doi:10.1016/j.jmb.2006.01.039 J. Mol. Biol. (2006) xx, 1–12

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Multiple Solvent Crystal Structures: Probing Binding Sites, Plasticity and Hydration

Carla Mattos[1]*, Cornelia R. Bellamacina[2], Ezra Peisach[2], Antonio Pereira[2] Dennis Vitkup[2], Gregory A. Petsko[2] and Dagmar Ringe[2]*

[1]*Department of Molecular and Structural Biochemistry, North Carolina State University Campus Box 7622, 128 Polk Hall, Raleigh, NC 27695, USA*

[2]*Rosenstiel Basic Medical Sciences Research Center Department of Chemistry and Biochemistry, Brandeis University, 415 South Street Waltham, MA 02254-9110 USA*

**Corresponding authors*

Multiple solvent crystal structures (MSCS) of porcine pancreatic elastase were used to map the binding surface the enzyme. Crystal structures of elastase in neat acetonitrile, 95% acetone, 55% dimethylformamide, 80% 5-hexene-1,2-diol, 80% isopropanol, 80% ethanol and 40% trifluoroethanol showed that the organic solvent molecules clustered in the active site, were found mostly unclustered in crystal contacts and in general did not bind elsewhere on the surface of elastase. Mixtures of 40% benzene or 40% cyclohexane in 50% isopropanol and 10% water showed no bound benzene or cyclohexane molecules, but did reveal bound isopropanol. The clusters of organic solvent probe molecules coincide with pockets occupied by known inhibitors. MSCS also reveal the areas of plasticity within the elastase binding site and allow for the visualization of a nearly complete first hydration shell. The pattern of organic solvent clusters determined by MSCS for elastase is consistent with patterns for hot spots in protein–ligand interactions determined from database analysis in general. The MSCS method allows probing of hot spots, plasticity and hydration simultaneously, providing a powerful complementary strategy to guide computational methods currently in development for binding site determination, ligand docking and design.

Keywords: elastase; solvent mapping; organic solvents; protein binding sites; multiple solvent crystal structures

Introduction

One of the current challenges in structural biology is to understand the general features that guide protein–ligand interactions, allowing the prediction of these interactions given the structure of the unbound components.[1] The binding process is thought to occur in a step that involves recognition, followed by rearrangements that optimize packing at the interface.[2–4] Therefore, an important aspect in understanding interfaces and the formation of complexes is to determine what distinguishes ligand-binding sites from other areas on protein surfaces, addressing the question posed a decade ago: what makes a binding site a binding site?[5] Indeed, the last decade has seen intense efforts to characterize the features that distinguish binding sites from other areas on protein surfaces. Initial database analysis determined that binding sites exhibit no general patterns of hydrophobicity, shape or charge,[6] although for small ligands there is a correlation with shape. Furthermore, it quickly became clear that plasticity plays a major role in protein–ligand interactions.[7] The dynamic component of ligand binding as well as the thermodynamic contribution of solvation effects have contributed to the difficulty of predicting *a priori* specific residues or binding pockets that are hot spots for ligand binding affinity.[8] In addition, the role of bound water molecules in mediating protein–ligand interaction cannot be ignored.[9] In spite of the difficulties, however, a combination of database analysis,[4] experimental observations on individual complexes,[10] the development of computational methods for predicting the location of binding sites[11] and docking of binding partners[12] have contributed to significant advancements in this area of research.

The multiple solvent crystal structures (MSCS, elastase models from cross-linked crystals approach

Abbreviations used: MSCS, multiple solvent crystal structures.

E-mail addresses of the corresponding authors: carla_mattos@ncsu.edu; ringe@brandeis.edu

0022-2836/$ - see front matter © 2006 Published by Elsevier Ltd.

provides a robust experimental method to locate and characterize ligand binding sites on proteins using organic solvents.[13,14] The method is fully developed here using porcine pancreatic elastase as a model enzyme in various solvent conditions: XLINK, aqueous solution; ACN, neat acetonitrile; DMF, 55% dimethylformamide; HEX, 80% 5-hexene-1,2-diol; TFE1, 40% 2,2,2-trifluoroethanol (cross-linked crystals were not transferred to distilled water); TFE2, 40% 2,2,2-trifluoroethanol; ETH, 80% ethanol; ACE, 95% acetone; ISO, 80% isopropanol; IBZ, 50% isopropanol, 40% benzene; ICY, 50% isopropanol; 40% cyclohexane). As different protein crystals have been successfully transferred to organic solvents for X-ray structure determination,[15–21] the MSCS method is gaining recognition as a method to study ligand binding sites when well diffracting crystals are available, as the organic solvent molecules probe the surface in a way that is not possible within the constraints of a larger molecule that occupies the entire binding site. The crystal structures of thermolysin in different concentrations of isopropanol[20] and in three other organic solvents[21] reveal that the organic solvents cluster in the known active site and appear in some areas of crystal contacts. Furthermore, the authors used the MSCS method to qualitatively rank the affinity of different thermolysin subsites for isopropanol.[20] The work on thermolysin includes a comparison of the experimental organic solvent binding sites with those obtained computationally using the multiple copy simultaneous search (MCSS)[22] and the GRID[23] methods. Only poor agreement was obtained between the experimental and computational results. Indeed, these calculations are unrealistic because they do not include solvation effects and the plasticity of the protein structure.[24] More recently, a computational counterpart to the MSCS method has shown great success in predicting the location of the primary binding pockets in enzymes.[25,26]

One long-term use of MSCS might be in ligand design, where the organic solvents provide experimental positions for functional groups that can be incorporated into larger ligands.[27–29] A strategy for ligand design that has been discussed for many years in molecular modeling approaches relies on the idea that functional groups can be optimized independently for different regions of a protein binding site.[22,23] These functional groups can then be linked to form a ligand with high affinity and specificity to the target protein.[27,28] The protein binding affinity of the resulting molecule will be, in principle, the product of the binding constants for the individual fragments plus a term that accounts for changes in binding affinity due to the linker portion of the larger ligand.[30] Though viable in principle, this strategy has met with only limited success, notably in cases when fragment positions were obtained experimentally,[30] because a major difficulty with the computational linked-fragment based approach is the lack of reliability in predicting optimal binding modes for the fragments tested. The use of organic solvent binding as an experimental approach to identify fragments for ligand design has been previously suggested as an immediate consequence of the MSCS method.[13,17] The focus in the present study is to further explore ways in which MSCS can be used to more fully characterize the surfaces of proteins. In addition to binding at specific sites, the organic solvents provide changes in the protein environment, inducing structural adjustments in areas of plasticity and influencing the way in which surface water molecules interact with the protein.[31] The MSCS method provides a unique experimental approach to characterize active sites of enzymes, while probing protein plasticity and surface hydration simultaneously throughout the entire structure. This is a critical step in understanding the complex protein template targeted for ligand design.

