Journal of Computer-Aided Molecular Design, 6 (1992) 223–233
ESCOM

J-CAMD 158

An approach to computer-aided inhibitor design: Application to cathepsin L

Sucha Sudarsanam, G. Duke Virca, Carl J. March and Subhashini Srinivasan*

Department of Protein Chemistry, Immunex Corporation, 51 University Street, Seattle, WA 98101, U.S.A.

Received 3 July 1991
Accepted 23 January 1992

Key words: Brevotoxin; Comparative modeling; Enzyme-inhibitor complex; Papain; Cysteine-protease

SUMMARY

We have developed an approach to search for molecules that can be used as lead compounds in designing an inhibitor for a given proteolytic enzyme when the 3D structure of a homologous protein is known. This approach is based on taking the cast of the binding pocket of the protease and comparing its dimensions with that of the dimensions of small molecules. Herein the 3D structure of papain is used to model cathepsin L using the comparative modeling technique. The cast of the binding pocket is computed using the crystal structure of papain because the structures of papain and the model of cathepsin L are found to be similar at the binding site. The dimensions of the cast of the binding site of papain are used to screen for molecules from the Cambridge Structural Database (CSD) of small molecules. Twenty molecules out of the 80 000 small molecules in the CSD are found to have dimensions that are accommodated by the papain binding pocket. Visual comparison of the shapes of the cast and the 20 screened molecules resulted in identifying brevotoxin b, a toxin isolated from the 'red tide' dinoflagellate *Ptycho brevis* (previously classified as *Gymonodium breve*), as the structure that best fits the binding pocket of papain. We tested the proteolytic activity of papain and cathepsin L in the presence of brevotoxin b and found inhibition of papain and cathepsin L with K_is of 25 μM and 0.6 μM, respectively. We also compare our method with a more elaborate method in the literature, by presenting our results on the computer search for inhibitors of the HIV-1 protease.

INTRODUCTION

Inhibiting the activity of a proteolytic enzyme would require a molecule to interact with the enzyme's active site in a manner similar to the lock-and-key mechanism. Interaction between two molecules, such as an enzyme and its inhibitor, is extremely complex. However, the requirements to achieve such interactions can be divided in two broad classes, steric and electronic. The steric part of the interaction has to do with the complementarity of the shapes of the interacting molecules near the active site. Based on this criterion, the design of an inhibitor can be accomplished

* To whom correspondence should be addressed.

0920-654X/$ 5.00 © 1992 ESCOM Science Publishers B.V.

in two steps. The first step would involve finding a molecule A which has a shape similar to the shape of the binding pocket of the target molecule. Conformation of the binding of molecule A to the target enzyme would suggest this molecule as a lead compound. Analogues of A with increased potency can then be made by modifying the electronic structure in a manner that complements the electronic structure of the enzyme binding pocket.

Although ligand and binding pocket are of complex geometrical shape, Kuntz et al. [1] have suggested a novel approach to define the shapes of the binding pockets on the surface of proteins of known structure. Using this approach it has been shown that haloperidol, a commercially available compound, has a K_i of ~ 100 μM against HIV-protease [2]. While the crystal structure of HIV-protease has been solved with and without substrate, it is of interest to know whether such precise structural information is required to simply define the binding pocket shape. This is of major concern because obtaining the 3D structure of any protein using crystallographic techniques is a laborious and time-consuming process. On the other hand, the comparative modeling technique proposed by Greer [3], offers a reasonable starting point by utilizing structures of homologous proteins in the absence of precise structural information about the molecule of interest.

There are at least 5 cysteine protease families. Papain is the prototype of the papain family of cysteine proteases. Sequence homology between papain and other members of this family, e.g. cathepsins L, B and H, chymopapain, bromelain, ficin and asclepains strongly suggests that this family of proteins shares a similar tertiary structure. Cathepsin L is a lysosomal enzyme present in a wide variety of cells and thought to play an important role in normal intracellular protein turnover [4,5], bone remodeling [6], and prohormone activation [7]. Cathepsins have been implicated in a number of disease processes such as cancer metastasis [8–10], arthritis [11], periodontal disease [12,13], muscular dystrophy [14,15], emphysema [16,17] and glomerulonephritis [18]. Cathepsin L, if inappropriately regulated, is potentially the most damaging of the intracellular cathepsins as it has been shown to degrade a number of structural tissue components such as collagen [19], elastin [20] and basement membrane [21,22]. Because of its potential for causing tissue destruction we sought to target this protease for inhibition. Its membership in the papain superfamily enabled us to use the known structure of papain [23] as the starting point for modeling the cathepsin L substrate binding pocket. In this paper we describe a simple modeling approach towards the design of protease inhibitors and demonstrate the approach in finding lead compounds for cathepsin L based on crystallographic structural information available for the binding pocket of papain.

