Water molecules participate in proteinase-inhibitor interactions: Crystal structures of Leu^{18}, Ala^{18}, and Gly^{18} variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B

KUI HUANG,[1] WUYUAN LU,[2] STEPHEN ANDERSON,[3] MICHAEL LASKOWSKI, JR.,[2] AND MICHAEL N.G. JAMES[1]

[1] Medical Research Council of Canada Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada
[2] Department of Chemistry, Purdue University, West Lafayette, Indiana 47907-1393
[3] Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, New Jersey 08854

Abstract

Crystal structures of the complexes of *Streptomyces griseus* proteinase B (SGPB) with three P$_1$ variants of turkey ovomucoid inhibitor third domain (OMTKY3), Leu^{18}, Ala^{18}, and Gly^{18}, have been determined and refined to high resolution. Comparisons among these structures and of each with native, uncomplexed SGPB reveal that each complex features a unique solvent structure in the S$_1$ binding pocket. The number and relative positions of water molecules bound in the S$_1$ binding pocket vary according to the size of the side chain of the P$_1$ residue. Water molecules in the S$_1$ binding pocket of SGPB are redistributed in response to the complex formation, probably to optimize hydrogen bonds between the enzyme and the inhibitor. There are extensive water-mediated hydrogen bonds in the interfaces of the complexes. In all complexes, Asn 36 of OMTKY3 participates in forming hydrogen bonds, via water molecules, with residues lining the S$_1$ binding pocket of SGPB. For a homologous series of aliphatic straight side chains, Gly^{18}, Ala^{18}, Abu^{18}, Ape^{18}, Ahx^{18}, and Ahp^{18} variants, the binding free energy is a linear function of the hydrophobic surface area buried in the interface of the corresponding complexes. The resulting constant of proportionality is 34.1 cal mol^{-1} Å^{-2}. These structures confirm that the binding of OMTKY3 to the preformed S$_1$ pocket in SGPB involves no substantial structural disturbances that commonly occur in the site-directed mutagenesis studies of interior residues in other proteins, thus providing one of the most reliable assessments of the contribution of the hydrophobic effect to protein–complex stability.

Keywords: association constant; crystal structure; hydrophobic effect; ovomucoid third domain; serine proteinase; water-mediated hydrogen bond

In recent years, many proteins from diverse sources have been shown to be inhibitors of a variety of proteinases. Thus far, 20 or so families of protein–proteinase inhibitors have been identified (reviewed by Bode & Huber, 1992). Proteinase inhibitors are classified into different families on the basis of their target enzymes, their inactivation mechanisms, their primary sequences, and their tertiary structures. Protein inhibitors of the serine proteinases exhibit a standard mechanism of inhibition (with the serpin and hirudin families as the only known exceptions), despite their unrelated sequences and three-dimensional structures (Laskowski & Kato, 1980). These so-called canonical inhibitors bind to their cognate enzymes in the same man-

ner as a good substrate, but are cleaved extremely slowly. The inhibitors resemble good substrates in two aspects: they display extremely tight binding (with K_a up to 10^{12} M^{-1}) and have very strict specificity (as measured by k_{cat}/K_m) for their cognate enzymes. X-ray structures of these substrate-like inhibitors have revealed that the reactive-site loops of the inhibitors are chemically and geometrically complementary to the active sites of their cognate enzymes and that the complexes are noncovalent (reviewed by Read & James, 1986). These intrinsic structural features of protein inhibitors confer them with exceptionally high binding affinities for their cognate enzymes.

The primary specificity determinant of a given inhibitor is the residue at the P_1 position, the residue contributing the carbonyl portion to the reactive-site peptide bond (nomenclature of Schechter & Berger, 1967). Substitution of the P_1 residue results in a dramatic change in K_a, by six orders of magnitude in some cases, corresponding to free energies of binding of up to 8 kcal/mol, as displayed by the variants of the ovomucoid inhibitor third domain from turkey (Bigler et al., 1993; Lu, 1994). OMTKY3 is one of the best-studied protein inhibitors. Native OMTKY3 has 56 amino acids; the peptide bond between Leu 18 and Glu 19 is the scissile bond and thus is termed the reactive-site bond. OMTKY3 is a potent inhibitor of chymotrypsin ($K_a = 1.9 \times 10^{11}$ M^{-1}). When the P_1 residue, Leu 18, of OMTKY3 is replaced by lysine, the resulting variant becomes a strong inhibitor of trypsin. Similarly, the Glu[18] variant is a strong inhibitor of the Glu-specific *Streptomyces griseus* proteinase (Komiyama et al., 1991). The Arg[15]-Lys[17]-Arg[18] variant of OMTKY3 is a good inhibitor of human furin (Lu et al., 1993).

A wealth of kinetic and thermodynamic data makes OMTKY3 an attractive model to study the inhibitory mechanism used by protein inhibitors. An overexpression system for OMTKY3 has been established (Lu et al., 1993). This facilitates the generation and purification of various mutants of interest. So far, the P_1 residue, which is Leu 18 in the naturally occurring OMTKY3, has been substituted by the other 19 coded amino acids and 5 noncoded amino acids (Hse, Abu, Ape, Ahx, and Ahp). The K_a values of all these recombinant P_1 variants were measured for six proteinases: SGPA, SGPB, chymotrypsin, porcine pancreatic elastase, human leukocyte elastase, and subtilisin (Bigler et al., 1993; Lu, 1994). These data have established an extensive characterization of the P_1 preference of the six well-studied enzymes. Thorough analysis and correlation of these valuable data to detailed X-ray structures will greatly extend our knowledge about proteinase-inhibitor interactions and subsequently will benefit inhibitor design.

SGPB is a chymotrypsin-like proteinase with respect to its substrate specificity. Measurements of the binding equilibrium constants have shown that leucine is most favored at the P_1 position of OMTKY3, with a K_a of 5.6×10^{10} M^{-1} (Bigler et al., 1993; Lu, 1994). Proline is least favored, with a K_a of 3.6×10^4 M^{-1}. In addition, the P_1 position prefers amino acids with straight and γ-branched side chains over those of β-branched of an equal number of atoms. The S_1 binding pocket (the preformed cavity in the enzyme into which the P_1 residue inserts) of SGPB is charge-phobic, as the introduction of positively or negatively charged residues at the P_1 position shows deleterious effects on K_a. Interestingly, valine and lysine have completely different polarities and volumes, but display comparable K_a values and are essentially "isofunctional" as the P_1 residues of OMTKY3. Although modeling may provide insights for some of the results, most interpretations are far from quantitative and convincing.