Results

Porcine pancreatic elastase (from hereon referred to as elastase) is a serine protease of the trypsin family, with 240 amino acid residues. It is composed of two β-barrel domains, with the catalytic triad (Ser206, His60 and Asp108) residing in the cleft between the domains.[13,17,32,33] The fact that elastase has been well studied makes it a particularly appropriate model to explore the potential of the MSCS method to characterize the surfaces of enzymes in general.

Elastase crystals grown in aqueous solution were cross-linked with glutaraldehyde and transferred to a series of solutions containing high concentrations of organic solvents in water as described in Materials and Methods. The crystal structures of elastase in each of ten conditions were solved, revealing the way in which the organic solvent molecules interact on its surface. No cross-links were observed in any of the electron density maps. Furthermore, the cross-linking does not affect the overall structure of elastase. Each type of solvent molecule binds to the surface of the protein in a unique fashion, reflecting its shape and chemical properties.

For the present analysis nine crystal structures of elastase were added to the one previously solved in neat acetonitrile (ACN).[17] We present here the crystal structures of elastase solved in the presence of 95% acetone (ACE), 55% dimethylformamide (DMF), 80% 5-hexene-1,2-diol (HEX), 80% isopropanol (IPR), 80% ethanol (ETH) and two structures solved in 40% trifluoroethanol (TFE1 and TFE2). To obtain the TFE1 model, the cross-linking phosphate buffer solution containing the elastase crystal was exchanged directly with a 40% trifluoroethanol solution, while for TFE2 the crystal was first exchanged into distilled and de-ionized water. In addition, a crystal structure was solved in a mixture of 40% benzene, 50% isopropanol and 10% water

(IBZ) and one in 40% cyclohexane, 50% isopropanol and 10% water (ICY). Isopropanol was bound to elastase in these last two structures, but there were no bound benzene or cyclohexane molecules observed in the electron density maps. For the MSCS analysis, each of the ten crystal structures was superimposed onto the previously published structure of cross-linked elastase in aqueous solution.[17]

Comparison between the elastase models

The elastase crystals in organic solvents are isomorphous with those in aqueous solution. They have the symmetry of space group $P2_12_12_1$ and unit cell dimensions of approximately $a=51$ Å, $b=58$ Å and $c=75$ Å. The unit cell parameters, data collection and refinement statistics are presented in Table 1 for each of the nine previously unpublished structures. All of the elastase models contain a calcium ion and one sulfate ion in the positions observed in previous structures of elastase solved in aqueous solution[34] (Table 1). A second sulfate ion is found in the oxy-anion hole in the ACE, DMF and HEX models. In the other structures, either a water molecule or one of the organic solvent molecules is found at this site. The number of water molecules found in each model is given in Table 1.

The root-mean-square deviations for pairs of MSCS structures range between 0.12 and 0.33 based on the C^{α} positions of elastase residues. This confirms the visual observation that overall there is no significant difference between elastase models in the various solvent environments. When comparing the structures in detail, it becomes clear that they vary primarily in the conformations of side-chains that have high *B*-factors in aqueous solution. However, a couple of notable exceptions involving main-chain atoms do occur. Residues 24–27 (RNSW) and 122–123 (SY), two regions that interact within the structure, adopt one of two very distinct backbone and side-chain conformations. In the presence of solvents of higher dielectric constants (XLINK, ACN, DMF, HEX, and TFE2/water) several of the charged or polar groups are exposed, while in environments of lower dielectric constant (TFE1/buffer, ETH, ACE, IPR, IBZ, and ICY) the resulting conformation maximizes H-bonding between the protein atoms, with polar groups turned away from the protein surface.

The organic solvent binding sites

In general, only a few organic solvent molecules were found in each of the models (Table 1), which is a far cry from the protein surface being solvated by the organic solvents. Figure 1 shows a ribbon diagram of elastase with 38 organic solvent molecules taken collectively from the ten superimposed crystal structures of elastase solved under different conditions (see Table 2 for a list of sites occupied in each model). There are 16 unique sites for the organic solvents. Molecules occupying the same site have the same number in all coordinate sets. The binding sites are numbered consecutively from 1001 to 1016 and the organic solvent molecules are numbered accordingly in the models. There are seven sites occupied by more than one type of organic molecule, numbered 1001 to 1007 (four of these are in the active site). The other nine sites, 1008 through 1016, bind a single organic solvent molecule (two in the active site). Organic solvent binding sites outside of the active site are located mostly at or near crystal contacts. There are two crystal contact areas that are particularly populated by the organic solvent probes. The first crystal contact group includes sulfate ion 290, which is invariably found in elastase crystal structures. The organic solvent binding sites in this general area of crystal contacts are 1005, 1015 and 1016 (see Table 2 for the identity of the solvents bound at these sites). The organic solvent sites in the other crystal contact area are 1007, 1012 and 1014. Solvent sites 1010, 1011 and 1013 are also in crystal contacts, but site 1006 is not. At this site, Lys234 interacts with either an acetone or an isopropanol molecule through its N^{ε} group. In addition to H-bonding with Lys234, the probes at site 1006 are in van der Waal's contact with Leu227, Val231 and the aliphatic portion of Arg233. This site is 8.5 Å from the S_4 subsite (1002).

The active site

The active site of elastase consists of pockets that bind four or five amino acid side-chains before (S_1–S_4 or S_5) and three after (S_1'–S_3') the scissile peptide bond.[32,35] These subsites have been well characterized through kinetic experiments and several crystal structures of elastase/inhibitor complexes have been published.[32,36–39] Small aliphatic amino acid residues were shown to be favored in the S_1 and S_4 subsites, with S_2 preferring Lys. The S_3 subsite is highly exposed to solvent and has a less distinct preference for any given amino acid residue.

The organic solvent molecular probes provide a map of the protein surface that distinguishes the active site from all other areas. Clustering of organic solvent molecules occurs prominently in the active site, where five of the known subsites on the protease are clearly delineated. Organic solvent molecules are found in S_1 (1001), S_4 (1002), the oxyanion hole (1003) and two sites on the leaving group side of the catalytic triad likely to be the S_1' (1008) and S_3' (1004) binding pockets. Four of these subsites are observed to bind at least three different types of organic solvent molecules. In addition, there is a single acetonitrile molecule at the entrance to the S_2 (1009) pocket[17] making a total of six binding sites for organic solvents in the active site.

