An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins†

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY · FEBRUARY 1996

Impact Factor: 2.78 · DOI: 10.1021/jp952516o

An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins†

Karen N. Allen,‡ Cornelia R. Bellamacina, Xiaochun Ding, Constance J. Jeffery, Carla Mattos, Gregory A. Petsko, and Dagmar Ringe*

Rosenstiel Basic Medical Sciences Research Center, Departments of Chemistry and Biochemistry, Brandeis University, 415 South Street, Waltham, Massachusetts 02254-9110

Received: August 28, 1995; In Final Form: December 1, 1995⊗

Porcine pancreatic elastase has been used as the model enzyme in the design and development of a crystallographic method that allows mapping of the binding surface of a protein by solving its crystal structure in a variety of organic solvents. The ultimate goal of this method is to aid in the process of drug design, where each of the chosen organic molecules represents a given functional group in a larger inhibitor molecule. This method of multiple solvent crystal structures (MSCS) has a theoretical counterpart in the method of multiply copy simultaneous search (MCSS) (Miranker, A.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29−34) and is the first experimental method that can be used as a check to the theory. The MSCS method is presented here with acetonitrile as the probe organic solvent. The procedure involved does not cause significant changes in the structure of elastase as compared to the structure in aqueous solution, and the positions found for the acetonitrile molecules in the active site are compared to those of similar functional groups belonging to known inhibitors bound to elastase.

Introduction

The strategies used in drug design are divided into two approaches. The first involves trial and error, where drugs are discovered by accident or as a result of testing an immense number of compounds. The second approach, which has gained increasing importance, is the so-called rational or structure-based drug design. The latter approach is dependent on the availability of a crystal or NMR structure which reveals the binding pockets of the protein for the design of ligands that have a complementary surface. Most of the methods used to design such compounds rely on modeling techniques and on theoretical programs that probe the available binding pockets. Examples of these theoretical techniques are found in programs such as DOCK[1] which chooses compounds that have shape complementary to the binding pocket of interest, GRID[2] which explores the potential energy surface of a protein model and attempts to find regions of minimum potential, and MCSS[3] where thousands of copies of a given small molecule or functional group are minimized onto the surface of the protein, using an empirical potential energy function to generate a so-called functionality map. One of the major limitations of these methods is that the protein is generally taken to be completely rigid, and solvation is not taken into account. In the case of MCSS, for example, hundreds of energy minima are obtained for a given functional group,[4] and it is not trivial to decide which minima are relevant for drug design. It is very difficult to assess the relative affinities of different sites for water and the probe molecule.

An experimental method for probing the protein surface in a manner analogous to the MCSS theoretical procedure[3] is suggested by the protein structures solved in neat organic solvents. The goal of these experiments was to understand the structural factors involved in nonaqueous catalysis.[5] To this end, the structure of subtilisin Carlsburg was solved in acetonitrile,[6] and that of γ-chymotrypsin was solved in hexane.[7] In both cases it was established that no changes occur in the overall structure of the protein as it is transferred into the organic solvent, and the changes that do occur involve side chain reorientations. Much was learned from these two crystal structures about the possible reasons for enhanced thermostability and changes in substrate specificity of the enzymes in organic solvents.

In the present article, the results of solving the crystal structures of enzymes in organic solvents are focused on the development of a new drug design methodology: multiple solvent crystal structures (MSCS). This methodology takes advantage of the fact that organic solvents contain functional groups that are representative of those found in more complicated compounds. One type of solvent is used at a time, either neat or mixed with water, so that the competition of the probe molecule for the binding sites on the surface of the protein is only with water, rather than with molecules of a second solvent. It also allows a direct comparison of the binding sites of the probe molecule in the crystal structure with those obtained by the MCSS method. Once the MSCS technique is well established, one may want to consider mixtures of organic solvents to address the problem of competition for a given binding site.

The MSCS method is designed to probe the entire surface of the target protein and therefore allows the discovery of subsites on the enzyme that are not occupied by substrates or their analogs. With the MSCS method it is possible to go beyond what is learned by inhibitor studies which focus on the active site alone. Ligands can be designed which interact in the substrate binding sites, inhibiting the catalytic activity of the enzyme, and in new sites which are not occupied by substrates, conferring greater specificity for the target enzyme. Mechanism-based inactivators can react with other proteins which use the same chemical mechanism, and substrate analogs can bind to other proteins which catalyze different reactions with the same substrate. In either case the inhibitor will be less specific than

* To whom correspondence should be addressed.

† This research was supported in part by the Biophysics Program at Brandeis University, NIH Grant T32GMO7596 (C.R.B.), and the Lucille P. Markey Foundation, NIH Grant GM-26788, American Cancer Society Fellowship PF3560 (K.N.A.), Cystic Fibrosis Foundation Postdoctoral Fellowship F629 (C.J.J). C.M. is a Biomedical Research Fellow at the Bunting Institute of Radcliffe College, supported by the Burroughs Wellcome Fund.