Crystal structures of OMTKY3 have been determined in complexes with SGPB (Read et al., 1983), α-chymotrypsin (Fujinaga et al., 1987), and human leukocyte elastase (Bode et al., 1986). The crystal structures of two closely related inhibitors, ovomucoid inhibitor third domains from Japanese quail and from silver pheasant, have been solved both in free, intact form (Weber et al., 1981; Papamokos et al., 1982; Bode et al., 1985) and in the scissile bond-cleaved form (Musil et al., 1991). We have undertaken crystal structure determinations of all the recombinant P_1 variants of OMTKY3 (excluding noncoded amino acid variants) in complex with SGPB. The scope of our project is to provide detailed structural data to assist in the explanation of the kinetic and thermodynamic data. One of our ultimate objectives is to be able to predict the free energy of binding for a given inhibitor and proteinase by computational simulation, and thereafter to evaluate the contribution of each individual residue involved in the complex formation. Accurate prediction of binding constants demands correct modeling of the conformational change induced by the protein–ligand association and realistic estimation of the interaction energy in that system (reviewed by Kollman, 1994). Therefore, precise information concerning possible conformational changes and the solvation state of interacting residues is required and thus far can only be obtained reliably from detailed X-ray crystallographic studies.

In this study, we report the crystal structures of three OMTKY3 variants, Leu[18], Ala[18], and Gly[18], in complex with SGPB. These variants span a large range of K_a values (4.7 × 10^3-fold change) in their association with SGPB (5.6 × 10^{10} M^{-1}, 3.6 × 10^8 M^{-1}, and 1.2 × 10^7 M^{-1}, respectively). In addition to the substitutions at the P_1 position, these recombinant variants differ from native OMTKY3 in that the first five residues of the N-terminus have been removed. Kinetic measurements have shown that the deletion of the first five residues has no effect on K_a (Wieczorek et al., 1987). The previously reported structure of native OMTKY3 in complex with SGPB also showed that the first six residues could not be located reliably in the electron density map, presumably because they were disordered or highly mobile (Read et al., 1983). We have also redetermined the structure of the recombinant Leu[18] variant (residues 6–56) to ensure that small structural differences in the active site regions displayed by the different variants were not due to the truncation of the N-terminus. We show that there are small but important conformational changes in the active site of SGPB upon the association with the OMTKY3 variants. In adjusting to the size of the P_1 residue, water molecules bound in the S_1 pocket are redistributed, presumably to optimize the enzyme–inhibitor interactions. Water-mediated hydrogen bonding networks, as observed in the complexes of DNA with DNA-binding proteins (Kim & Burley, 1994; Shakked et al., 1994), peptides with MHC class I molecules (Madden et al., 1992), and peptides with oligopeptide binding protein OppA (Tame et al., 1994), are also present in our enzyme–inhibitor complexes.

Results

Quality of the complex structures

The crystallographic R-factors and the final stereochemical parameters of the current models are listed in Table 1. The final R-factors for the complexes of SGPB with the Leu[18], Ala[18],

Table 1. *Statistics of structural refinement*

Refinement	Leu[18] complex	Ala[18] complex	Gly[18] complex
No. of reflections used	17,165	30,412	13,120
Resolution range (Å)	20.0–1.8	20.0–1.4	20.0–1.9
R-factor[a] (%)	13.9	17.1	13.2
No. of protein atoms	1,697	1,694	1,693
No. of water atoms	169	182	168
No. of phosphate anion	1	1	1
RMS deviation from ideal stereochemistry			
Bond distance (Å)	0.022	0.023	0.023
Bond angle (°)	2.4	2.4	2.5
Planar groups (Å)	0.022	0.022	0.021

[a] R-factor $= \Sigma_{hkl} \|F_o| - |F_c\| / \Sigma_{hkl} |F_o|$.

and Gly[18] variants are 13.9% (1.8 Å), 17.1% (1.4 Å), and 13.2% (1.9 Å), respectively. Figure 1 is a representative electron density map at 1.4 Å resolution, showing the reactive-site loop and water molecules bound in the S₁ binding pocket in the complex of SGPB and the Ala[18] OMTKY3 variant.

N-terminal truncation of OMTKY3 has no effect on the reactive-site conformation

All the 51 residues of the OMTKY3 variants can be located in the electron density map. Figure 2 shows the superposition of the main-chain atoms of native OMTKY3 (residues 1–56, Read et al., 1983) with the Leu[18] variant of OMTKY3 (residues 6–56). The RMS difference (RMSD) for the 200 common main-chain atoms (residues 7–56) is 0.18 Å. There are significant shifts (greater than 2 times the RMSD) in the region of the N-terminus (Asp 7–Pro 12) and the neighboring residues of the α-helix (Cys 38–Glu 43) disulfide-bridged to the N-terminus. The largest shift is 0.8 Å at the N atom of residue Asp 7. Elsewhere there are virtually no noticeable changes in other regions, including the reactive-site loop. The three variants used in this study show better agreement with one another. The RMSD of the 200 back-

bone atoms is 0.15 Å for Leu[18] versus Ala[18] variants (the largest difference is 0.52 Å at Pro 14 O), 0.14 Å for Leu[18] versus Gly[18] variants (the largest difference is 0.50 Å at Pro 14 O), and 0.12 Å for Ala[18] versus Gly[18] variants (the largest difference is 0.27 Å at Ser 9 O). These RMSDs are close to the experimental positional errors of crystallographically determined structures at the corresponding resolutions. It is interesting to note that, although the first six residues of the N-terminus are disordered in the previously reported structure, their presence, however, has an impact on the conformation of the inhibitor. Nevertheless, the conformations of the reactive-site loops are not affected by the N-terminal truncations, which is consistent with the fact that the K_a's for native OMTKY3 and the truncated Leu[18] variant of OMTKY3 are virtually identical (Wieczorek et al., 1987).

Conformational changes of one substrate binding loop of SGPB

Figure 3 is a main-chain atom superposition of the uncomplexed SGPB (crystallized at its enzymatically optimal pH, pH 7.4, in a cubic crystal form; refined at 1.8 Å resolution) (Codding et al., 1974; H. Blanchard & M.N.G. James, in prep.), with SGPB in the complexed state with the Leu[18] variant of OMTKY3. The RMSD for the 740 main-chain atoms of SGPB is 0.27 Å. The most significant difference between the two structures is the shift of the residues forming one side of the S₁ specificity pocket (Gly 215–Ser 217). This segment of polypeptide makes only a few direct contacts with the P₁ residue (the closest contact is 3.8 Å between Gly 215 C$^\alpha$ of SGPB and Leu 18 C$^{\delta 1}$ of OMTKY3). In response to the binding of the inhibitor, this loop moves from the native position in a concerted manner closer to the side chain of Leu 18 of OMTKY3 as if drawn by it. The C atom of Gly 216 of SGPB makes the largest shift, 0.7 Å toward the side chain of Leu 18 of OMTKY3.

It should be noted that this conformational change in SGPB takes place only in the Leu[18] variant complex. There are two SGPB molecules per asymmetric unit in the uncomplexed SGPB; in them the conformations of the S₁ binding loops (Gly 215–Ser 217) are virtually identical. In the complexes of SGPB with the Ala[18] and Gly[18] variants, loop Gly 215–Ser 217 remains in the same position as in the uncomplexed SGPB.