While there is a common set of hydrophobic interactions with organic solvents within each of the S subsites, there is significant variability in the polar interactions, depending on the H-bonding potential

Table 1. Data collection and refinement statistics

	HEX	ETH	TFE1	TFE2	IPR	IBZ	ICY	ACE	DMF
A. *Data collection*									
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters									
a, *b*, *c* (Å)	52.60, 58.27, 75.84	52.45, 58.58, 75.32	52.60, 58.07, 75.46	51.88, 57.84, 75.24	52.49, 58.24, 75.39	52.13, 57.92, 74.91	52.02, 57.92, 74.84	52.20, 58.12, 75.05	52.54, 58.01, 75.66
$\alpha=\beta=\gamma$ (deg.)	90	90	90	90	90	90	90	90	90
Resolution (Å)	∞ to 2.2	∞ to 2.0	∞ to 1.9	∞ to 1.8	∞ to 2.2	∞ to 1.9	∞ to 1.9	∞ to 2.0	∞ to 2.2
Temperature (°C)	25	25	0	0	25	25	25	0	4
Number of unique reflections (% complete)	19,334 (85)	14,637 (48)	14,070 (40)	25,281 (62)	20,624 (91)	21,613 (63)	22,509 (65)	12,617 (42)	11,178 (44)
B. *Refinement*									
R factor (%)	16.5	18.4	17.1	16.7	15.4	16.1	16.4	16.9	16.5
R-free (%)	20.6	22.4	21.6	19.3	19.8	20.0	19.7	22.9	22.6
Restraints (rms observed)									
Bond lengths (Å)	0.005	0.005	0.005	0.004	0.005	0.004	0.004	0.005	0.005
Bond angles (deg.)	1.16	1.23	1.23	1.22	1.16	1.22	1.24	1.20	1.18
Dihedral angles (deg.)	27	26	25	25	26	25	26	26	26
Improper angles (deg.)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7
Total no. protein atoms	1822	1822	1822	1822	1822	1822	1822	1822	1822
Number of water molecules	149	134	179	178	149	161	150	126	147
Number of sulfate ions	2	1	1	1	1	1	1	2	2
Number of calcium ions	1	1	1	1	1	1	1	1	1
Number of solvent molecules	2	8	4	2	3	3	5	4	5

$R=\sum_h |F_o(h)-F_c(h)|/\sum_h F_o(h)$.

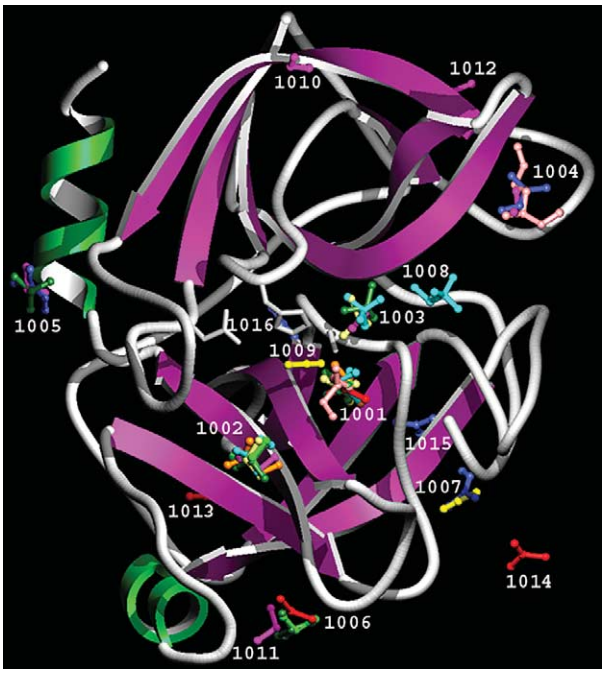

Figure 1. Organic solvent binding sites. Ribbon diagram of elastase showing the binding sites for organic solvent molecules in a common frame of reference. Each site is numbered as described in the text. The number of the sites occupied by organic solvent molecules in each of the models is given in Table 2. The catalytic triad is shown explicitly in the cleft between the two β-barrel domains: Ser203, His60, Asp108 are shown in gray. The β-strands are shown in purple and the two α-helices are shown in green. The organic solvent molecules are color-coded as follows: HEX, salmon; ETH, hot pink; TFE1, cyan; TFE2, orange; IPR, light green; IBZ, green; ICY, dark green; ACE, red; DMF, blue; ACN, yellow. Figures 1–4 were made using the program MOLSCRIPT.[52]

of the molecule involved and its orientation within the pocket. The hydrophobic interactions with solvent molecules in S_1 are provided by three residues that line the bottom of this pocket: Thr221 ($C^{\gamma 2}$), Val224 and Thr236 ($C^{\gamma 2}$). In general the organic solvent molecules are oriented in such a way that their more hydrophobic areas reach deep into the S_1 pocket, with their polar atoms H-bonding to atoms closest to the surface. The S_4 subsite is lined by Phe223, Val103, Ala104, the $C^{\gamma 2}$ atom of Thr182 and the aliphatic portion of Arg226. Both the S_1 and S_4 subsites provide extensive hydrophobic contacts to the bound molecules, consistent with the preference for substrates with apolar residues in these pockets.[34,35] Two Leu residues provide the hydrophobic contacts within each of the S_1' and S_3' pockets. Leu156 contributes to both pockets, while Leu149 is found in S_1' and Leu77 is found in S_3'. In contrast, the interactions within the oxyanion hole are primarily polar in nature. The models HEX, ACE and DMF have a sulfate ion bound in the oxyanion hole, as reported for the structure in neat acetonitrile.[17] The sulfate ions were not included in Figure 1 for clarity. Of the organic solvents, only the alcohols ethanol (ETH), trifluoroethanol (TFE1) and isopropanol (ICY) bind in the oxyanion hole (Figure 1). In all three cases the OH group H-bonds with the backbone N atoms of Gly201 and Ser203 as well as with the O^{γ} of the active site Ser203, mimicking the interactions of the water molecule found at this site in the native and cross-linked structures solved in aqueous solution.[17] In the TFE1 model the hydrophobic trifluoromethyl group is in van der Waal's contact with the trifluoromethyl group of the molecule bound in the S_1' site.