COMPUTATIONAL METHODS

The availability of the crystal structure of papain is the key to this approach. Given the structure of this enzyme, the design of an inhibitor consists of 3 main parts: (i) computation of the shape of the target site on the surface of the protein, i.e. its binding pocket, (ii) searching for organic molecules that would fit the shape of the target site and (iii) testing the inhibition of the targeted enzyme with the selected organic molecules.

A 3D model* for cathepsin L was built using the homology modeling program from Biosym

*The coordinates for this model have been deposited with the Journal of Computer-Aided Molecular Design, and are available on request.

Technologies (San Diego, CA). The crystal structure of papain and the resulting cathepsin L model have similar structures at the binding sites. Since binding pockets are not rigid and can accommodate small changes in shape, we used the binding pocket of papain to search for lead molecules for cathepsin L. It should be noted that the electrostatics of the binding pockets of papain and cathepsin L are very different and will be used in improving the specificity of lead molecules directed towards the two enzymes. We discuss the various steps involved in our algorithm below.

(A) Representation of an arbitrary three-dimensional shape

In general, molecular shapes are complex geometrical forms. There is no precise way of describing molecular shapes mathematically. For the purpose of computation and visualization, each atom in a molecule is often represented by a sphere, the radius of such a sphere being equal to the atom's van der Waals radius. Thus, the molecule itself is represented as a collection of spheres. In our problem, we are concerned with the cast of a pocket on the surface of a protein. A detailed approach to describe possible shapes of the pocket exists in the literature [1]. We have developed a simple but powerful algorithm to describe arbitrary molecular shapes and use this information in analyzing the binding sites. Our approach assumes that the cast of a pocket in the protein and that of the potential binding molecule can be approximated to that of ellipsoids (see Appendix). This approximation preserves information about the dimensions of the pocket as well as the orientation of the pocket. The ellipsoid thus defined fits the shape of the cast or the molecule in a least-squares sense.

To define the shapes of the potential inhibitor lead compounds we compute an ellipsoid for every molecule in the CSD [24,25] and the information is stored in an 'ellipsoid database'. Note that this needs to be done only once. We then compute an ellipsoid for the target cast of the binding pocket on the surface of the protein. Thus, searching for a potential binding molecule is reduced to searching the ellipsoid database for given cast dimensions.

(B) Computation of protein pocket shape

Pockets on the surface of a protein are defined as sites accessible for a probe sphere of radius greater than 3 Å but not accessible for a probe sphere greater than 4 Å. Locations of the center of the probe sphere at such sites constitute the collection of pockets on the entire surface of the protein. This collection consists of many constellations, of which usually the biggest one is the binding pocket. These constellations are visually examined to isolate the binding pocket. Figure 1 shows the binding pocket of papain computed by the above method and Fig. 2 shows the binding pocket of HIV-1 protease.

(C) Searching the ellipsoid database

The protein molecule is examined for binding sites according to the procedure described in step B. Once the binding site has been located, an ellipsoid representing the binding pocket is computed. The ellipsoid database is then searched for small molecules with dimensions similar to the dimensions of the cast within predefined limits. The limits described here result in the selection of a small number (<200) of molecules out of the database (CSD) of 80 000 molecules. This more manageable number of molecules can now be examined visually. Figure 3 shows some sample compounds screened from the CSD for the papain binding pocket.