‡ Current address: Structural Biology Group, Department of Physiology, Boston University School of Medicine, 80 East Concord Street, Boston, MA 02118-2394.

⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1996.

0022-3654/96/20100-2605$12.00/0

desired. Several examples of ligands that occupy binding sites on enzymes which are not substrate subsites have been observed.[8−11] Elastase was chosen as the model enzyme to develop the MSCS method. It is an interesting target for drug design, since it is a serine protease directly involved in diseases including emphysema,[12] arthritis,[13] and pancreatitis.[14] Furthermore, it is an enzyme which is commercially available and which is easy to crystallize into mechanically stable crystals that diffract to very high resolution in aqueous solvent (1.6 Å). It is an ideal enzyme for structure-based drug design because it does not undergo any conformational changes upon the binding of substrates or inhibitors. The degree of variability between elastase structures previously solved in aqueous solution can serve as a standard in monitoring significant changes which may occur due to the effect of the organic solvent in the MSCS experiments. The inhibitor and substrate bound structures also provide information about ligand binding to elastase which allows for retrospective analysis of the significance of the binding pockets for the probe organic molecules.

In this first article the MSCS methodology is presented with acetonitrile as the single probe molecule. The foundations of the methodology are established using the structure of elastase in acetonitrile compared to that of the native protein and to the cross-linked crystal structure in aqueous solution. As will be seen, the crystals must be stabilized by cross-linking before being transferred to the organic solvents. It is shown that neither the cross-linking nor transfer into acetonitrile significantly affects the overall structure of the protein. The positions of the acetonitrile molecules found in the active site are compared to the locations of the functional groups belonging to various inhibitors bound to the enzyme. A second paper is being prepared containing the application of this method to elastase.

Materials and Methods

Porcine pancreatic elastase was purchased from Worthington Biochemicals and used without further purification. Glutaraldehyde was purchased as an 8% solution in distilled water from Electron Microscopy Sciences. Since glutaraldehyde undergoes polymerization and other side reactions, the concentration of the glutaraldehyde was considered to be equal to that calculated by the company only after initially opening each vial. Thereafter, the glutaraldehyde was purified by allowing it to sit 12 h at 4 °C on activated charcoal and then filtered. The concentration was then determined from the absorbance at 280 nm and compared to the absorbance of the freshly opened vial. Note that contaminants of the glutaraldehyde which were removed by the activated charcoal are observed at 235 nm. All other chemicals were of analytical grade or of the highest quality commercially available.

Crystals of elastase were grown by the vapor diffusion sitting drop method, using published crystallization conditions[15] with minor adjustments. The crystals were grown in 18−24 μL drops containing a 1:1 mixture of protein solution and reservoir solution. Freezed-dried elastase was dissolved in 10 mM sodium acetate buffer, pH 5.0, to a concentration which varied from 9.0 to 14.0 mg/mL. The reservoir solution consisted of 10−15 μL of 20−50 mM sodium sulfate and 10 mM sodium acetate buffer, pH 5.0. Orthorhombic crystals with the symmetry of the space group $P2_12_12_1$ were obtained, with unit cell dimensions $a = 51.1$ Å, $b = 58.1$ Å, and $c = 75.3$ Å and angles $\alpha = \beta = \gamma = 90°$. All crystals tested were isomorphous with each other and with the crystals used to obtain the native elastase model from which phases were taken for solving the structures presented in this article.

Native Elastase. X-ray diffraction data for native elastase were collected at 4 °C to 1.6 Å resolution on an R-Axis II phosphoimaging plate mounted on a Rigaku RU200 rotating anode generator operating at 50 kV and 100 mA. Radiation was monochromatized with a graphite crystal. The data were reduced with the program Process.[16,17] The program XPLOR was used for initial phase calculations and positional least-squares refinement.[16,17] The initial phases were calculated from the coordinates of native elastase refined at 1.65 Å resolution[15] from which all solvent molecules were removed. After application of initial phases to the measured data from 10 to 1.6 Å and the first round of refinement, the R factor was 25.8%. A difference Fourier electron density map with coefficients ($2F_{obs} - F_{calc}$) showed the initial model of the protein fit the electron density well. An Evans & Sutherland PS300 graphics terminal running the program FRODO[18] was used for manual rebuilding of the protein and solvent. A difference Fourier electron density map with coefficients ($F_{obs} - F_{calc}$) was calculated and used to check the position of water molecules and ions at the 3 σ contour. The initial waters were obtained from a list of waters which occurred in two structures of elastase complexed to trifluoroacetyl−dipeptide−anilide inhibitors.[9] Additional water molecules were added during four more rounds of XPLOR positional refinement and manual rebuilding to the ($2F_{obs} - F_{calc}$) electron density map contoured at the 1σ level. The final R factor versus all data was 17.8% with good geometry for the model as indicated in Table 1.