The binding of organic solvents in the S subsites coincides with the way these sites are occupied by elastase inhibitors. Figure 2 shows the inhibitor

Table 2. Data collection and processing conditions

Organic solvent	% Volume	Equipment[a]	Temperature (°C)	Data processing	Phases (PDB code)	Binding sites (see Figure 2)
HEX	80	3	25	Xengen	1ELA	1001, 1004
ETH	80	3	25	Xengen	1ELA	1001, 1002, 1003, 1004, 1005, 1010, 1011, 1012
TFE1	40	2	4	XDS	1ELC	1001, 1002, 1003, 1008
TFE2	40	1	4	Denzo	1ELC	1001, 1002
IPR	80	3	25	Xengen	1ELA	1001, 1002
IBZ	40/50/10	1	25	Process	1ELA	1001, 1002, 1006
ICY	40/50/10	1	25	Process	1ELA	1001, 1002, 1003, 1005, 1006
ACE	95	1	0	Process	3EST	1001, 1013, 1014, 1006
DMF	55	2	4	XDS	3EST	1004, 1005, 1007, 1015,1016
ACN	100	1	4	Process	3EST	1001, 1007, 1009

[a] Three different setups were used for data collection, designated 1, 2 and 3 as follows: (1) R-Axis II phosphoimaging plate mounted on a Rigaku RU200 rotating anode generator operating at 50 kV and 100 mA. (2) Siemens X100-A area detector mounted on an Eliot GX-6 rotating anode operating at 30 kV and 30 mA. (3) Siemens multi-wire area detector with Argonne "Mad" Interface mounted on a Rigaku RU200 generator operating at 50 kV and 100 mA.

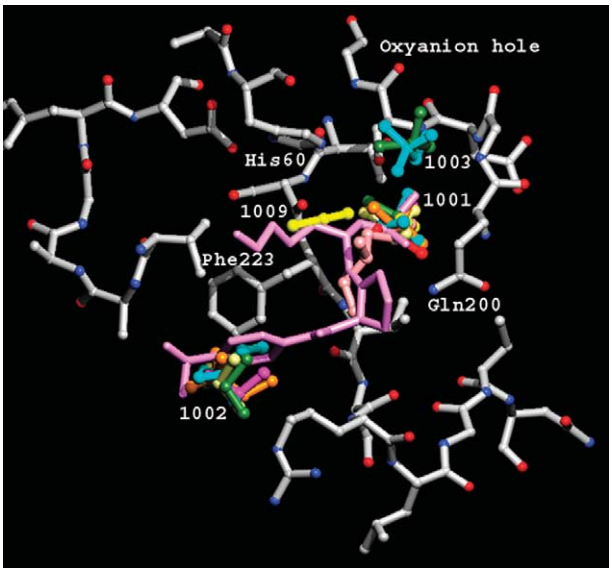

Figure 2. Superposition of organic solvents and inhibitor in the active site of elastase. The organic solvent binding sites are numbered and some of the active site residues are labeled. The protein atoms are depicted with N in blue, O in red and C in gray. The trifluoroacetyl-Lys-Pro-*p*-isopropylanilide inhibitor is in pink and the organic solvents are color-coded as in Figure 1.

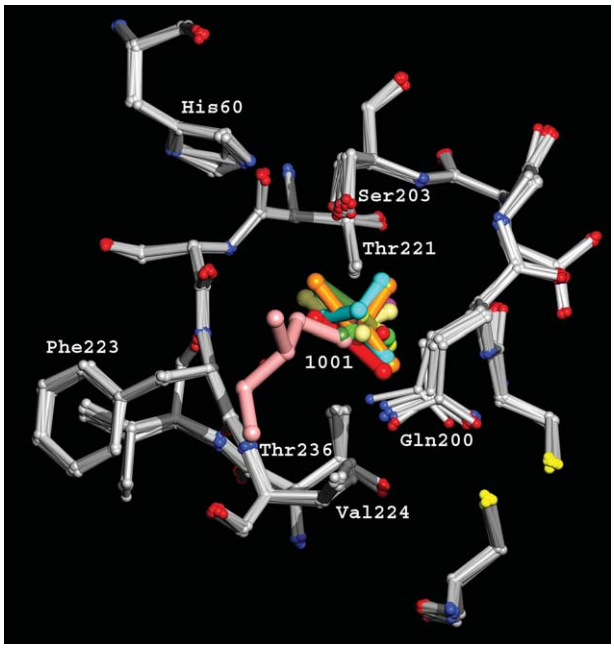

Figure 3. Plasticity in the active site of elastase. The S_1 site is shown with the MSCS models superimposed. Some of the active site residues are labeled. The protein atoms are depicted with N in blue, O in red, S in yellow and C in gray. The organic solvent molecules in the S_1 subsite (1001) are color-coded as in Figure 1.

trifluoroacetyl-Lys-Pro-*p*-isopropylanilide[32] superimposed on the organic solvent molecules observed in that region. The inhibitor binds with the trifluoroacetyl group in the S_1 subsite, the Lys residue in the S_2, Pro in S_3 and the anilide group in S_4. The functional groups of the inhibitor interact with the same protein atoms described above for the solvents. The protein interactions that are common to all ligands provide a core of hydrophobic residues whose positions are invariant from one structure to another, whereas the interactions that vary between ligands form the surrounding portions of the pocket that show plasticity in accommodating each one. Two such residues in the S_1 subsite are Gln200 and Ser203. Figure 3 shows the MSCS models of elastase clustered by least-squares superposition and focused on the primary specificity pocket S_1. The strands containing Thr221, Val224 and Thr236 cluster so tightly that it is nearly impossible to distinguish the individual models in Figure 3. This is also the case for the backbone of residues 200 through 203. The side-chains of Gln200 and Ser203, however, are in slightly different positions in each of the models and provide two key areas of plasticity on either side of the ligand within the binding pocket. A similar situation is observed in the S_4 subsite, with residues Phe223, Val103, Ala104 and Trp179 well clustered, while Arg226 and Glu65 provide plasticity.

Water binding sites

Table 1 shows that while few organic solvent molecules were found in each of the models, the number of crystallographic water molecules per structure was similar to that found for the structure solved in aqueous solution, with a range of 126 to 179 water molecules in the MSCS models. When all 11 models are superimposed over 400 unique water-binding sites are observed. It is clear that the superimposed models provide an enhanced picture of the first hydration shell surrounding the protein compared to what is normally obtained with a single model, even at high resolution. An analysis of the hydration pattern revealed by MSCS distinguishes four categories of water binding sites according to the way water molecules interact with the protein: buried, channel, surface and crystal contact.[33] The different types of water molecules can be distinguished by their average atomic *B*-factors, which correlate with disorder and occupancy of the sites. The average *B*-factor for buried water molecules is 16 $Å^2$ (similar for atoms found in the interior of the protein), for channel it is 33 $Å^2$, and for surface it is 38 $Å^2$. It is this last group of water binding sites that most often changes location from one structure to another, mediating local adaptation of the protein structure to the bulk solvent environment.[31]

Interestingly, all of the water molecules found in the active site are of the surface type. Each of the models individually has somewhere between five and 11 crystallographic water molecules in the active site. However, when all 11 structures are taken together they total 21 unique sites. Figure 4 shows these water molecules superimposed on the inhibitor discussed above. It is striking how well