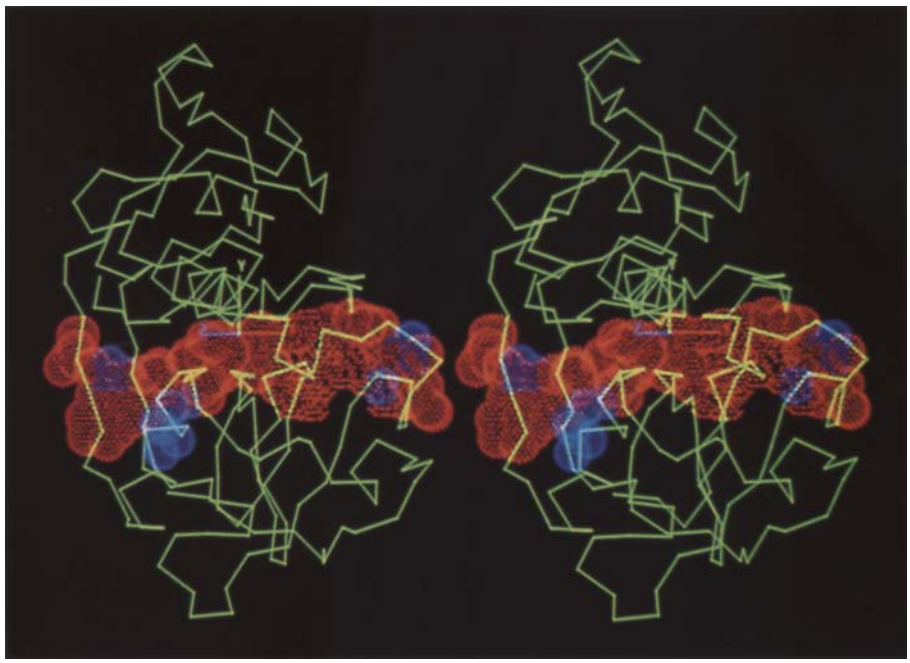

Fig. 1. Stereoview of the binding pocket (red and blue) and C-α trace of papain (green).

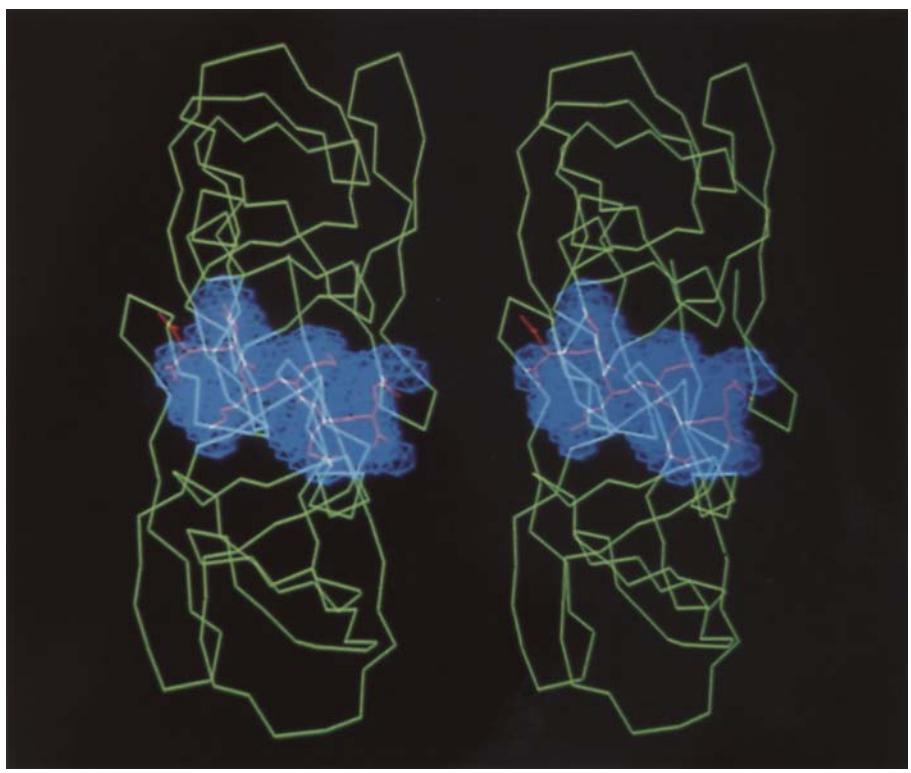

Fig. 2. Stereoview of the binding pocket (blue) and C-α trace (green) of HIV-1 protease and substrate (pink) as bound :he crystal structure.