Cross-Linked Elastase. All manipulations of the crystals were performed in a siliconized concave watch glass. Native elastase crystals were cross-linked with glutaraldehyde following a modification of the general method of Quiocho and Richards.[19−21] The crystals were transferred from the crystallization mother liquor to 200 μL of a high salt solution (100 mM sodium sulfate in 10 mM sodium acetate, pH 5.0). This buffer was then exchanged with a cross-linking buffer (100 mM sodium sulfate, 100 mM sodium phosphate, pH 7.5) by the addition and removal of 100 μL aliquots (20 times). Glutaraldehyde was added to a final concentration of 1.5 vol %, and the crystal was soaked for 90 min. The concentration of glutaraldehyde was then raised to 3.0%, and the cross-linking reaction was allowed to proceed for an additional 90 min. Once the cross-linking procedure was completed, the buffer was exchanged with distilled water and the crystal mounted for data collection.

X-ray diffraction data were collected at 4 °C to 1.9 Å resolution with a Siemens X-100A area detector, using Cu Kα radiation from a Eliot GX6 rotating anode operating at 30 kV and 30 mA, with a 0.5 mm focusing cup and 0.3 mm collimation. The data were reduced with the program XDS.[22] Initial phases were calculated from the model of elastase complexed to trifluoroacetyl−Lys−Pro−*p*-isopropylanilide (TFA−Lys−Pro−ISO) solved at 1.8 Å resolution.[9] The solvent molecules and the inhibitor were omitted from this initial model. After application of phases to the measured data from 10 to 1.9 Å and initial round of refinement, the R factor was 29.5%. The protein was rebuilt manually into a difference Fourier electron density map calculated using the coefficients ($3F_{obs} - 2F_{calc}$). XPLOR positional refinement alternating with manual refitting to a difference Fourier electron density map using the coefficients ($2F_{obs} - F_{calc}$) was performed for six cycles. The water molecules added in the second cycle of refinement were taken from positions that were found in the TFA−Lys−Pro−ISO inhibited elastase structure.[9] In the next round of refinement these water molecules were retained only if they were still visible in an electron density map using coefficients ($2F_{obs}$

TABLE 1: Data Collection and Refinement Statistics

	aqueous PPE	cross-linked aqueous	PPE in acetonitrile
		Data Collection	
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
unit cell (Å, deg)	$a = 51.11$, $b = 58.05$, $c = 75.26$ $\alpha = \beta = \gamma = 90.0$	$a = 51.05$, $b = 58.12$, $c = 75.28$ $\alpha = \beta = \gamma = 90.0$	$a = 51.76$, $b = 58.02$, $c = 75.53$ $\alpha = \beta = \gamma = 90.0$
resolution (Å)	∞ to 1.6	∞ to 1.9	∞ to 2.2
temperature (°C)	4	4	4
number of unique reflections	27886	11558	10441
completeness (%)	92	66	88
		Refinement	
resolution range (Å)	10.0−1.6	10.0−1.9	10.0−2.2
final *R* factor (%)[a]	17.8	15.8	16.1
restraints (rms observed)			
bond lengths (Å)	0.013	0.014	0.015
bond angles (deg)	2.62	2.81	3.00
dihedral angles (deg)	29	27	27
improper angles (deg)	1.0	1.1	1.2
average *B* factor (Å^2)			
main chain	16	11	17
side chain	20	14	20
water	31	26	31
total number of protein atoms	1822	1822	1822
number of waters	155	165	127
number of sulfate ions	2	1	2
number of calcium ions	1	1	1
number of acetonitrile molecules	0	0	9

[a] R factor $= \Sigma_{\mathbf{h}}|F_o(\mathbf{h}) - F_c(\mathbf{h})|/\Sigma_{\mathbf{h}}F_o(\mathbf{h})$.

$- F_{calc}$) at the 1σ contour level or higher. In subsequent refinement, using an electron density map calculated with the coefficients ($F_{obs} - F_{calc}$), water molecules were added by manually fitting to density at the 3σ contour level. The final *R* factor versus all data was 15.8% with good geometry for the model (see Table 1).

Cross-Linked Elastase in 100% Acetonitrile. Native elastase crystals were first transferred to the cross-linking buffer as described above. An 8% glutaraldehyde solution was added stepwise to a final concentration of 4%. The cross-linking reaction was allowed to proceed overnight. The crystals were subsequently washed several times with an excess of the cross-linking buffer without glutaraldehyde, followed by several washings with 10 mM sodium acetate buffer, pH 5.0. The sodium acetate buffer was replaced with anhydrous acetonitrile in a stepwise fashion until the crystals were sitting in neat acetonitrile.