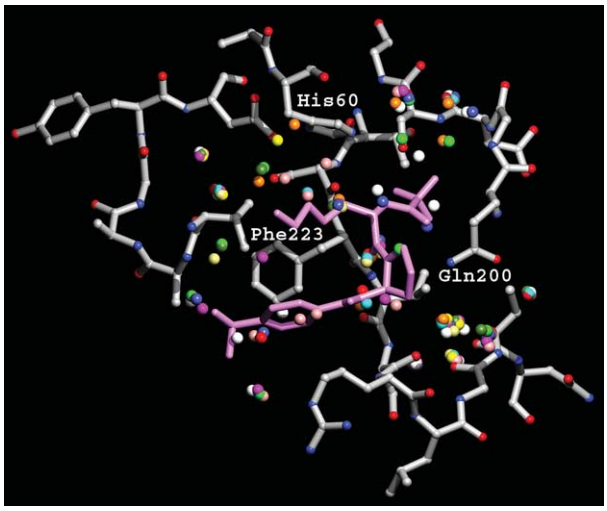

Figure 4. Crystallographic water molecules in the active site. The same region of the active site is shown as in Figures 2 and 3, with the same color code for protein atoms and organic solvent molecules. Water molecules are superimposed on the trifluoroactyl-Lys-Pro-*p*-isopropylanilide (pink). The water molecules are color-coded according to the model from which they were taken: XLINK, white; HEX, salmon; ETH, hot pink; TFE1, cyan; TFE2, orange; IPR, light green; IBZ, green; ICY, dark green; ACE, red; DMF, blue; ACN, yellow.

these crystallographically visible water molecules collectively trace the location of the inhibitor binding sites. The active site of elastase is well hydrated, with crystallographically visible water molecules sampling all of the hydrogen bonds available to ligands. Although the average position of a water molecule varies with the solvent conditions in which the protein is immersed, it is exactly this variation that increases the scope of the information available from the MSCS models. This allows a view of how closely the hydration networks form an imprint of the inhibitors and presumably of the substrate in the active site.

Discussion

The MSCS method provides an experimental approach for locating and characterizing binding sites on protein surfaces. Its power lies in the collective analysis of several (typically five to ten) superimposed crystal structures of the protein, each solved in the presence of a high concentration of a particular organic solvent, to compensate for the fact that the probe molecules bind with relatively low affinity to most sites. Elastase is a good representative of the group of extracellular enzymes that have typically been used in solvent mapping experiments to date. These proteins have well defined binding sites that do not undergo large conformational changes upon substrate binding and therefore can be thought to represent the type of protein–ligand interactions embodied by the "lock and key" model. In addition to analyzing the patterns of organic solvents clustered in the active site, the MSCS analysis presented here for elastase also focuses on the patterns of plasticity and on the hydration properties of the site, providing a more complete picture of the components involved in the binding process.

Characterizing the active site

The results presented here for elastase fully establish the MSCS method as a powerful tool to locate binding sites on enzymes. The distribution of organic solvent molecules unambiguously distinguishes the binding pockets for side-chains of the peptide substrate from all other areas on the protein surface. The organic solvent molecules cluster in four of the seven elastase pockets known to be important for binding. The side-chain binding pockets are lined by hydrophobic residues surrounded by polar groups that interact in a specific way with each type of molecule found in the pockets. In this manner a variety of functional groups can be accommodated in the pockets, with apolar regions facing inward and polar functional groups taking advantage of a subset of the available H-bonding interactions offered by protein atoms closer to the surface within the site. The rotational freedom available to a small organic molecule distinguishes it from larger, more specific ligands that must simultaneously satisfy a greater number of constraints within an extended binding site. In this sense, a diverse group of small molecules can collectively serve as true probes of the possible interactions available within a binding site. Rotational freedom is reflected in computational solvent mapping in that each type of molecule is observed to bind in multiple conformations within a site, typically having a common binding interaction for the apolar portion of the molecule and a variety of orientations for the polar groups.[26] The entire cluster, rather than individual docking conformers, is used to obtain an average free energy of interaction, thus accounting at least partially for the entropic effects of binding. In the electron density maps produced in the experimental solvent mapping by MSCS only a single conformation is observed for each solvent, although in some cases high *B*-factors for the bound organic solvent molecules can be indicative of local disorder in the structure. The fact that a major peak in the electron density is observed for a given conformer indicates that it is likely present at least 40% to 50% of the time, but does not rule out the presence of less populated conformations. This is one of the ways in which the computational solvent mapping provides complementary information to MSCS. Interestingly, while each distinct organic solvent molecule has its preferential set of interactions observed experimentally with protein atoms, computational solvent mapping results for lysozyme and thermolysin indicate that multiple sets of H-bonding

interactions often exist for a particular solvent.[25] The fairly diverse options of polar interactions within a binding site make possible the binding of small molecules of different shapes and functional groups. Conversely, it is possible to experimentally detect the diversity of polar interactions in a site by the number of distinct types of organic molecules found to cluster within it in the MSCS experiments.

The fact that clustering of different organic solvent molecules is observed by MSCS primarily at the known binding site of elastase is in striking contrast to the abundant water binding sites throughout the entire surface of the protein. At first sight it is remarkable that the organic solvent molecules, present in most cases at higher %V than water, is unable to solvate the protein in any way. In fact, the organic solvents are only able to compete with water where the protein evolved to interact with ligands or other proteins. From this perspective, the organic solvents behave much more like ligands than like solvents. It is probably this inability to compete with water outside ligand binding pockets, combined with a diverse set of available polar interactions at binding sites that makes binding of organic solvents so distinctly prominent in the active sites of enzymes. It is their behavior as ligands, rather than solvents, that makes small organic molecules powerful locators of binding sites on protein surfaces.

Protein plasticity

It is well known that there is a great degree of plasticity in the binding sites of proteins.[7,40] Plasticity does not necessarily involve large rearrangement of main-chain or side-chain conformation, but can also reflect small movements of one or more residues in response to ligand binding or to a residue mutation that changes the local environment at a given site. Plasticity in response to ligand binding has been observed in many enzymes and was described for elastase in particular as the "subtle induced fit of the active site as a result of ligand binding".[34] By superimposing the protein models taken from MSCS as shown in Figure 3, one can use experimental solvent mapping to delineate areas of plasticity in the active site. For the S_1 subsite in elastase, used here as the illustrative example, these areas include the side-chains of residues Gln200 and Ser203, which undergo subtle changes in position in response to the various bound organic solvent molecules. This kind of information can be extremely important in ligand docking and in the development of ligand design strategies. The lack of this information has been shown to be a limiting factor in computational solvent mapping, which is often more successful in predicting consensus binding sites when the conformation found in the complex (rather than the apo form of the protein) is used as a template for the calculations.[26] It is often costly to include protein flexibility in computational studies of ligand docking in binding sites, especially when a large number of compounds need to be tested. This expense could be greatly reduced if one could use the information from MSCS to allow only selected parts of the structure to be dynamic during docking or ligand design.