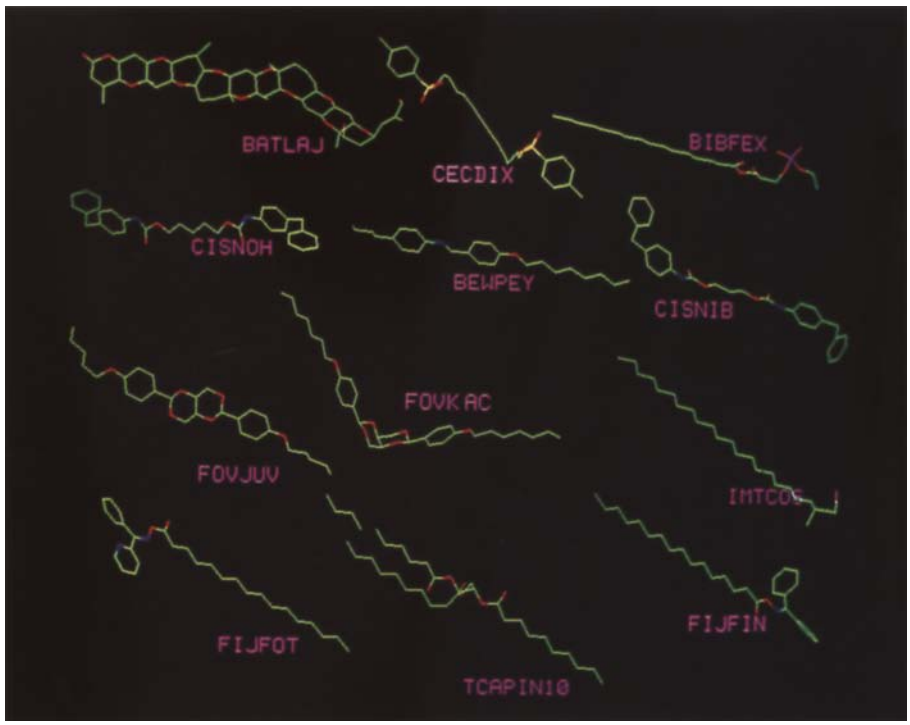

Fig. 3. Samples of small molecules searched from the Cambridge Structural Database for the binding pocket of papain.

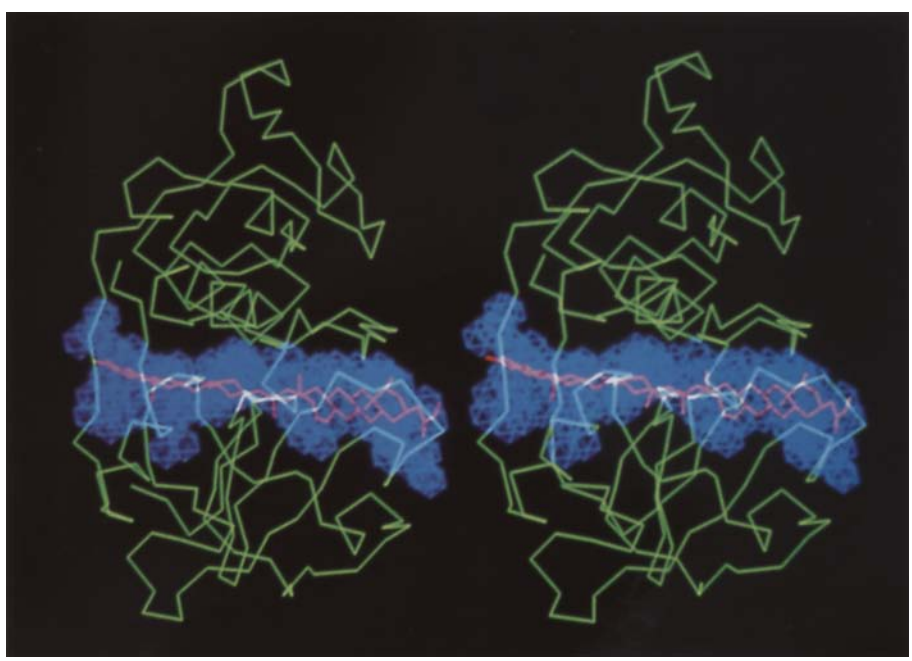

Fig. 4. Stereoview of brevotoxin b (pink), the binding pocket (blue) and C-α trace (green) of papain.

(D) Docking of selected molecules to the binding pocket

All the molecules selected (step C) are displayed along with the binding pocket on the graphics screen. These molecules are then docked visually in the binding pocket of the protein. Based on this visual examination, a final list of molecules is generated for further modification and testing. We used this approach to search for an inhibitor for both HIV-protease and cathepsin L.

RESULTS

(A) HIV-protease

The semi-axes of the least-squares ellipsoid representing the cast of the binding pocket are: a = 5.1 Å, b = 2.7 Å and c = 1.5 Å and those of the substrate as complexed with the protease are: a = 6.0 Å, b = 2.4 Å and c = 1.2 Å.

The ellipsoid database was searched for molecules which satisfy the following criteria:

$$4.5 \text{ Å} \leq a \leq 6.0 \text{ Å}, b \leq 2.7 \text{ Å and } c \leq 1.5 \text{ Å}.$$

The upper and lower bounds for a, b and c are chosen to allow for flexibility in the binding pocket. The lower bounds for b and c are ignored as they fall very close to atomic dimensions. This search resulted in about 150 molecules. This list includes bromperidol (CSD code: BIBSEK) and haloperidol (CSD code: HALOPB), which are present in the hit list of 200 molecules found by using a more elaborate method in the literature [2]. The semi-axes of the ellipsoid for BIBSEK are 4.981 Å, 1.207 Å and 0.949 Å. The semi-axes for HALOPB are 5.610 Å, 1.036 Å and 0.711 Å. The dimensions for both BIBSEK and HALOPB fall within the bounds of the search. Haloperidol has been shown to have a K_i of ~ 100 μM to HIV-protease [3].