X-ray diffraction data were collected at 4 °C to 2.2 Å resolution with a Siemens X-100A area detector, using Cu Kα radiation from a Eliot GX6 rotating anode operating at 30 kV and 30 mA, with a 0.5 mm focusing cup and 0.3 mm collimation. The data were reduced with the program XDS.[22] The program XPLOR was used for initial phase calculations and positional refinement.[16,17] The initial phases were calculated from the coordinates of native elastase refined at 1.65 Å resolution[15] from which all solvent molecules were removed. After the first round of refinement, the *R* factor dropped to 22.7%. A difference Fourier electron density map with coefficients ($2F_{obs} - F_{calc}$) was calculated, and the model of the protein was observed to fit the electron density very well. A difference Fourier electron density map with coefficients ($F_{obs} - F_{calc}$) was calculated and used to build the water molecules and ions that had been excluded from the initial model. Only waters with electron density at the 3σ contour level or higher were retained. Additional water molecules were fitted using the program WATERHUNTER (S. Sugio, personal communication) and checked in a difference Fourier electron density map with coefficients ($F_{obs} - F_{calc}$). The program WATERHUNTER places water based on the sigma level of the electron density map. The water positions are then checked on a graphics system

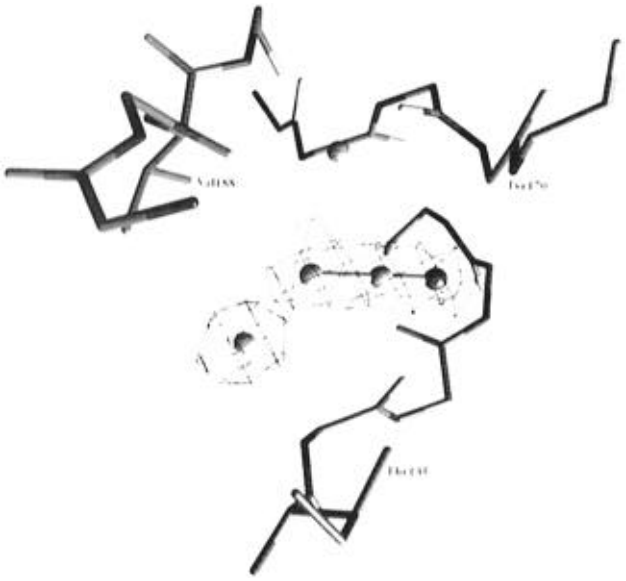

Figure 1. A $2F_o - F_c$ electron density map contoured at the 1.2σ level shows a distinct ellipsoidal density for acetonitrile 707 and a spherical density for a nearby water molecule. The protein backbone of the binding pocket is represented by a coil, with no individual atoms shown. The nitrogen atom of the acetonitrile molecule is shown in dark gray and the carbons are in a lighter gray. Molscript[26] was used in the preparation of this figure.

for H-bonding partners on the surface of the protein. At this point, using a model that contained the well-refined protein atoms and the crystallographic water molecules, several acetonitrile molecules were easily distinguished in the electron density map. An acetonitrile molecule has a large, ellipsoidal electron density, whereas a water molecule fits into a smaller, spherical electron density (Figure 1). The model of acetonitrile was fit to an electron density map calculated with coefficients ($2F_{obs} - F_{calc}$) contoured at the 1.2σ level. All but four of the acetonitrile molecules are fully within electron density at the 1.4σ level. The ambiguity in the orientation of the acetonitrile within the density was resolved by optimizing hydrogen bonding between the N of acetonitrile and a donor on the protein or

crystallographic water molecules. Four subsequent cycles of refinement and rebuilding resulted in a structure with a final *R* factor of 16.1% versus all data. The final refinement statistics are given in Table 1.

Results and Discussion

The MSCS method is based on the analysis of a series of crystal structures of the target protein in several different organic solvents, representative of functional groups which may be used as part of a larger ligand. The surface of elastase is being mapped by this method using various organic solvents. Although in the present article only the structure in acetonitrile is reported to illustrate the general methodology, the others will be included in a subsequent paper describing the application of the MSCS method.

General Method. Proteins are generally not stable in most organic solvents, and therefore it is necessary to cross-link the crystals in aqueous solution before transferring them into the various organic solvents of interest. The cross-linking is done with gluteraldehyde, which randomly links lysine side chains within a single molecule, or between two neighboring molecules in the crystal. During the development of the cross-linking protocol it was found that most crystals that were cross-linked in 1.5% or 4% glutaraldehyde retained their mechanical integrity and diffracted to high resolution as observed by X-ray precession photography. Neither a change in the space group nor significant changes in the unit cell parameters were observed for any of the cross-linked aqueous or cross-linked, organic soaked crystals. Thus, for each one of the structures the enzyme is first crystallized in aqueous solution and then cross-linked with gluteraldehyde. The concentration of the cross-linker may vary with different proteins, but the conditions should be such that the cross-linking is random and sufficient to maintain the crystal integrity in organic solvents. Concentrations of glutaraldehyde over 4% and combinations of other cross-linking agents resulted in the crystals shattering, formation of oil droplets around crystals, or loss of diffraction. It may be argued that the cross-linking could affect the structure of the protein, and to show that this is not the case, it must be compared to the native structure. This is an important comparison which validates the assumption carried throughout the application of the MSCS method, that the binding pockets found for organic solvent molecules in the cross-linked structure are the same as those which would be found in the native enzyme.