Plasticity can also be observed at the level of the entire protein structure. The MSCS provide a view of the protein in very different environments. Through observing how the protein structure adapts to the global changes, it becomes clear which areas of the structure are malleable and contribute to the changes in conformation that optimize structure given the properties of the solvent and the constraints of the crystal environment. The bulk effects of organic solvents on protein structure have been reviewed[14] and a study of apolar solvents on the conformation of the switch II region of Ras[41] provides a fundamental framework for explaining the local conformational changes seen in the elastase residues 24–27 and 122–123, where polar interactions between protein atoms are optimized in solvents of lower dielectric constants, while in the higher dielectric environment polar groups are often extended toward the solvent. Remarkably, with the exception of the region involving residues 24–27/122–123, the backbone of elastase is largely unperturbed by the changes in solvent environment. Rather, it is the first hydration shell that is the primary mediator between the bulk solvent and the protein, changing positions with changes in the solvent and probably helping to maintain the integrity of the native state.[31] Thus, the MSCS method provides a view of plasticity in the water structure as well as in the protein.

It has been observed through database analysis that enzyme complexes tend to involve anchor residues that interact with structurally constrained portions of the binding partner.[3] These constrained or anchor areas are thought to be involved in the first step of recognition, which is hypothesized to be followed by an induced fit process involving surrounding residues to form the high-affinity complex. The pattern of organic solvents in the binding site of elastase can be viewed as depicting binding sites for anchor residues at a few locations in the more extended active site. The pockets mapped experimentally have areas that do not change with the organic solvent environment surrounded by plasticity regions as exemplified in Figure 3. The probe molecules cluster in the active site in well-defined hot spots. Rather than forming a continuous surface, the hot spots are distributed throughout the binding site in pockets lined by hydrophobic residues. This is also the pattern observed in database analysis of protein complexes, although the hot spot residues are not always hydrophobic in nature.[4] The important point here is that patterns essential to binding site obtained through database analysis are reflected in the structures of elastase in the presence of organic solvents. The MSCS method is able to depict both hot spot residues and the plasticity areas that compose the active site of elastase, providing a consistent picture of some essential binding site properties.

Hydration

The 21 active site water molecules observed collectively in the MSCS experiment form H-bonding interactions with virtually all available polar groups in the active site. This includes the walls and openings to the pockets as well as the β-strand that forms a classic antiparallel interaction with peptide substrates or a parallel interaction with some inhibitors in the active site.[32] The pattern of crystallographic water molecules delineating the polar groups in the active site mimics the shape of the bound inhibitors. This leads to the idea that while the organic solvents may identify areas of hydrophobic hot spots on the surfaces of enzymes, the water molecule pattern surrounding the hot spots may provide clues to the general shape of the active site that can be occupied by a ligand. If so, these two components of MSCS are indeed highly complementary. All of the crystallographic water molecules observed in the active site of elastase must be released upon complex formation. This would be expected to contribute an entropically favorable component to binding.

Conclusions

A complete MSCS analysis of a protein surface takes into account patterns of organic solvent molecules indicating the location of hot spots, the areas of plasticity observed when superimposing the protein models and the distribution of crystallographically visible water molecules. The results presented here provide a detailed map of the active site of elastase containing all of these components and consistent with binding site properties deciphered form database analysis of hundreds of proteins. Because the properties being probed have been shown to be component features of binding sites, MSCS is poised as a powerful method for analyzing the active sites of enzymes in general. Hot spots, plasticity and certain patterns of hydration converge in active sites. The MSCS method probes these features simultaneously, and therefore can serve not only to locate the binding sites, but also to offer a fairly complete characterization of their essential properties. The diversity of polar and hydrophobic interactions within a protein binding site can be mapped in context with the plasticity of the active site and the malleability of the water structure. This information is critical in filling the missing links currently existing in computational ligand design methods.

Materials and Methods

Crystal growth, cross-linking and solvent soaks

Porcine pancreatic elastase was purchased from either Calbiochem, Inc. or Worthington Biochemicals and used without further purification. Crystals were grown and cross-linked with gluteraldehyde as described.[17] The cross-linking buffer (100 mM sodium sulfate, 100 mM sodium phosphate, pH 7.5) was slowly exchanged with distilled water in a stepwise fashion. At the end of this process the cross-linked elastase crystals were immersed in 200 μl of distilled water. Cross-linked crystals were transferred to different organic solvents and the solvent concentration decreased 5% from neat by the addition of water, to a maximum concentration that was not detrimental to the diffraction of X-rays. A new crystal was used for each test. Once the maximum viable concentration of a particular solvent was determined, the distilled water containing the cross-linked elastase crystals was exchanged with the organic solvent solution in a stepwise fashion by repeated removal of 20 μl aliquots followed by addition of the same amount of the final solution every 5 to 10 min. The final step in the process involved the transfer of the crystals to a fresh solution containing the desired concentration of the organic solvent. The one exception to the general protocol above was the TFE1 structure, for which the cross-linking phosphate buffer solution containing the elastase crystal was exchanged directly with a 40% trifluoroethanol solution, rather than being first transferred to distilled and de-ionized water. The solvents for which data were collected are shown in Table 1 and described in Results.

Data collection

Crystals were mounted in quartz capillaries for X-ray diffraction data collection. Data were collected to a maximum resolution ranging from 2.2 Å to 1.8 Å on one of three detector/generator sources: (1) R-Axis II phosphoimaging plate mounted on a Rigaku RU200 rotating anode generator operating at 50 kV and 100 mA; (2) Siemens X100-A area detector mounted on an Eliot GX-6 rotating anode operating at 30 kV and 30 mA; (3) Siemens multi-wire area detector with Argonne "Mad" Interface mounted on a Rigaku RU200 generator operating at 50 kV and 100 mA. The data were reduced using the program Process,[42] XDS,[43] Xengen[44] or Denzo.[45] The data collection temperature and equipment, as well as the program used for data processing, are given in Table 2 for each of the structures used in the present MSCS analysis.