(B) Cathepsin L

The semi-axes of the least-squares ellipsoid representing the cast of the binding pocket are: a = 9.4 Å, b = 2.3 Å and c = 2.2 Å. The ellipsoid database was searched for molecules which satisfy the following criteria:

$$8.0 \leq a < 10.0, b \leq 3.0 \text{ Å and } c \leq 3.0 \text{ Å}.$$

Again, the upper and lower bounds for a, b and c are chosen to allow for flexibility in the binding pocket. The lower bounds for b and c are ignored as they fall very close to atomic dimensions. Visual docking of the molecules to the binding pocket resulted in selecting brevotoxin b [26] (CSD code: BATLAJ) as the molecule that best fits into the pocket. The semi-axes of the ellipsoid for BATLAJ are 8.540 Å, 1.889 Å and 1.266 Å. Figure 4 shows brevotoxin b in the binding pocket of papain. Figure 5 compares the binding pocket and brevotoxin b in 3 orthogonal views. We also give a sampling of the organic molecules from this search (see Table 1).

INHIBITION OF CATHEPSIN L AND PAPAIN BY BREVOTOXIN B

Methods and materials

Human procathepsin L was expressed in *S. cerevisiae* using the yeast expression vector pIXY-

406 containing a human procathepsin L cDNA isolated from a human peripheral blood monocyte cDNA library constructed in the bacterial plasmid vector pDC302 [27,28]. Procathepsin L was acid activated by preincubation in activation buffer (120 mM sodium acetate, 20 mM acetic acid, 1 mM EDTA, 2.5 mM DTT, pH 5.5) for 2 h at 37°C. Papain (2× crystallized) was purchased from Sigma (St. Louis, MO 63178). Cathepsin L and papain activities were measured with the colorimetric substrates Z-Phe-Arg-pNA and Ac-Phe-Gly-pNA, respectively (Bachem Bioscience, Philadelphia, PA 19104). All assays were performed in the pH 5.5 activation buffer (see above). Brevotoxin b (PbTx-2) was purchased from Chiral Corp., Miami, FL 33146. A 1-mM stock solution of brevotoxin b was made up in ethanol and diluted in activation buffer prior to assay.

Results

The apparent K_i, $K_{i(app)}$, was determined using a modification of the 'progress curve' method [29]. This approach measures the uninhibited rate of substrate hydrolysis (V_0) and compares it with the rate of substrate hydrolysis in the presence of inhibitor (V_i). The ratio of the rates of uninhibited to inhibited substrate turnover is used to determine the $K_{i(app)}$ from the following relationship.

$$V_0/V_i = 1 + [I] / K_{i(app)} \quad (1)$$

where [I] is the concentration of inhibitor and $K_{i(app)}$ the equilibrium inhibition constant in the presence of substrate. The true equilibrium constant K_i is then calculated from:

$$K_i = K_{i(app)} / 1 + [S] / K_m \quad (2)$$

TABLE 1
MOLECULES OBTAINED FROM SEARCHING THE ELLIPSOID DATABASE FOR CATHEPSIN L BINDING POCKET (CSD CODES)

BATLAJ	brevotoxin b
BEWPEY	*N*-(4-*n*-octyloxybenzylidene)-4′-butylaniline
BIBFEX	3-palmitoyl-DL-glycerol-1-phosphorylethanolamine
CECDIX	1,12-bis(tosyloxy)-dodeca-5,7-diyne//Dodeca-5,7-diyne-1,12-diyl-1-bis-*p*-toluenesulfonate
CERMER	8,16-bis(3-(4-(4,5-dihydroxy-6-methyl-2-tetrahydropyranyloxy)-5-ethyl-2-hydroxy-6-methyl-2-tetrahydropyranyl)-2-hydroxy-1-methylbutyl)-7,15-dimethyl-1,9-dioxa-cyclohexadeca-3,5,11,13-tetraene-2,10-dione dihydrate
CISNIB	1,3-propanediyl-bis(4-(phenylmethyl)-phenylcarbamate)
CISNOH	1,5-pentanediyl-bis(4-(phenylmethyl)-phenylcarbamate)
FIJFIN	*O*-palmitoylbenzophenone oxime
FIJFOT	(E)-*O*-palmitoyl phenyl 2-pyridyl ketone oxime
FOVJUV	2e,6e-bis(4-pentyloxyphenyl)-trans-1,3,5,7-tetraoxadecalin
FOVKAC	2e,6e-bis(4-heptyloxyphenyl)-cis-1,3,5,7-tetraoxadecalin
GAHMAD	4-hydroxyethyldimethylammonio-pentyloxybenzene-1-azo-1′-4′-hexyl oxybenzene bromide
GLPHYC10	glucosyl-phytosphingosine hydrochloride monohydrate
IMTCOS	(-)-1-iodo-3-methyltricosane
PRSTER	*n*-propyl stearate//Propyl octadecanoate
STRPAB	collagen proline hydroxylase inhibitor P-1894B toluene solvate
TCAPIN10	tricaprin