The cross-linking solution is removed during several washes with the crystallization buffer or distilled water (especially for nonpolar solvents), and the crystal is transferred in a stepwise fashion into the organic solvent at its final concentration (>99% in the case of elastase in acetonitrile). During the organic solvent soaking procedure, crystals sometimes developed small cracks and fissures, but these did not affect diffraction significantly.

Depending on the solvent being used, the crystal may not diffract in neat solvent, and a given amount of water must be retained. The precise proportion of water needed can be obtained by stepwise decreasing the percent organic solvent until diffraction is observed. For elastase, the lowest organic solvent concentration was with DMF at 55% in water (unpublished results). The fact that several DMF molecules are found to bind to the surface of elastase under these conditions is an indication that for the MSCS method it is not crucial that the structures be solved in neat organic solvent. This is not the case for structural determination in which the goal is to understand the structural basis for enzymatic reactions in organic solvents.[6,7]

The model for the protein, stripped of all of the crystallographic water molecules, is used as the starting model for refinement of the structures in organic solvents. After visually checking the electron density for the entire protein and refining with protein atoms alone, one then includes all of the unambiguous water molecules. Only after refinement with all unambiguous water molecules does one include the organic solvents in the model for the final rounds of refinement.

TABLE 2: Rms Deviations between Each Pair of Elastase Structures (Å)[a]

	native	cross-linked aqueous	cross-linked acetonitrile
native		0.76	0.47
cross-linked aqueous	0.13		0.70
cross-linked acetonitrile	0.19	0.20	

[a] The upper right corner contains the rms for all atoms, and the lower left corner contains the rms for main chain atoms only.

Acetonitrile as the Probe Molecule. Acetonitrile was chosen as the example probe molecule to illustrate the MSCS methodology. It is a probe for possible sites of interactions of amphipathic nitrogen-containing species. It is miscible in aqueous buffer, and its ellipsoidal shape can easily be distinguished from a spherical water molecule in a high-resolution crystal structure. A comparison between the protein structure in aqueous solution and in acetonitrile is important to establish that the solvent does not significantly change the structure of the protein. This is essential because the functional groups represented by the organic solvent should be introduced to the same binding pockets in the aqueous protein, where a drug is to bind.

Table 2 contains the root-mean-square (rms) deviation between pairs of the three structures presented in this article. The lower left corner of the table contains the rms deviation including the backbone atoms only, whereas the upper right contains the rms deviation between all protein atoms in the two structures being compared. It is clear that the backbone of elastase is very similar in all three structures. The greater rms values obtained when all atoms are included result from a few freely rotating side chains located on the protein surface. Both the main chain and overall rms deviations obtained for these structures are typical of rms deviations obtained between any two elastase structures solved in aqueous solvent. The average rms deviation between 10 structures of elastase complexed to various inhibitors or substrates in aqueous solution was found to be 0.2 Å for main chain atoms and 0.6 Å for all protein atoms.[9,23] The rms deviations reported in Table 2 establish that neither the cross-linking alone nor infusion with acetonitrile after cross-linking caused any significant changes in the structure of elastase. Furthermore, the changes that do occur are not due to the presence of the acetonitrile but reflect changes that are usually found between any two independently solved structures of elastase.

There are two sulfate ions and a calcium normally found in the structure of native elastase.[15] One of the sulfates is in the active site and is not present in the structures of elastase complexed to various ligands.[23] The other sulfate ion and the calcium are both found in most of the elastase structures reported to date. The native elastase structure which is being compared to the cross-linked elastase and elastase in acetonitrile contains all three ions at the positions previously reported. The cross-linked elastase crystal in aqueous solution, however, was washed with distilled water. This procedure resulted in the active site sulfate being replaced by a water molecule which forms a hydrogen bond with the Oγ atom of the active site Ser 203 (2.7 Å). The cross-linked crystal which was subsequently transferred to acetonitrile was not washed with distilled water,

Figure 2. Elastase structure represented by the ribbon diagram,[27] where the overall secondary structural features are shown as a solid helix, strand, or coil. Elastase contains 240 amino acid residues, and it is composed of two β-barrel domains. The catalytic triad (Asp 108, His 60, and Ser 203) and the nine acetonitrile molecules are shown explicitly. The nitrogen atom of the acetonitrile molecule is in dark gray, and the carbons are in a lighter gray. Molscript[26] was used in the preparation of this figure.

and the three ions remain in their respective positions on the surface of the protein.