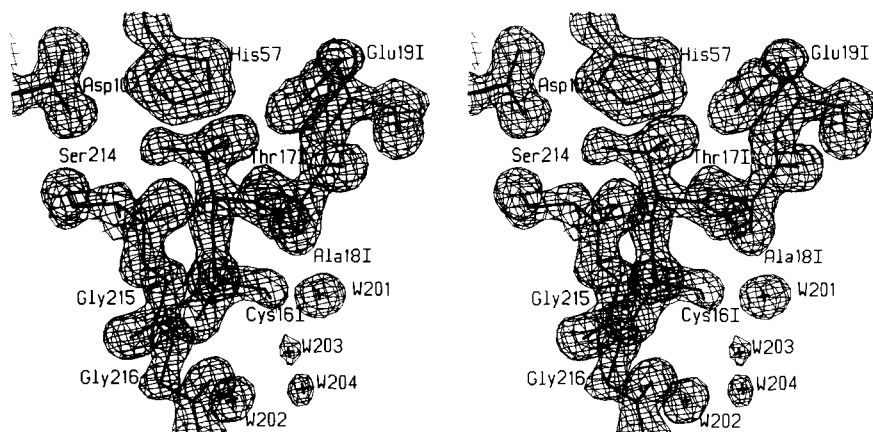**Fig. 1.** Electron density map showing part of the active site of the complex of SGPB with the Ala[18] variant of OMTKY3. The map shows the catalytic His 57 of SGPB, the residues flanking the scissile bond of OMTKY3, and four water molecules bound in the S₁ binding pocket. The map is computed with coefficients $(2m|F_o| - d|F_c|)\exp(i\alpha_{cal})$ (Read, 1986) and calculated phases and is contoured at the 1.5σ level.
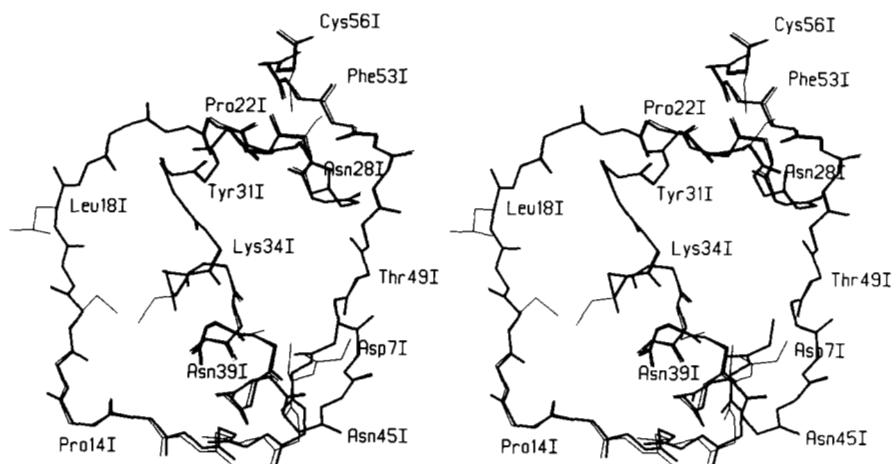
Fig. 2. Superposition of the Leu[18] variant of OMTKY3 (thick lines) with native OMTKY3 (thin lines) (Read et al., 1983). Only main-chain atoms and disulfide bridges are shown. The RMS deviation for the 200 common main-chain atoms is 0.18 Å. There is no noticeable difference in the reactive-site loop.

A phosphate binding site in the interaction region

A difference map (coefficients $(|F_o| - |F_c|)\exp(i\alpha_{cal})$), where $|F_o|$ and $|F_c|$ are, respectively, the observed and calculated structure factor amplitudes, and α_{cal} the calculated phases) unambiguously indicates that a phosphate binding site is present (Fig. 4). It is identified in all the three complexes, in the peripheral region of the complex interface, at a position equivalent to W32 in the previously determined structure (Read et al., 1983). The phosphate is refined with a half occupancy and B-factors around 40 Å^2. Figure 4 shows the coordination of the phosphate anion. The phosphoryl oxygen atoms form hydrogen bonds to O^η of Tyr 32, N^ϵ of Arg 42 of SGPB, and to water molecules W146 and W147. It should be noted that the observed phosphate binding site may merely be an artifact of the crystallization conditions (50 mM of phosphate buffer was used). Phosphate was not present in the measurement of the binding constants (Bigler et al., 1993; Lu, 1994). The presence of a partially occupied phosphate anion in the complex should not have a significant effect on the strength of the binding affinity. In ad-

dition, its presence in these structural studies should have little if any effect on the active-site conformations.

Comparison of common water molecules in the complexes

In order to compare solvent structures in the S_1 pocket of different complexes, it is necessary to set up a distance criterion within which water molecules from different crystal structures are considered equivalent. For this purpose, crystallographically determined water molecules were placed into three groups on the basis of their refined thermal factors (B-factors). The three B-factor shells include all water molecules, those with B-factors lower than 1.5 $B_{average}$, and those with B-factors lower than $B_{average}$, respectively. Water molecules within the same B-factor shell were then compared in a pairwise manner. Table 2 gives the RMSDs of the superpositions. The average B-factors are approximately 13.5 Å^2 for main-chain atoms, 17.5 Å^2 for side-chain atoms, and 37 Å^2 for solvent molecules of the three complexes, respectively.
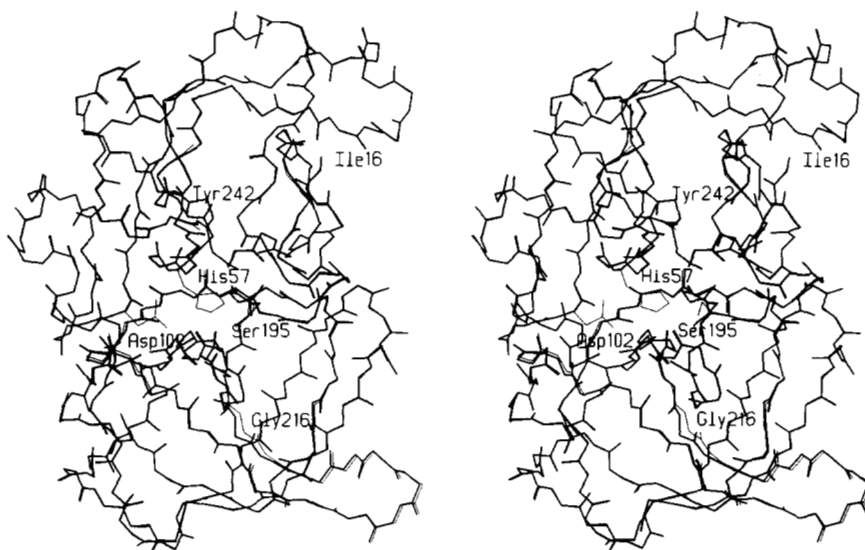