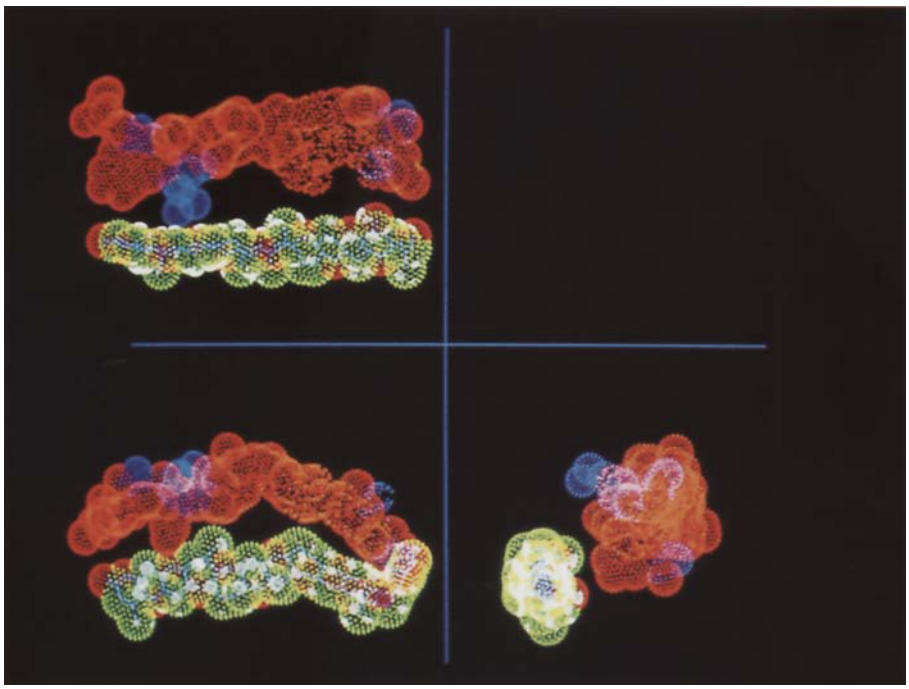

Fig. 5. Orthogonal projection of the binding pocket of papain (red and blue) and brevotoxin b (color coded by atoms) shown side by side.

The K_m values for cathepsin L with Z-Phe-Arg-pNA and papain with Ac-Phe-Gly-pNA were determined graphically by standard methods [30–32].

Figure 6 shows the results of cathepsin L assays performed in the absence (V_0) or presence (V_i) of 3 different concentrations of brevotoxin b. From these rate assays V_0/V_i ratios were determined and from these the following $K_{i(app)}$ values were calculated: 23 μM (at a brevotoxin b concentration of 20 μM), 17 μM (brevotoxin b at 10 μM) and 25 μM (brevotoxin b at 5 μM). In order to calculate the true K_i we determined the K_m for Z-Phe-Arg-pNA with cathepsin L to be 28 μM (data not shown). Using these values for $K_{i(app)}$ and K_m in Eq. 2, the mean K_i for the inhibition of cathepsin L by brevotoxin b is 0.6 μM. Data generated in similar experiments performed with papain and brevotoxin b resulted in a mean K_i of 24 μM (data not shown).

DISCUSSION

Knowledge of the 3D structure of the enzyme is necessary to arrive at the shape of the binding pocket. However, here we emphasize that the approximate shape of the binding pocket from X-ray crystallographic studies of homologous proteins could very well be used in this approach for inhibitor design. This approximation is valid considering the multiple substrate specificities of most enzymes. In fact, the search for a tight-fitting ligand could limit the repertoire of chemistry of potential lead compounds. Generally the binding pocket is computed from a crystal structure that has the substrate bound in the pocket. The search for a tight-fitting ligand to this pocket