Acetonitrile Binding Sites. Nine acetonitrile molecules were found on the surface of elastase (Figure 2), providing the first set of binding pockets which will subsequently be added to those found for other solvents in the application of the MSCS method. Two of the acetonitrile molecules are in the active site (numbered 701 and 702 in the coordinates deposited in the PDB). Three others (703, 704, and 705) are involved in crystal contacts and probably would not be found outside of the crystal environment. The remaining four (706−709) are scattered in different regions of the elastase surface. The average *B* factor for the nine acetonitrile molecules is 35.8 Å^2, comparable to the average *B* factor for structural water molecules (31.0 Å^2). Of the nine acetonitrile molecules, seven have *B* factors which are similar to the average *B* factor for waters. Acetonitrile 701 has a lower *B* factor of 18.4 Å^2. Acetonitrile 705 has a comparatively high *B* factor of 66.6 Å^2 but was retained because of the strong electron density to support its inclusion in the model.

The two acetonitrile molecules found in the active site are positioned in the S_1 and S_2 subsites,[23] and both are within 4.0 Å of the sulfate ion found in the oxyanion hole.

The S_1 subsite is known to prefer small aliphatic amino acids such as Ala or Val. It is therefore not surprising that the CH_3 group of acetonitrile 701 is directed toward the interior of the S_1 subsite, interacting with the protein as previously observed for small nonpolar amino acids. The N atom of acetonitrile 701 interacts mainly with the Oγ atom of Ser 203 (2.8 Å), and it is also 3.9 Å from one of the oxygen atoms of the sulfate ion. Acetonitrile molecule 702 binds in the S_2 subsite with its hydrophobic portion (CH_3−C) stacking against the active site histidine 60 (∼3.5 Å). The N atom of acetonitrile 702 is 2.8 Å from the sulfate O atom that interacts with N of acetonitrile 701.

The three acetonitrile molecules found at crystal contacts bind in pockets that have contributions from two molecules in the unit cell. In the coordinates deposited in the Protein Data Bank, only one molecule of elastase is included. For that reason it may seem that a single elastase molecule provides this binding pocket. Although these acetonitrile molecules are not in true binding pockets on the surface of elastase, their presence is justifiable in the context of the crystalline environment.

Acetonitrile 703 binds with its N atom interacting with a guanidinium nitrogen atom of Arg 196 (2.8 Å) and with the backbone N atom of residue 164 (3.0 Å). It is 3.0 Å from a water molecule which is not found in either the native elastase or cross-linked elastase structure in aqueous solution. This water molecule is 2.9 Å from the Oγ_1 atom of Thr 20 and 2.9 Å from the carbonyl oxygen atom of Gly 82 of a symmetry-related molecule. Acetonitrile 703 is involved through this unique water molecule in an H-bonding network which spans between the two molecules in the crystal contact. The hydrophobic portion of this acetonitrile molecule interacts with Tyr 123 of the symmetry-related molecule (3.5 Å). Both in the native and in the cross-linked structures of elastase in aqueous solution there is a water molecule at the position occupied by acetonitrile 703.

Acetonitrile 704 is the second of the three acetonitrile molecules found in a crystal contact. Its N atom is 2.9 Å from the side chain nitrogen atom of Asn 212 and 3.2 Å from a symmetry-related water molecule. This water molecule, in turn, forms hydrogen bonds to the symmetry-related residues Val 94 and Ile 92. (The distance from the water to the backbone N of Val 94 and O of Ile 92 is 3.1 Å.) As was the case for acetonitrile 703, there is an H-bonding network, which involves acetonitrile 704, between two symmetry-related molecules in the unit cell. The hydrophobic contact formed by the CH_3 group of acetonitrile 704 is with the aliphatic portion of Lys 91 of the symmetry-related molecule. Acetonitrile 704 occupies a position coinciding with that of a water molecule in the native elastase structure.

Acetonitrile 705 is the third and final probe molecule found in a crystal contact. Its N atom makes an H bond to a water molecule (3.4 Å), which in turn forms H bonds to O of Gly 38 (3.3 Å) and to a guanidinium nitrogen of Arg 36 (3.3 Å). The fact that acetonitrile 705 does not form an H bond directly with the protein, but does so through a water molecule, may account for the relatively high *B* factor obtained for this acetonitrile molecule. Arg 36 is involved in a crystal contact, with the other guanidinium N 3.2 Å from the side chain oxygen atom of Glu 102 of a symmetry-related molecule. The hydrophobic portion of acetonitrile 705 stacks against the aliphatic portion of the symmetry-related Lys 184 side chain (4.0 Å), making a good van der Waals interaction. There is no water molecule occupying the position of acetonitrile 705 in either the native or the cross-linked elastase structure.