Fig. 3. Superposition of the uncomplexed SGPB (thin lines) (H. Blanchard & M.N.G. James, in prep.) with SGPB in complex with the Leu[18] variant (thick lines). Residue Gly 216 is labeled to highlight the conformational change of the segment Gly 215–Ser 217. Gly 216 C$^\alpha$ of SGPB moves 0.7 Å toward the side chain of Leu 18 of OMTKY3.
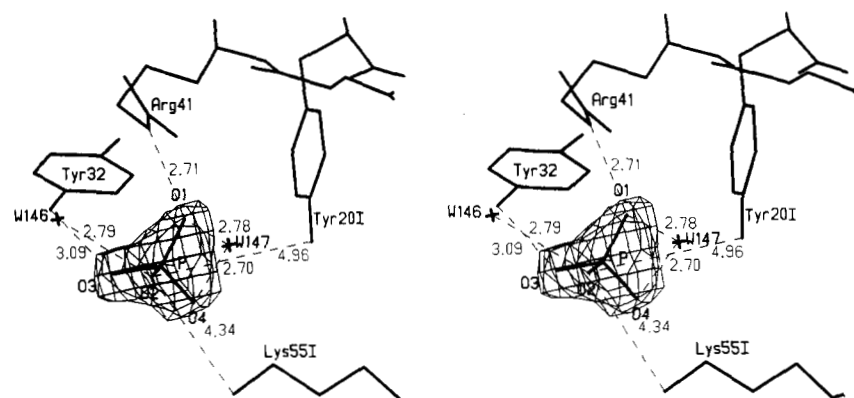
Fig. 4. Phosphate binding site in the Leu[18] variant complex, superimposed on a difference electron density map. The map contour level is 3σ. Dashed lines represent possible hydrogen bonds. Residues of OMTKY3 are labeled with an "I" following the sequence number to distinguish them from those of SGPB.

To facilitate pairwise comparisons of water molecules, the numbering of water molecules in the three complexes was unified, following that of the previously reported structure (Read et al., 1983). A water molecule was assigned if there was associated density for it at a 1.0σ level on the refined $2|F_o| - |F_c|$ map and if it had a refined B-factor lower than 60.0 Å^2. If a water molecule was present with an identical ID in all the three complexes, it was considered as a common water molecule in the three complexes. There are 146 such common water molecules that are observed in all the three complexes, with the largest RMSD of 0.41 Å (Ala[18] versus Gly[18] variants, Table 2). Therefore, a positional difference of 0.8 Å (2 times the RMSD) is significant enough and will be used as the distance cutoff in the following analysis of solvent structures. When comparing two structures, water molecules that superimposed within 0.8 Å are considered bound to an identical site and assigned an identical ID; differences greater than 0.8 Å are considered bound to different positions if different hydrogen bonding patterns are also observed.

Redistribution of water molecules
in the S₁ binding pocket

Comparisons among the three complexes and of each with the uncomplexed SGPB structure (H. Blanchard & M.N.G. James,

in prep.) reveal significant differences in the solvent distributions in the S₁ binding pocket. Figure 5 (Kinemage 1) represents the superposition of the four structures, centering on the S₁ binding pocket. Water molecules bound in the S₁ binding pocket are summarized in Table 3. The numbering system for the solvent molecules of the uncomplexed SGPB follows that of H. Blanchard and M.N.G. James (in prep.), which is independent from that of the three complexes. In the S₁ binding pocket of the uncomplexed SGPB, there are five water molecules, B438, B440, B471, B529, and B543. These five water molecules are conserved in the two noncrystallographically related molecules in the asymmetric unit of the uncomplexed SGPB (the RMSD for the five water molecules is 0.46 Å), therefore only one of the two SGPB molecules has been used in the structural comparisons and following discussion.

Upon binding of the Leu[18] variant of OMTKY3, three water molecules (B438, B471, and B543) of the uncomplexed SGPB are excluded from the S₁ pocket by the side chain of Leu 18 of OMTKY3 (Fig. 5; Kinemage 1). The remaining two water molecules (B440 and B529) are actually equivalent to W42 and W71 of the Leu[18] variant complex. In addition, two water molecules, W15 and W67 of the Leu[18] variant complex, that were not observed in the uncomplexed SGPB, are present in the S₁ binding pocket.

Table 2. *Comparisons of solvent structures in the three complexes*

	Average B-factor (Å^2)			Number of water molecules		
	Main-chain	Side-chain	Water molecules	Total	With $B < 1.5\,B_{ave}$	With $B < B_{ave}$
Leu[18] variant complex	13.28	17.75	36.80	169	150	88
Ala[18] variant complex	13.71	17.68	37.47	182	174	86
Gly[18] variant complex	13.49	17.38	36.63	168	161	83
Number of common water molecules[a]				146	124	64
RMSDs (Å)						
Leu[18] versus Ala[18]				0.38	0.3	0.24
Leu[18] versus Gly[18]				0.34	0.29	0.20
Ala[18] versus Gly[18]				0.41	0.36	0.25

[a] Numbering of water molecules in the three complexes follows that of the previously reported structure (Read et al., 1983), provided there is associated electron density for it at a $1.0\,\sigma$ level in the refined $2|F_o| - |F_c|$ map and a refined B-factor lower than 60.0 Å^2. If a water molecule is present in the three complexes with an identical ID, it is considered as a common water molecule.
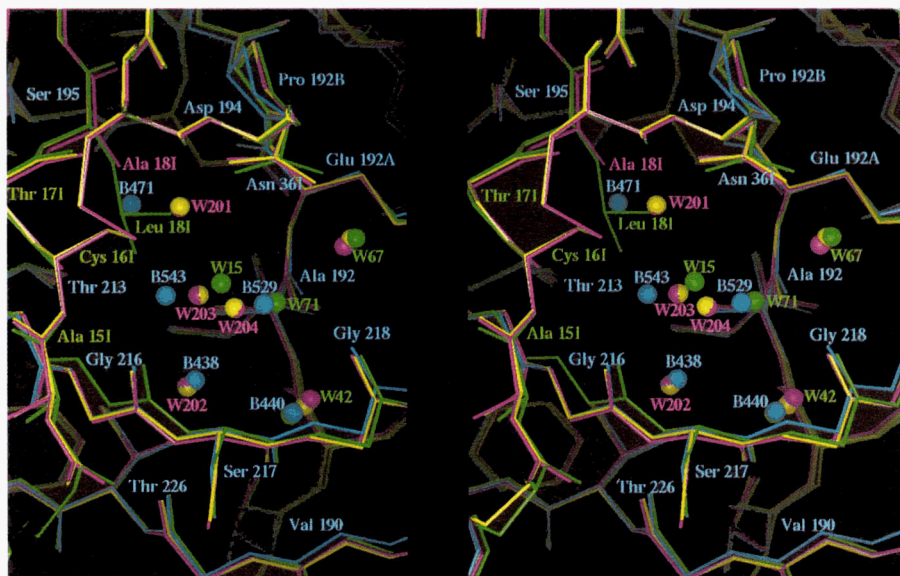
Fig. 5. Redistribution of water molecules bound in the S_1 binding pocket. Superposition of four structures: the uncomplexed SGPB (in cyan), complexes of SGPB with the Leu[18] variant (in green), the Ala[18] variant (in magenta), and the Gly[18] variant (in yellow). Residues of OMTKY3 are labeled with an "I" following the sequence number to distinguish them from those of SGPB. The picture is centered on the S_1 binding pocket to show the different positions of bound water molecules in each SGPB structure.

