhydrase. Two orientations are shown: the orientation that is found by the search procedure (green), and the orientation after 500 cycles of AMBER[16,17] minimization (purple). The protein and its molecular surface before minimization are shown. The structure is not altered much during the minimization.

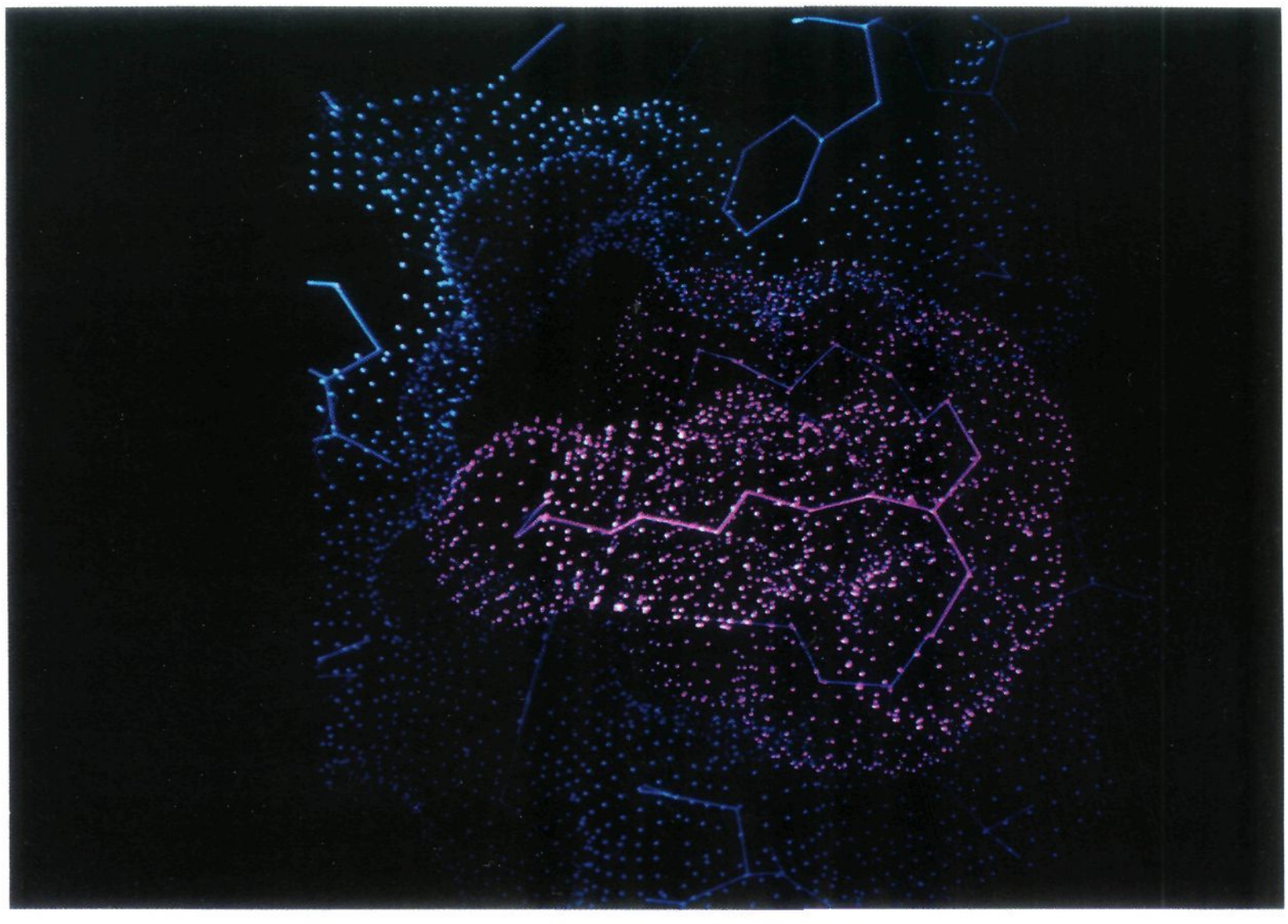


Figure 7. Molecule 8 docked in the carbonic anhydrase active site. Carbonic anhydrase is in blue and its molecular surface is displayed. Molecule 8 is in purple and is shown with its van der Waals surface. Molecule 8 is shown after 500 cycles of AMBER[16,17] minimization.