The remaining four acetonitrile molecules are found in binding pockets formed on the surface of a single elastase molecule. Acetonitrile 706 is positioned with its N atom 3.0 Å from the N atom of Ser 246 and 3.1 Å from the side chain Oγ atom of the same serine. Acetonitrile 706 replaces two water molecules which are found in the native enzyme. In the cross-linked structure only one of these water molecules is present. Acetonitrile molecule 707 is found with its N atom 2.8 Å from the backbone nitrogen atom of Tyr 170 and 3.3 Å from a water molecule. Its CH_3 group is 4.1 Å from a side chain methyl group of Val 188. This acetonitrile replaces two water

Figure 3. Bound TFA−Lys−Pro−ISO inhibitor structure[23] superimposed on the elastase in acetonitrile structure. Only the active site region is shown, with two acetonitrile molecules. The protein backbone is represented by coil, and some of the active site side chains are shown explicitly. The nitrogen atom of the acetonitrile molecule is in dark gray, and the carbons are in a medium shade of gray. The inhibitor N atoms are dark gray, the inhibitor C and F atoms are a medium shade of gray, and the inhibitor O atoms are light gray. Hydrogen atoms are not shown, since their positions are not observable in electron density maps. Molscript[26] was used in the preparation of this figure.

molecules found both in the native elastase and in the cross-linked elastase structures. Acetonitrile 708 binds to elastase with its N atom 3.8 Å from a guanidinium nitrogen atom of Arg 51 and 3.8 Å from a water molecule which is also found in both the native and cross-linked elastase structures. Acetonitrile 708, however, does not replace any observed water molecule in either the native or cross-linked elastase structure. Acetonitrile 709 is in a pocket where its N atom is 3.0 Å from the backbone N atom of Gly 18, and its CH_3 group is 3.6 Å from the β carbon atom of Thr 150. Other protein atoms, such as backbone carbonyl oxygens, are within 4.0 Å of acetonitrile 709, lining the cavity in which it binds. This acetonitrile molecule replaces a water molecule which is present in both the native and cross-linked structures of elastase in aqueous solution.

Comparison with Inhibitor Structures. The position of molecules within the active site of elastase in acetonitrile is consistent with what is known about preferences within each of the available subsites. To illustrate this notion, it is useful to superimpose the acetonitrile structure with two of the TFA−dipeptide−anilide inhibitors[23] and with the acetyl−Ala−Pro−Val−fluoro-*n*-phenylelthyl amide (APVP) inhibitor.[24] The latter inhibitor forms a tetrahedral adduct with elastase. In the TFA−Lys−Pro−ISO/elastase complex there is an acetate molecule in the oxyanion hole, in a position similar to that occupied by the sulfate ion in the acetonitrile structure (see Figure 3). Both the acetate and the sulfate may approximately mimic the interactions of the tetrahedral intermediate in the oxyanion hole. The trifluoroacetyl group in the TFA−Lys−Pro−ISO inhibitor is found in the S_1 subsite where acetonitrile 701 is positioned. The acetonitrile molecule 702 approximately coincides with the N, Cα, and Cβ atoms of the inhibitor Lys residue. These atoms differ by respectively 1.6, 1.0, and 0.54 Å from the N, C, and CH_3 groups of the acetonitrile molecule.

In the second inhibitor complex, TFA−Lys−Phe−ISO, the TFA group is in the oxyanion hole and occupies approximately the position where the sulfate ion is seen in the acetonitrile structure. The positions of the carbon of the CF_3 group in the inhibitor complex and the S of the sulfate in the acetonitrile structure differ by 0.56 Å, and the fluorine atoms are oriented to interact with the N 201 and N 203 in the oxyanion hole. The Phe residue of the inhibitor is found in the S_1 subsite. The N of acetonitrile 701 in this site roughly coincides with the center of the phenylalanine ring in the inhibitor structure. The C−CH_3 portion of this acetonitrile is oriented toward the interior of the subsite. The lysine of this inhibitor has the same position relative to the acetonitrile in the S_2 subsite as does the Lys in TFA−Lys−Pro−ISO.

The APVP inhibitor binds to elastase with the Val, Pro, and Ala residues occupying the S_1, S_2, and S_3 subsites, respectively. A tetrahedral adduct is formed with Ser 203, and the remaining portion of the molecule binds in the S_1' and S_2' subsites, where it forms weak interactions with the protein.[24] As in the TFA−Lys−Phe−ISO inhibitor, the S of the sulfate group is 0.67 Å from the CF_2 group. The N atom of acetonitrile molecule 701 is 0.47 Å from the β carbon atom of the inhibitor Val residue and binds with its hydrophobic CH_3−C portion further into the S_2 pocket. The position of the N of acetonitrile 702 differs by 1.0 Å from the backbone nitrogen of the inhibitor Pro residue, and the CH_3−C superimposes with the proline ring.

Although a single probe molecule is not sufficient to guide the design of new inhibitors, the two acetonitrile molecules, together with the sulfate ion found in the active site, coincide in general with the type of functional groups found in similar positions in the known inhibitor molecules. Clearly, a negatively charged or electronegative group is favored in the oxyanion hole. This is demonstrated by the persistent appearance of the sulfate group in elastase structures not complexed to an inhibitor or substrate. The binding of acetonitrile in the S_1 pocket is consistent with the hydrophobic environment of this subsite. Finally, in the three inhibitors considered above, the position of the N of the backbone of the P_2 residue is found close to that of the N of acetontitrile 702, and the side chain of the P_2 residue goes into the S_2 pocket in the general direction indicated by the hydrophobic portion of the acetonitrile probe.