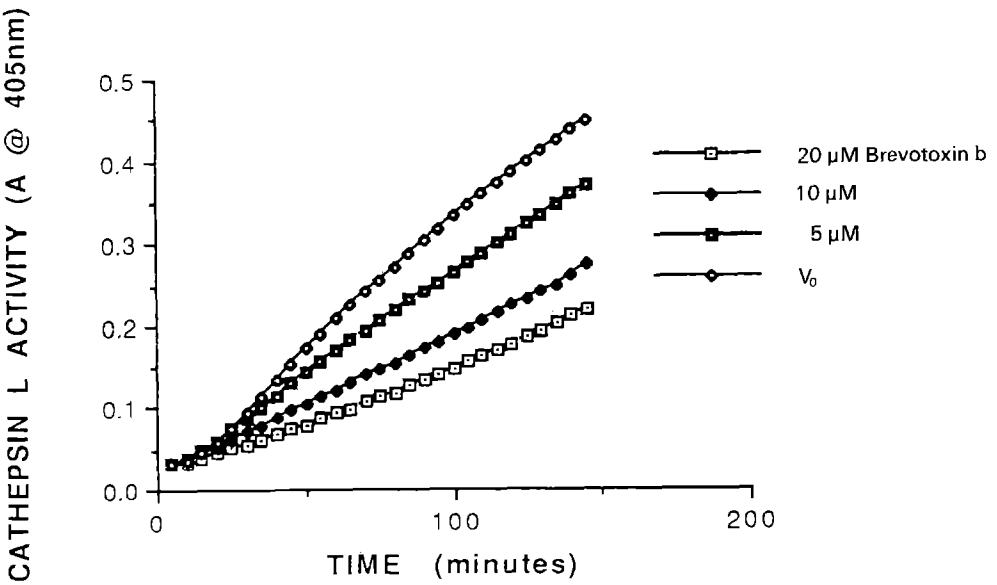

Fig. 6. Inhibition of cathepsin L by brevotoxin b.

would be too restrictive. The bounds on the semi-axes in our search allow for some flexibility in the binding pocket. The approach given in this work thus can find ligands that can fill the entire binding pocket, increasing the contact area between enzyme and ligand. Hydrophobic interactions of greater contact region overcome entropic considerations and potential electrostatic interactions with water. The increased contact area provides more sites on the ligand for possible modification to enhance specificity. The approach taken by Kuntz et al. [1], although more accurate, does not make an attempt to fill the binding pocket of the enzyme with the ligand. The reasonable inhibition of papain and cathepsin L by brevotoxin b, taking into account the coarseness in our approach, may result from the ligand filling of the entire binding pocket.

The method outlined in this work, although tested for proteolytic enzymes could very well be used in the search of inhibitors for other enzymes such as kinases, RNases, etc. The 3D structures of other enzymes are not as well studied as those of proteolytic enzymes. However, the limited number of crystal structures reported for kinases and RNase reveal deep binding pockets near the active sites, thus potentially broadening the scope of this method in the design of inhibitors for other enzymes. For example, the recently reported crystal structure of HIV-RNase has a deep binding pocket near the active site which could be exploited to design a therapeutically important inhibitor.

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., Seibel, G.L., Kuntz, I.D., Furth, P.S., Alvarez, J.C., Ortiz de Montellano, P.R., DeCamp, D.L., Babe, L.M. and Craik, C.S., Proc. Natl. Acad. Sci. USA, 87 (1990) 6644.
3 Greer, J., Proteins: Structure, Function and Genetics, 7 (1990) 317.