The complexes of the Ala[18] and Gly[18] variants display an almost identical solvent structure in the S_1 binding pocket (Fig. 5; Kinemage 1). Upon binding of the Ala[18] (Gly[18]) variant to free SGPB, three water molecules (B471, B529, and B543) are excluded from the S_1 binding pocket. Water molecule B438, which was excluded by the Leu[18] variant of OMTKY3, is in a position identical to W202 of the Ala[18] and Gly[18] variant complexes. Whereas water molecules W42 and W67 are conserved among the Leu[18], Ala[18], and Gly[18] variant complexes, W15 and W71 are not observed in the Ala[18] and Gly[18] variant complexes. Three water molecules (W201, W203, and W204), none of which is observed in either the Leu[18] variant complex or the uncomplexed SGPB, are newly introduced to the Ala[18] and Gly[18] variant complexes. W42 (B440) is the only water molecule that is conserved in the uncomplexed SGPB and the three complexed SGPB molecules.

Water molecule B438 of the uncomplexed SGPB, which is also present in the Ala[18] and Gly[18] variant complexes (equivalent to W202), is excluded from the Leu[18] variant complex (Fig. 5; Kinemage 1). This event may be coupled with the relatively small conformational change of loop Gly 215–Ser 217 of SGPB in the Leu[18] variant complex. Gly 216 C of SGPB advances by 0.7 Å

toward the side chain of Leu 18 of OMTKY3 in the complex. Consequently, B438 is excluded from the S_1 binding pocket to avoid the otherwise too close contact (<3.0 Å) with the C atom of Gly 216. An alternative explanation might be that the exclusion of water molecule B438 by the side chain of Leu 18 of OMTKY3 results in a vacant space, thereby triggering the inward shift of that loop of SGPB.

Water molecules participate in the enzyme–inhibitor association

In an attempt to appreciate the redistribution of water molecules in the S_1 binding pocket, we have examined the possible hydrogen bonds involving these water molecules. We found that all the water molecules that are newly introduced to the S_1 binding pocket are engaged in bridging hydrogen bonds between SGPB and OMTKY3. Table 4 lists all the possible hydrogen bonds observed in the interface of the complexes of SGPB and OMTKY3 variants, as well as those observed in the S_1 binding pocket of the uncomplexed SGPB. Figure 6 (Kinemage 2) shows the possible hydrogen bonds between the Leu[18] variant of OMTKY3 and SGPB, bridged by water molecules W15, W67, and W71.

Table 3. *Water molecules bound in the S_1 binding pocket of SGPB*

Crystal structure	Bound water molecules[a]				
Uncomplexed SGPB	B440	B529	B438	B471	B543
B-factor (Å^2)	41.91	64.75	47.50	65.04	55.17
Leu[18] variant complex	W42 (B440)	W71 (B529)	W15	W67	
B-factor (Å^2)	26.60	27.20	17.60	21.57	
Ala[18] variant complex	W42 (B440)	W67	W201	W202 (B438)	W203	W204
B-factor (Å^2)	17.78	26.31	17.68	22.05	37.65	41.68
Gly[18] variant complex	W42 (B440)	W67,	W201	W202 (B438)	W203	W204
B-factor (Å^2)	21.35	23.53	22.05	31.41	53.55	44.43

[a] Numbers in parentheses denote equivalent water molecules in the uncomplexed SGPB (H. Blanchard & M.N.G. James, in prep.).

Asn 36 N$^{\delta 2}$ of OMTKY3 forms indirect hydrogen bonds with the carbonyl oxygen of Gly 216 of SGPB, via water molecule W15. Asn 36 O$^{\delta 1}$ of OMTKY3 forms indirect hydrogen bonds with the carbonyl oxygen atoms of Ala 192 and Gly 218 of SGPB, via water molecules W67 and W71. Both W15 and W67 are newly introduced to the S₁ binding pocket upon the binding of the Leu^{18} variant. W71, which is equivalent to B529 of the uncomplexed SGPB, migrates by a distance of 0.5 Å, presumably to optimize these hydrogen bonding interactions.

Figure 7 (Kinemage 3) shows the probable hydrogen bonds observed in the Ala^{18} and Gly^{18} variant complexes, focusing only on those mediated by water molecules in the S₁ binding pocket. The indirect hydrogen bonds between Asn 36 of OMTKY3 and the carbonyl-oxygen atoms of Gly 216, Ala 192, and Gly 218 of SGPB, as observed in the Leu^{18} variant complex, are also present in the Ala^{18} and Gly^{18} variant complexes. They occur via water molecules W201, W202, W203, W204, and W67. Note that water molecules W201, W203, and W204 are only present in the Ala^{18} and Gly^{18} variant complexes; there are no equivalent solvent sites in either the uncomplexed SGPB or the Leu^{18} variant complex. These water molecules are introduced to the S₁ binding pocket to fill the vacant space resulting from the small side chains of Ala 18 and Gly 18 and, perhaps more importantly, to bridge the hydrogen bonding interactions between SGPB and OMTKY3.

Correlation of binding free energy with the hydrophobic surface area buried in the complex interface

SGPB is a chymotrypsin-like serine proteinase with respect to its substrate specificity. The S₁ binding pocket in SGPB is hydrophobic in nature and relatively large in size. Therefore, P₁ residues with bulky hydrophobic side chains are preferred over those with small or polar side chains. Noncoded amino acids, Abu, Ape, Ahx, and Ahp, were introduced into the P₁ position of OMTKY3 to probe systematically the binding profile of the S₁ pocket. The negative binding free energy ($-\Delta\Delta G^0$) increases with the increasing number of side-chain atoms of the P₁ residue (Lu, 1994; Table 5). Due to the relatively low yields and high

cost of chemical synthesis, the available quantities of the noncoded amino acid variants (Abu, Ape, Ahx, and Ahp) were not sufficient for conducting crystallization trials. Therefore, we have resorted to atomic models of the complexes of SGPB with these noncoded variants, in an effort to understand the S₁ specificity profile of SGPB. The torsion angles with which the P₁ side chains were constructed are listed in Table 5. The P₁ side chains of these variants can adopt favorable rotamers without causing any unacceptably close contacts with SGPB. The models were built on the basis of the Ala^{18} variant complex. The inward shift of residues Gly 214–Gly 216, as observed in the Leu^{18} variant complex, does not cause any unfavorable contacts and has negligible effects on the subsequent calculation of accessible surface area. A few water molecules bound in the S₁ pocket, such as W15 and W71 in the Leu^{18} variant complex, or W201, W203, and W204 in the Ala^{18}/Gly^{18} variant complex, are expected to be displaced by Ahp and possibly by Ahx.