Third, we can deal with a limited number of rotatable bonds using the method developed by DesJarlais et al.[14] Alternative approaches may use distance geometry methods[27] for exploring the candidate conformation and a robust optimization method such as the ellipsoid algorithm may be helpful.[28] If one is interested in designing inhibitors or antagonists, keeping the protein site rigid during the docking is not a great limitation. As long as a ligand binds tightly enough to *some* conformation of the protein, it will prevent substrate binding, and it should have an inhibitory effect. The problem of designing substrates or agonists is clearly more difficult. It does require knowledge of the active conformation of the receptor.

A second limitation is the bias inherent in the set of molecules in the shape database. For example, there are a large number of macrocyclic and polycyclic hydrocarbons in the set of molecules used in the work reported. A number of these have similar shapes. It would be useful to choose the molecules included in the database carefully to provide the greatest variety in shape possible, but with attention to molecules that are synthetically accessible. We are currently working on this aspect of the problem.

We are aware that most targets of rational drug design are not receptors of known structure. We can extend the shape search method to such problems by using methods of pharmacophoric alignment.[29] In order to characterize an unknown binding site, one may start with a set of known ligands, each of which can be assigned on active conformation. If these can be superimposed, their combined volume approximates the shape of the binding site. We can then generate a receptor site mode complementary in shape to the ligands. Thus, only the algorithm for characterizing the shape of receptor site need to be different from that described in this paper. We have had some success with this alternate application, which will be reported elsewhere.

The success of this method will depend strongly on how well we are able to take the next step and map receptor properties onto the candidate atom positions. We are currently at work on a semiautomatic method of suggesting chemically reasonable molecules. It is likely that certain molecular frameworks that are found in the shape search may provide atom positions that allow for better interaction with the receptor than others. Synthetic accessibility must be considered at this point as well.

Despite its limitations, we expect that the shape-fitting approach will be a useful way of obtaining a wide variety of structures that fit a given receptor binding site, and may be used for more detailed calculational and experimental tests. An automatic search of this kind is likely to be more thorough than visual examination. Through such a search, the user is presented with a variety of structure that may be quite different from known inhibitors and substrates. The structures that are found may also be modified so that their chemical nature is more compatible with the site. The results of these efforts should give insight into the general problem of predicting receptor–ligand interactions.

The programs described in this paper are available by contacting IDK; the Cambridge Crystallographic Database is available through the Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, England or through William Duax, Medical Foundation of Buffalo, Research Laboratories, 73 High Street, Buffalo, NY 14203; the Protein Data Bank is available through Thomas Koetzle, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973.

Acknowledgment. We thank the Computer Graphics Laboratory (supported by National Institutes of Health RR-01081) for use of their facilities for parts of this project. Partial support was received from the National Institutes of Health GM-31497 (IDK), GM-07175 (RLD), and GM-29072 to Peter Kollman. The Floating Point Systems 264 array processor was purchased with grant support from National Science Foundation DMB-84-13762 and National Institutes of Health RR-02441. Assistance from the Macromolecular Workbench supported by the Defense Advanced Research Projects Agency under contract N00014-86-K0757 administered by the Office of Naval Research is also appreciated.

Registry No. Papain, 9001-73-4; carbonic anhydrase, 9001-03-0.

(20) van Staveren, C. J.; den Hartog, H. J., Jr.; Reinhoudt, D. N.; Uiterwijk, J. W. H. M.; Kruise, L.; Harkema, S. *J. Chem. Soc., Chem. Commun.* **1984**, 1409.
(21) Groth, P. *Acta Chem. Scand., Ser. A* **1984**, *A38*, 342.
(22) Durant, F.; Renard, P.; Evrard, G. *Bull. Soc. Chim. Belg.* **1984**, *93*, 419.
(23) Givens, R. S.; Venkatramanan, M. K.; Figard, J. *Tetrahedron Lett.* **1984**, *25*, 2187.
(24) Hirayama, F.; Zabel, V.; Saenger, W.; Vogtle, F. *Acta Crystallogr., Sect. C.: Crystal Struct. Commun.* **1985**, *C41*, 61.
(25) Dietrich, B.; Guilhem, J.; Lehn, J.-M.; Pascard, C.; Sonveaux, E. *Helv. Chim. Acta* **1984**, *67*, 91.
(26) Nakagawa, M.; Endo, M.; Tanaka, N.; Gen-Pei, L. *Tetrahedron Lett.* **1984**, *25*, 3227.
(27) Crippen, G. M. *Distance Geometry and Conformational Calculations*; Research Studies Press, Wiley: New York, 1981.
(28) Billeter, M.; Havel, T.; Kuntz, I. D. *Biopolymers* **1987**, *26*, 777.
(29) Marshall, G. R. "Computer Graphics and Receptor Modelling", in *Quantitative Approach to Drug Design*; Dearden, J. R., Ed.; Elsevier: Amsterdam, 1984.