4 Ritonja, A., Popovic, T., Kotnic, M., Machleidt, W. and Turk, V., FEBS Lett., 228 (1984) 341.
5 Bohley, P., Kirschke, H., Langer, J., Miehe, M., Riemann, S., Slama, Z., Schon, E., Weideranders, B. and Ansorge, S., In: Holzer, H. and Tschesche, H. (Eds.) Biological Functions of Proteins, Springer-Verlag, Berlin, 1979, p. 17.
6 Delaisse, J.M., Ledent, P., Eeckhout, Y. and Vaes, G., In: Turk, V. (Ed.) Cysteine Proteinases and Their Inhibitors, Walter de Gruyter, Berlin, 1986, p. 259.
7 Marx, J.L., Science, 235 (1987) 285.
8 Denhardt, D.T., Greenberg, A.H., Egan, S.E., Hamilton, R.T. and Wright, J.A., Oncogene, 2 (1987) 55.
9 Sloane, B.F. and Honn, K.V., Cancer Metastasis Rev., 3 (1984) 249.
10 Rozhin, J., Wade, R.L., Honn, K.V. and Sloane, B.F., Biochem. Biophys. Res. Commun., 164 (1989) 556.
11 VanNoorden, C.J.F., Smith, R.E. and Rasnick, D., J. Rheumatol., 15 (1988) 1525.
12 Lah, T., Skaleric, U., Babnik, J. and Turk, V., J. Periodontal Res., 20 (1985) 458.
13 Lah, T., Skaleric, U., Babnik, J. and Turk, V., J. Periodontal Res., 21 (1986) 504.
14 Sano, M., Wada, Y., Ii, K., Katunama, N. and Tsukagoshi, H., Acta Neuropathol., 75 (1988) 217.
15 Kominami, E., Bando, Y., Ii, K., Hizawa, K. and Katunama, N., J. Biochem. (Tokyo), 96 (1984) 1841.
16 Mason, R.W., Johnson, D.A., Barrett, A.J. and Chapman, H.A., Biochem. J., 233 (1986) 925.
17 Johnson, D.A., Barrett, A.J. and Mason, R.W., J. Biol. Chem., 261 (1986) 14748.
18 Thomas, G.J. and Davies, M., Biochim. Biophys. Acta, 990 (1989) 246.
19 Kirschke, H., Kembhavi, A.A., Bohley, P. and Barrett, A.J., Biochem. J., 201 (1982) 367.
20 Tsuchida, K., Aihara, H., Isogai, K., Hanada, K. and Kato, H., Hoppe-Seyler's Z. Physiol. Chem., 367 (1981) 39.
21 Matsukura, U., Okitani, A., Nishimuro, T. and Kato, H., Biochim. Biophys. Acta, 662 (1981) 41.
22 Baricos, W.H., Zhou, Y., Mason, R.W. and Barrett, A.J., Biochem. J., 252 (1988) 301.
23 Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A. and Drenth, J., J. Mol. Biol., 179 (1984) 233.
24 Allen, F.H., Kennard, O., Motherwell, W.D.S., Town, W.G. and Watson, D.G., J. Chem. Doc., 13 (1973) 119.
25 Allen, F.H., Bellard, S., 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 Crytallogr. B, 35 (1979) 2331.
26 Lin, Y.Y., Risk, M., Ray, S.M., Van Engen, D., Clardy, J., Golik, J., James, J.C. and Nakanishi, K., J. Am. Chem. Soc., 103 (1981) 6773.
27 Mosley, B., Beckman, M.P., March, C.J., Idzerda, R.I., Gimpel, S.D., Vanden Bos, T., Friend, D., Alpert, A., Andersen, D., Jackson, J., Wignall, J.M., Smith, C., Gallis, B., Sims, J.E., Urdall, D., Widmer, M.B., Cosman, D. and Park, L.S., Cell, 59 (1989) 335.
28 Price, V., Mochizuki, D., March, C.J., Cosman, D., Deely, M.C., Klinke, R., Clevenger, W., Gillis, S., Baker, P. and Urdall, D., Gene, 55 (1987) 287.
29 Salvesean, G. and Nagese, H, In: Beynon, R.J. and Bond, J.S. (Eds.) Proteolytic Enzymes: A Practical Approach, IRL Press, Oxford, 1989, pp. 83–104.
30 Lineweaver, H. and Burke, D., J. Am. Chem. Soc., 56 (1934) 658.
31 Eadie, G.S., J. Biol. Chem., 146 (1942) 85.
32 Hofstee, B.N.J., Nature, 184 (1959) 1296.

APPENDIX

Given the coordinates of the binding pocket r_i (x, y, z), i = 1 ... n, the semi-axes of the least-squares ellipsoid representing the binding pocket can be obtained by diagonalizing the following symmetric matrix:

$$\begin{bmatrix} \langle u_x^2 \rangle & \langle u_x u_y \rangle & \langle u_x u_z \rangle \\ \langle u_y u_x \rangle & \langle u_y^2 \rangle & \langle u_y u_z \rangle \\ \langle u_z u_x \rangle & \langle u_z u_y \rangle & \langle u_z^2 \rangle \end{bmatrix}$$

where the diagonal elements are the second moments of x_i, y_i, z_i and are given by $\langle u_x^{2'} \rangle = \langle (x_i - < x_i >)^2 \rangle$ etc. The off-diagonal terms $\langle u_x u_y \rangle$, $\langle u_y u_z \rangle$ and $\langle u_z u_x \rangle$ represent the correlation between (x_i,y_i), (y_i,z_i) and (z_i, x_i). A typical off-diagonal element can be computed from

$$C_{xy} = \frac{\langle u_x u_y \rangle}{\sqrt{\langle u_x^2 \rangle \langle u_y^2 \rangle}}$$

where C_{xy} is the linear correlation coefficient and can be computed for a pair of vectors (x_i,y_i). The semimajor axes of the ellipsoid are the square roots of the eigenvalues of the above matrix. The corresponding eigenvectors orient the ellipsoid in its principal axis coordinate system.