The solvent-accessible surface area was calculated for each complex, the free SGPB and the free OMTKY3 variants using the corresponding coordinates. The coordinates for the free SGPB and the free OMTKY3 variants were extracted from their complexes. The total solvent-accessible surface area that is buried in the enzyme–inhibitor interface is defined as: $(S_E + S_I - S_{EI})$ Å², where S_E, S_I, and S_{EI} are the total solvent-accessible surface areas of the free SGPB, the free OMTKY3, and the SGPB-OMTKY3 complex, respectively.

Figure 8 is a plot of the binding free energy ($-\Delta\Delta G^0$, relative to that of the Gly^{18} variant) versus the buried hydrophobic surface area, for the six variants having straight aliphatic side chains. The increase in $-\Delta\Delta G^0$ is directly correlated with the increase in the hydrophobic surface area buried in the complex interface. The resulting constant of proportionality (the slope of the fit) is 34.1 cal mol^{-1} Å^{-2} for hydrophobic atoms (carbon and sulfur).

As expected, similar correlations are also found for complexes of the same set of variants with α-chymotrypsin and SGPA and have comparable constants of proportionality (36.9 cal mol^{-1} Å^{-2} for α-chymotrypsin and 40.4 cal mol^{-1} Å^{-2} for

Fig. 6. Water-mediated hydrogen bonds in the complex of SGPB with the Leu^{18} variant of OMTKY3. Dashed lines represent possible hydrogen bonds. Residues of OMTKY3 are labeled with an "I" following the sequence number to distinguish them from those of SGPB. The Leu^{18} variant complex (shown in yellow) is superimposed with the uncomplexed SGPB (shown in cyan), to emphasize that water molecules W15, W67, and W71, which are not present in the uncomplexed SGPB, are all engaged in bridging hydrogen bonds between SGPB and Asn 36 of OMTKY3.

Table 4. *Hydrogen bonding and nonbonded interactions (within 4.1 Å) in the interface of the complexes of SGPB and OMTKY3 variants*

H-bonding partners					H-bond distance (Å)			
					Leu[18] variant	Ala[18] variant	Gly[18] variant	Uncomplexed SGPB
SGPB		**OMTKY3**						
Thr 39	O	Arg 21I	$N^{\eta 2}$		3.16	2.93	2.95	
Arg 41	N	Tyr 20I	O		3.10	3.00	3.13	
Arg 41	O	Tyr 20I	N		2.96	2.89	3.00	
Tyr 171	O	Lys 13I	N^{ζ}		2.88	3.19	3.18	
Gly 193	N	X18I	O		2.50	2.61	2.47	
Asp 194	N	X18I	O		3.3[a]	3.23[a]	3.29[a]	
Ser 195	N	X18I	O		3.05	2.94	3.12	
Ser 195	O^{γ}	X18I	N		2.83	2.91	2.89	
Gly 216	N	Cys 16I	O		2.98	2.96	3.13	
Gly216	O	Cys 16I	N		2.99	2.83	2.80	
SGPB		**Water**						
Ala 192	N	W42	O		2.82	2.89	2.85	
Ala 192	O	W42	O		3.12	2.80	2.84	
Ala 192	O	W71	O		2.59	N/A[b]	N/A[b]	
Ala 192	O	W202	O		N/A	3.71	3.54	
Ala 192	O	W203	O		N/A	3.23	3.65	
Gly 216	O	W15	O		2.99	N/A	N/A	
Gly 216	O	W204	O		N/A	3.14	3.04	
Gly 216	N	W202	O		N/A	3.48	3.49	
Ser 217	O	W42	O		2.82	3.02	2.99	
Ser 217	O	W202	O		N/A	3.57	3.87	
Ser 217	N	W202	O		N/A	3.38	3.56	
Gly 218	O	W67	O		3.08	3.03	2.92	
Gly 218	O	W71	O		2.74	N/A	N/A	
Gly 218	O	W42	O		3.99	3.28	3.43	
Asn 219	N	W42	O		4.03	3.29	3.41	
Thr 225	O	W202	O		N/A	3.08	3.42	
Thr 226	$O^{\gamma 1}$	W42	O		3.45	2.81	3.96	
OMTKY3		**Water**						
Asn 36I	$N^{\delta 2}$	W15	O		3.17	N/A	N/A	
Asn 36I	$N^{\delta 2}$	W201	O		N/A	2.93	3.07	
Asn 36I	$O^{\delta 1}$	W67	O		2.86	2.79	3.08	
Asn 39I	N	W26	O		3.40	3.40	3.55	
Water		**Water**						
W15	O	W26	O		3.17	N/A	N/A	
W15	O	W71	O		2.90	N/A	N/A	
W26	O	W204	O		N/A	3.18	3.31	
W42	O	W202	O		N/A	3.20	3.25	
W67	O	W71	O		2.85	N/A	N/A	
W201	O	W203	O		N/A	2.69	2.56	
W202	O	W203	O		N/	2.75	2.52	
W204	O	W203	O		N/A	3.41	3.29	
Uncomplexed SGPB		**Water**						
Ala 192	N	B440 (W42)						3.03
Ala 192	O	B440 (W42)						3.20
Ala 192	O	B529 (W71)						3.14
Glu 192A	O	B471						3.21
Ser 217	O	B440 (W42)						2.72
Gly 218	O	B529 (W71)						2.61
Thr 225	O	B438 (W202)						3.16
Thr 226	$O^{\gamma 1}$	B438 (W202)						2.91
B438 (W42)	O	B440 (W42)						2.70
B438 (W42)	O	B529 (W71)						3.36
B438 (W42)	O	B543						2.39
B543	O	B471						3.39
B543	O	B529 (W71)						2.98

[a] Asp 194 N of SGPB is in close contact with the X18I O of OMTKY3, but the value of the N-H···O angle precludes the formation of a hydrogen bond. ASP 194 NH is hydrogen-bonded to Glu 192A CO.