Structure refinement

For all structures the program XPLOR[46] was used for initial phase calculation and least-squares positional refinement. Initial phases were calculated from the coordinates of elastase solved from crystals grown in aqueous solution, with PDB codes as specified in Table 2. All water molecules, counter ions and inhibitor atoms were removed from the initial models. The refinement strategy was to first make adjustments to protein atoms, then to add water molecules that appeared simultaneously in an electron density map with coefficients $2F_o - F_c$, contoured at the 1σ level, and in an electron density map with coefficients $F_o - F_c$ contoured at the 3σ level. A Silicon Graphics workstation running the program O[47] was used to manually rebuild the protein. In general, water molecules were only added to the model once the *R*-factor dropped to below 22%. Calcium and sulfate ions, as well as the first round of water molecules were selected from the coordinates of native elastase solved in aqueous solution (PDB code 1ELA). Additional water molecules were added manually by visual inspection using the program O. Only then were the organic

solvent molecules included in the model. The alcohol moiety in ethanol and isopropanol and the N atom in acetonitrile were positioned based on the chemical environment of the protein site, in order to optimize hydrogen-bonding interactions. With the exception of DMF and TFE, the topology and parameter files used to refine the organic solvent positions were based on similar functional groups already present in the XPLOR protein topology and parameter files.[46] The parameters for DMF used with XPLOR were developed by Jorgensen[48] and the parameters for TFE were taken from CHARMM22.[49]

As a final check, all organic solvent molecules were removed from each model and a slow-cool simulated annealing protocol was performed using CNS[50] with 10% of the data set aside for cross-validation.[51] The starting temperature for the simulated annealing calculations was 2500 K and the final temperature was set to 300 K, with a 25 K drop per annealing cycle. The resulting F_o-F_c (3σ) and $2F_o-F_c$ (1σ) electron density maps were used to check the protein structure and the water molecule positions, and to validate the location of the organic solvent molecules. Only the probe molecules for which there was clear difference density in the omit maps were kept. One or two rounds of positional and *B*-factor refinement with the organic solvent molecules included followed, with the entire model being checked and adjusted if necessary. The final *R*-factors and corresponding *R*-free values are shown in Table 1.

Coordinate numbering scheme

For the purpose of analysis, the DMF model was chosen as a reference onto which all other models were superimposed using the least-squares superposition function in the program O.[47] Every organic solvent molecule bound at a particular site on the protein has the same number in all of the coordinate sets. The numbering system for the probe molecules and how it relates to the pockets in the active site are described in Results.

Similarly, equivalent water molecules have the same residue numbers in all structures. Water molecules retained from the native elastase structure solved in aqueous solution (PDB code 3EST) have numbers 301–345. This includes 21 of the 23 buried water molecules.[33] The remaining two buried sites are numbered 406 and 418, respectively. Ten channel water molecules (305, 306, 308, 311, 316, 318, 319, 323, 326 and 332) and 14 surface type water molecules (325, 327, 329, 333, 335, 336, 337, 339 and 340–345) were also retained from the original coordinates and therefore fall into this numbering group. The active site water molecules, all of them classified as surface, are numbered 351–371. Channel water molecules that did not originate from the 3EST coordinate set are numbered 400–421, with the exception of 406 and 418, which are buried, as mentioned above. Crystal contact water molecules, those that are within 4 Å of a symmetry-related protein atom, are numbered 500–558. All of them have the characteristics of the surface water type with the exception of 500, 503, 549 and 554, which are of the buried type (buried between two protein molecules in the crystal). Surface water molecules that did not originate from the 3EST coordinate set, are away from crystal contacts and do not bind in the active site are numbered 600–892.

Protein Data Bank accession codes

The coordinates and structure factors for nine models of elastase soaked in organic solvents have been deposited in the Protein Data Bank. The accession codes are given using the abbreviation for each model as specified in the text and Tables: HEX PDB code 2FOE; ETH PDB code 2FOD; TFE1 PDB code 2FOG; TFE2 PDB code 2FOH; IPR PDB code 2FOF; IBZ PDB code 2FOA; ICY PDB code 2FOB; ACE PDB code 2FO9; DMF PDB code 2FOC.

Acknowledgements

We thank Karen Allen for help with the original soaks. We are grateful to Diana Griffith, Dmitri Ivanov, Cheryl Kreinbring and Marty Stanton for technical help during various stages of this project. Thanks also to Paul Swartz for redoing the Figures for publication. This research was supported, in part, by the NSF PECASE award to C.M. at NC State University, MCB-0237297.

References

1. Vajda, S. & Camacho, C. J. (2004). Protein–protein docking: is the glass half-full or half-empty? *Trends Biotechnol.* **22**, 110–116.
2. Selzer, T., Albeck, S. & Schreiber, G. (2000). Rational design of faster associating and tighter binding protein complexes. *Nature Struct. Biol.* **7**, 537–541.
3. Rajamani, D., Thiel, S., Vajda, S. & Camacho, C. J. (2004). Anchor residues in protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **101**, 11287–11292.
4. Keskin, O., Ma, B. & Nussinov, R. (2005). Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **345**, 1281–1294.
5. Ringe, D. (1995). What makes a binding site a binding site? *Curr. Opin. Struct. Biol.* **5**, 825–829.
6. Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
7. Najmanovich, R., Kuttner, J., Sobolev, V. & Edelman, M. (2000). Side chain flexibility in proteins upon ligand binding. *Proteins: Struct. Funct. Genet.* **39**, 261–268.
8. Ma, B., Wolfson, H. J. & Nussinov, R. (2001). Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr. Opin. Struct. Biol.* **11**, 364–369.
9. Papoian, G. A., Ulander, J. & Wolynes, P. G. (2003). Role of water mediated interactions in protein–protein recognition landscapes. *J. Am. Chem. Soc.* **125**, 9170–9178.
10. Reichmann, D., Rahat, O., Albeck, S., Meged, R., Dym, O. & Schreiber, G. (2005). The modular architecture of protein–protein binding interfaces. *Proc. Natl Acad. Sci. USA*, **102**, 57–62.
11. Kortvelyesi, T., Dennis, S., Silberstein, M., Brown, L., 3rd & Vajda, S. (2003). Algorithms for computational solvent mapping of proteins. *Proteins: Struct. Funct. Genet.* **51**, 340–351.
12. Cummings, M. D., DesJarlais, R. L., Gibbs, A. C., Mohan, V. & Jaeger, E. P. (2005). Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **48**, 962–976.
13. Mattos, C. & Ringe, D. (1996). Locating and characterizing binding sites on proteins. *Nature Biotechnol.* **14**, 595–599.