Conclusion

The MSCS method is a novel experimental approach to drug design which is complementary to already existing strategies. Although a protein was used to develop this methodology, it can be applied to any molecule which can provide binding pockets and is tractable to X-ray crystallographic structure determination. The MSCS method promises to have a direct application in drug design, where various probe molecules are used to guide the development of new inhibitors with high specificity to the target protein.[25] This does not mean that the binding positions of any given solvent alone (as presented in this article) will result in exact placement of similar functional groups within a larger ligand but that when several solvent positions are taken together they serve as an initial template from which the drug design process can begin.

The MSCS method does have the limitation that not all interesting probe molecules are liquids, and of those that are, only a portion can be tolerated by the cross-linked protein crystals without loss of diffraction. One of the great advantages of the crystallographic method is that it locates only a few (usually fewer than 10) binding sites for a given solvent. The MSCS method will therefore serve a second important role. It provides a basis for the further development and optimization of the theoretical drug design methodology embodied in the multiple copy simultaneous search (MCSS) program.[3] Empirical information obtained from the MSCS method is currently being incorporated into the MCSS program to improve selection of the relevant minima that represent the binding pockets used in the drug design process (Mattos and Karplus, personal communication). The MSCS method provides the only experimental method for discovering the existence of binding sites other than those where substrate and effector molecules bind.

Acknowledgment. We thank David H. Harrison and R. Lynn Rardin for their help in data collection and data reduction and Arnon Lavie for his help in developing the cross-linking procedures. Also many thanks to Eben Kunz for help preparing figures.

References and Notes

(1) Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. *Acc. Chem. Res.* **1994**, *27*, 117−123.
(2) Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849−857.
(3) Miranker, A.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29−34.
(4) Caflisch, A.; Miranker, A.; Karplus, M. *J. Med. Chem.* **1993**, *36*, 2142−2167.
(5) Klibanov, A. M. *Trends Biochem. Sci.* **1989**, *14*, 141−144.
(6) Fitzpatrick, P. A.; Steinmetz, A. C. U.; Ringe D.; Klibanov, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8653−8657.
(7) Yennawar, N. H.; Yennawar, H. P.; Farber, G. K. *Biochemistry* **1994**, *33*, 7326−7336.
(8) Mattos C.; Ringe, D. Multiple Binding Modes. In *QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers: Leiden, The Netherlands, 1993; pp 226−254.
(9) Mattos, C.; Rasmussen, B.; Ding, X.; Petsko, G. A.; Ringe, D. *Nature: Struct. Biol.* **1994**, *1*, 55−58.
(10) Rydel, T. J.; Tulinsky, A.; Bode, W.; Huber, R. *J. Mol. Biol.* **1991**, *221*, 583−601.
(11) Qiu X.; et al. *Biochemistry* **1991**, *31*, 11689.
(12) Powers, J. C. *Trends Biochem. Sci.* **1976**, *1*, 211−214.
(13) Buchardi, H. *Adv. Exp. Med. Biol.* **1984**, *167*, 319−333.
(14) Geokas, M. C.; Murphy, R.; McKenna, R. D. *Arch. Pathol.* **1968**, *86*, 117−126.
(15) Meyer, E.; Cole, G.; Radhakrishnan, R. *Acta Crystallogr., Sect. B* **1988**, *44*, 26−38.
(16) Brünger, A. T. *J. Mol. Biol.* **1988**, *203*, 803−816.
(17) Brünger, A. T.; Kuriyan, J.; Karplus, M. *Science* **1987**, *235*, 458−460.
(18) Jones, T. A. *Methods Enzymol.* **1985**, *115*, 157−171.
(19) Quiocho, F. A.; Richards, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *52*, 833−839.
(20) Quiocho, F. A.; Richards, F. M. *Biochemistry* **1967**, *5*, 4062−4076.
(21) Quiocho, F. A.; Bishop, W. H.; Richards, F. M. *Proc Natl. Acad. Sci. U.S.A.* **1967**, *57*, 525−537.
(22) Kabsch, W. *J. Appl. Crystallogr.* **1988**, *21*, 67−71, 916−924.
(23) Mattos C.; Giammona, D. A.; Petsko, G. A.; Ringe, D. *Biochemistry* **1995**, *34*, 3193−3203.
(24) Takahashi, L. H.; Radhakrishnan, R.; Rosenfield, R. E., Jr.; Meyer, E. F., Jr.; Trainor, D. A. *J. Am. Chem. Soc.* **1989**, *111*, 3368−3374.
(25) Peisach, E.; Casebier, D.; Gallinon, S. L.; Furth, P.; Petsko, G. A.; Hogan, G. C.; Ringe D. *Science* **1995**, *269*, 66−69.
(26) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.
(27) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167−339.

JP952516O