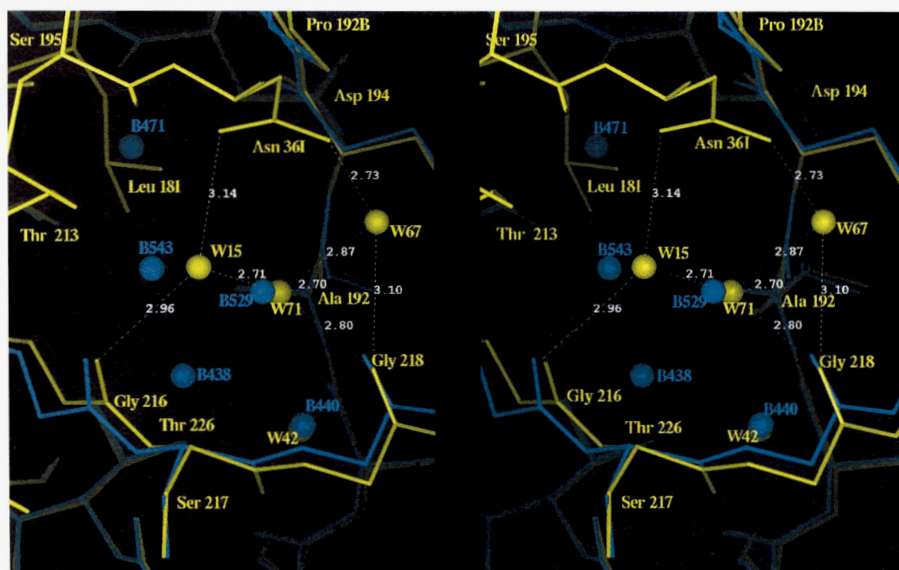[b] N/A refers to distances that are longer than 3.6 Å.

Fig. 7. Water-mediated hydrogen bonds in the complex of SGPB with the Ala[18] variant of OMTKY3. Dashed lines represent possible hydrogen bonds. Residues of OMTKY3 are labeled with an "I" following the sequence number to distinguish them from those of SGPB. The Ala[18] variant complex (shown in magenta) is superimposed with the uncomplexed SGPB (shown in cyan), to emphasize that water molecules W201, W203, and W204, which are not present in either the uncomplexed SGPB or the Leu[18] variant complex, are all engaged in bridging hydrogen bonds between SGPB and Asn 36 of OMTKY3.

SGPA, data not shown). SGPA exhibits a substrate specificity very similar to that of SGPB. The observed strong correlations attribute the K_a variations to the amounts of the hydrophobic surface areas that are in contact in the enzyme–inhibitor interface. It indicates that the primary determinant of the binding free energy is the desolvation energy of the hydrophobic P_1 side chain of the inhibitor and the hydrophobic S_1 binding pocket in these enzymes, when no conformational strains are involved.

Discussion

The combination of the available thermodynamic data with the atomic structures showing the interactions of the P_1 variants of OMTKY3 with SGPB provides an excellent system to study the contribution of hydrophobic interactions to protein–protein associations. The currently used measurements of the hydrophobic strength, by cavity-creating substitutions of interior residues, suffer from the complication of potential conformational disturbance of the protein structure (reviewed by Fersht & Serrano, 1993; Matthews, 1993). In contrary, the transfer of the P_1 res-

idue to the preformed, relatively rigid S_1 pocket involves only displacement and redistribution of water molecules, thus providing a clean measurement of the unit contribution of hydrophobic interactions (hydrophobic strength) to protein and protein–complex stability.

There has been a large discrepancy (reviewed by Fersht & Serrano 1993; Matthews, 1993) over the value of the hydrophobic strength between the microscopic measurements (20–30 cal mol^{-1} Å^{-2} as derived from solute partition experiments [Chothia, 1974; Eisenberg & McLachlan, 1986]) and macroscopic measurements (70–75 cal mol^{-1} Å^{-2} by surface tension measurement [Aveyard & Haydon, 1965]). Substitutions of residues in the hydrophobic cores of proteins and subsequent assessments of the unfolding energy of the mutant proteins gave an estimate of 45–60 cal mol^{-1} Å^{-2} (Kellis et al., 1989; Shortle et al., 1990), which is close to the macroscopic value. Theoretical calculations applying size-difference correction (De Young & Dill, 1990; Sharp et al., 1991) and curvature effects (Nicholls et al., 1991) tend to support the macroscopic measurement. Our estimate (34–40 cal mol^{-1} Å^{-2}) is higher than the solvation pa-

Table 5. *Binding equilibrium constants (Lu, 1994) and calculated buried surface areas for a selection of complexes of SGPB and OMTKY3 variants*

No. of atoms	P_1	K_a (M^{-1})	$-\Delta G^0$ (kcal mol^{-1})	$-\Delta\Delta G^0$ (kcal mol^{-1})	Surface area (Å^2) Apolar	Total	Torsion angles of the P_1 side chain χ_1	χ_2	χ_3	χ_4
0	Gly	1.2×10^7	9.49	0	662[a]	1,165[a]	—			
1	Ala	3.6×10^8	11.47	1.98	719[a]	1,226[a]	—			
2	Abu	2.3×10^9	12.55	3.06	745[b]	1,253[b]	$-56°$			
3	Ape	1.2×10^{10}	13.51	4.02	772[b]	1,277[b]	$-56°$	$-50°$		
4	Ahx	2.1×10^{10}	13.84	4.35	784[b]	1,292[b]	$-56°$	$-50°$	$170°$	
5	Ahp	2.8×10^{10}	14.00	4.51	801[b]	1,306[b]	$-56°$	$-50°$	$170°$	$164°$

[a] Calculated from crystal structure.
[b] Calculated from model structure.

rameters (20–30 cal mol^{-1} $Å^{-2}$) for hydrophobic atoms that were derived from experimental measurements of transferring free amino acids from water to an organic solvent (Chothia, 1974; Eisenberg & McLachlan, 1986), but lower than the hydrophobic strength (49–61 cal mol^{-1} $Å^{-2}$) estimated from substitutions of interior residues of barnase (Kellis et al., 1989). It would appear, therefore, that the S_1 pocket of SGPB is not as hydrophobic as the interior cavity in the core of a protein. This phenomenon can be understood by the unique chemical nature of the enzyme–inhibitor interface. Unlike the substitution-created interior cavities that are usually completely depleted of water molecules, the S_1 binding pocket is partially hydrated, as ordered water molecules are observed crystallographically. This partial hydration may account for the reduced hydrophobicity of the S_1 pocket. On the other hand, the apparently reduced hydrophobicity may also have to do with the fact that, when an interior residue is substituted by a smaller amino acid, the surrounding residues tend to relax to fill the substitution-resulting cavity, a process in which the change of free energy may be considerable. Consequently, the substitution-resulting change of the unfolding energy is highly position-dependent. For example, the changes in the unfolding energies associated with the Leu to Ala substitutions in T4 lysozyme vary considerably from site to site, ranging from 1.9 kcal/mol to 6 kcal/mol (Eriksson et al., 1992). In the case of the SGPB–OMTKY3 complex, burial of the P_1 side chain brings only displacement and/or redistribution of water molecules and minor, if any, conformational disturbances in the S_1 binding pocket. For this reason, the SGPB–OMTKY3 system should provide one of the most realistic measurements of the hydrophobic strength. The constants of proportionality determined for SGPB, α-chymostrypsin, and SGPA are comparable (34–40 cal mol^{-1} $Å^{-2}$), indicating that our assessment of the hydrophobic strength is reliable.