14. Mattos, C. & Ringe, D. (2001). Proteins in organic solvents. *Curr. Opin. Struct. Biol.* **11**, 761–764.
15. Fitzpatrick, P. A., Ringe, D. & Klibanov, A. M. (1994). X-ray crystal structure of cross-linked subtilisin Carlsberg in water *vs.* acetonitrile. *Biochem. Biophys. Res. Commun.* **198**, 675–681.
16. Yennawar, N. H., Yennawar, H. P. & Farber, G. K. (1994). X-ray crystal structure of gamma-chymotrypsin in hexane. *Biochemistry*, **33**, 7326–7336.
17. Allen, K. N., Bellamacina, C. R., Ding, X., Jeffery, C. J., Mattos, C., Petsko, G. A. & Ringe, D. (1996). An experimental approach to mapping the binding surfaces of crystalline proteins. *J. Phys. Chem.* **100**, 2605–2611.
18. Schmitke, J. L., Stern, L. J. & Klibanov, A. M. (1997). The crystal structure of subtilisin Carlsberg in anhydrous dioxane and its comparison with those in water and acetonitrile. *Proc. Natl Acad. Sci. USA*, **94**, 4250–4255.
19. Wang, T. & Zhou, J. (1998). 3DFS: a new 3D flexible searching system for use in drug design. *J. Chem. Inf. Comput. Sci.* **38**, 71–77.
20. English, A. C., Done, S. H., Caves, L. S., Groom, C. R. & Hubbard, R. E. (1999). Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins: Struct. Funct. Genet.* **37**, 628–640.
21. English, A. C., Groom, C. R. & Hubbard, R. E. (2001). Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.* **14**, 47–59.
22. Miranker, A. & Karplus, M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method (MCSS). *Proteins: Struct. Funct. Genet.* **11**, 29–34.
23. Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849–857.
24. Smith, G. R. & Sternberg, M. J. E. (2002). Prediction of protein–protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12**, 28–35.
25. Dennis, S., Kortvelyesi, T. & Vajda, S. (2002). Computational mapping identifies the binding sites of organic solvents on proteins. *Proc. Natl Acad. Sci. USA*, **99**, 4290–4295.
26. Silberstein, M., Dennis, S., Brown, L., Kortvelyesi, T., Clodfelter, K. & Vajda, S. (2003). Identification of substrate binding sites in enzymes by computational solvent mapping. *J. Mol. Biol.* **332**, 1095–1113.
27. Eisen, M. B., Wiley, D. C., Karplus, M. & Hubbard, R. E. (1994). HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins: Struct. Funct. Genet.* **19**, 199–221.
28. Verlinde, C. L., Rudenko, G. & Hol, W. G. (1992). In search of new lead compounds for trypanosomiasis drug design: a protein structure-based linked-fragment approach. *J. Comput. Aided Mol. Des.* **6**, 131–147.
29. Huc, I. & Lehn, J. M. (1997). Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly [published erratum appears in Proc Natl Acad Sci U S A 1997 Jul 22;94(15):8272]. *Proc. Natl Acad. Sci. USA*, **94**, 2106–2110.
30. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science*, **274**, 1531–1534.
31. Mattos, C. (2002). Protein–water interactions in a dynamic world. *Trends Biochem. Sci.* **27**, 203208.
32. Mattos, C., Giammona, D. A., Petsko, G. A. & Ringe, D. (1995). Structural analysis of the active site of porcine pancreatic elastase based on the X-ray crystal structures of complexes with trifluoroacetyl-dipeptide-anilide inhibitors. *Biochemistry*, **34**, 3193–3203.
33. Mattos, C. & Ringe, D. (2001). Solvent structure. In *International Tables for Crystallography* (Rossman, M. G. & Arnold, E., eds), vol. F, pp. 623–640, Kluwer Academic Publishers, Dordrecht.
34. Meyer, E., Cole, G., Radhakrishnan, R. & Epp, O. (1988). Structure of native porcine pancreatic elastase at 1.65 Å resolutions. *Acta Crystallog. sect. B*, **44**, 26–38.
35. Atlas, D. (1975). The active site of porcine elastase. *J. Mol. Biol.* **93**, 39–53.
36. Shotton, D. M., White, N. J. & Watson, H. C. (1972). Conformational changes and inhibitor binding at the active site of elastase. *Cold Spring Harbor Symp. Quant. Biol.* **36**, 91–105.
37. Meyer, E. F., Jr, Radhakrishnan, R., Cole, G. M. & Presta, L. G. (1986). Structure of the product complex of acetyl-Ala-Pro-Ala with porcine pancreatic elastase at 1.65 Å resolution. *J. Mol. Biol.* **189**, 533–539.
38. Takahashi, L. H., Radhakrishnan, R., Rosenfield, R. E., Jr, Meyer, E. F., Jr, Trainor, D. A. & Stein, M. (1988). X-ray diffraction analysis of the inhibition of porcine pancreatic elastase by a peptidyl trifluoromethylketone. *J. Mol. Biol.* **201**, 423–428.
39. Takahashi, L. H., Radhakrishnan, R., Rosenfield, R. E., Jr & Meyer, E. F., Jr (1989). Crystallographic analysis of the inhibition of porcine pancreatic elastase by a peptidyl boronic acid: structure of a reaction intermediate. *Biochemistry*, **28**, 7610–7617.
40. Atwell, S., Ultsch, M., De Vos, A. M. & Wells, J. A. (1997). Structural plasticity in a remodeled protein–protein interface. *Science*, **278**, 1125–1128.
41. Buhrman, G., de Serrano, V. & Mattos, C. (2003). Organic solvents order the dynamic switch II in Ras crystals. *Structure (Camb)*, **11**, 747–751.
42. Brunger, A. T. (1988). Crystallographic refinement by simulated annealing. Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816.
43. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallog.* **26**, 795–800.
44. Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. & Selemme, F. R. (1987). Use of an imaging proportional counter in macromolecular crystallography. *J. Appl. Crystallog.* **20**, 383–387.
45. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology: Macromolecular Crystallography, Part A* (Carter, C. W. & Sweet, R. M., eds), vol. 276, pp. 307–326, Academic Press, New York.
46. Brunger, A. T. (1992). *X-PLOR: A System for X-ray Crystallography and NMR*, Yale University Press, Hew Haven, CT.
47. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
48. Jorgensen, W. L. & Swenson, C. J. (1985). Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides. *J. Am. Chem. Soc.* **107**, 569–578.

49. MacKerell, A. D., Jr, Bashford, D., Bellott, M., Dunbrack, R. L., Jr, Evanseck, J., Field, M. J. *et al*. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
50. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., DeLano, W. L. *et al*. (1998). Crystallographic & NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
51. Kleywegt, G. J. & Brunger, A. T. (1996). Checking your imagination: applications of the free *R* value. *Structure*, **4**, 897–904.
52. Kraulis, P. G. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Edited by R. Huber

(Received 3 September 2005; received in revised form 21 December 2005; accepted 5 January 2006)