The homologous series of aliphatic, straight-chain P_1 variants, Gly^{18}, Ala^{18}, Abu^{18}, Ape^{18}, Ahx^{18}, and Ahp^{18}, provides a good system to estimate the contribution of a methyl group

to protein–protein association and stabilization. The energetic contribution of a methyl group estimated from Gly to Ala substitution is higher than that estimated from Ala to Abu substitution (2.0 kcal/mol versus 1.1 kcal/mol). The discrepancy of 0.9 kcal/mol may be due to the additional conformational flexibility of the P_1 glycine residue relative to those amino acids having side chains. However, the observed correlations between the binding energy and the buried surface area suggest that the effect of the additional conformational variability of glycine is actually minimal. Otherwise, the Gly^{18} variant would be off the correlation curve by a considerable amount. The crystal structures of the Gly^{18} and the Ala^{18} variant complexes are virtually indistinguishable, thus glycine does not seem to behave differently from the other amino acids. It is mainly the relatively large increase in the buried hydrophobic surface area (an 8.6% increase for Gly to Ala substitution versus the 3.6% increase for Ala to Abu substitution) that accounts for the increase in binding energy with the addition of a methyl group to glycine.

The observed differences in the solvent structures for the different complexes lead to the conclusion that, during the process of SGPB–OMTKY3 complex formation, bound water molecules in the S_1 binding pocket are not simply displaced by the inhibitor, but that there is an accompanying redistribution of solvent binding sites, synchronous with the conformational change of loop Gly 215–Ser 217 of SGPB. We believe that these two events are to avoid unfavorable nonbonded contacts and, perhaps more importantly, to reinforce enzyme–inhibitor interaction by introducing hydrogen bonds between SGPB and OMTKY3. Water-mediated hydrogen bonding interactions are implicated in stabilizing a variety of other molecular complexes. Water-mediated hydrogen bonds have been observed in the recognition and formation of the DNA and trp repressor complex (Shakked et al., 1994), and the DNA and TATA binding protein complex (Kim & Burley, 1994). In these protein–DNA complexes, bound water molecules become part of the recognition motif of the DNA sequence. In the case of protein and peptide association, it has been shown that water molecules play important roles in forming stable complexes of the OppA protein, an oligopeptide-binding protein, with peptides of varying sequences and lengths (Tame et al., 1994). In the study of L-arabinose-binding protein, it was found that bound water molecules can modulate the specificity and affinity of the protein toward different sugar substrates (Quiocho et al., 1989). Additionally, Madden et al. (1992) have observed that a few conserved water molecules participate in forming key hydrogen bonds between the class I MHC molecule HLA-B27 and its bound peptide. Furthermore, Arnez and Steitz (1994) have proposed that pseudo-uridinylation may stabilize $tRNA^{Gln}$ through water-mediated linking of the pseudouridine to the backbone phosphate. The absence of these bound water molecules in the unmodified $tRNA^{(Gln)}$ structure may account for the reduced thermal stability of the unmodified $tRNA^{Gln}$. In our study, we have demonstrated that water molecules provide key interactions linking SGPB and the several OMTKY3 inhibitors. Taken together, the roles played by water molecules in protein–ligand recognition and association are extensive and should not be underestimated.

The water-mediated hydrogen bonds observed in the three complex structures show that Asn 36 of OMTKY3 is invariably involved in forming water-bridged hydrogen bonds to residues of SGPB. Thus, a mutation of Asn 36 to the isosteric amino acid leucine could provide a measure of the impact on binding by the observed water-mediated hydrogen bonds. However, the ener-
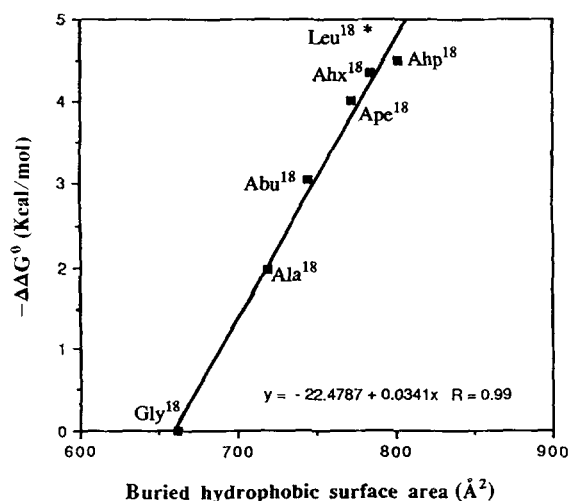

Fig. 8. Linear correlation of the binding free energy $-\Delta\Delta G^0$ with the hydrophobic surface area buried in the interface of a selection of complexes of SGPB with OMTKY3 variants. The constant of proportionality is 34.1 cal mol^{-1} $Å^{-2}$. The Leu^{18} variant (position shown by an "*") was not included in the line fitting.

getic contributions of these bound water molecules are not expected to be high. Comparisons among avian ovomucoid inhibitor third domains from different species indicate that mutations Asn 36 to Asp, Asn 36 to Ser, and Asn 36 to Ala do not result in much variation in the binding with SGPB (less than 0.9 kcal/mol; Lu, 1994). The functions of these bound water molecules perhaps are to fill the vacant space left in the interface and to satisfy the hydrogen bonding potentials of protein atoms. The gain of favorable enthalpic energy by forming hydrogen bonds may be comparable to the loss of unfavorable entropic energy by ordering the water molecules in the interface. The most desirable situation would be that the ligand displaces all the previously bound water molecules (entropically favorable), while concomitantly forming hydrogen bonds or van der Waals contacts with the enzyme (enthalpically favorable).

Protein–protein associations usually involve large contact surface areas (600–1,800 Å²) and from 10 to 30 residues from either side. Mutagenesis studies in many systems have demonstrated that not all residues involved in the associations are energetically important. Usually, only a few of the residues are critical for tight binding. For example, two tryptophan residues constitute the "functional epitope" of the human growth hormone receptor, accounting for nearly three-quarters of the binding free energy (Clarkson & Wells, 1995). The functional epitopes of the hormone receptor are analogous to the specificity determinant, the P_1 residue of OMTKY3, in terms of energy contributions as well as their geometrical disposition relative to the main body of the protein (Trp 104 and Trp 169 project from two adjacent loops). Large hydrophobic amino acids are increasingly implicated in protein–protein association. The crystal structure of N cadherin suggests that a tryptophan residue may play an important role in cadherin dimerization, a process thought to enhance cell adhesion (Shapiro et al., 1995). Accordingly, large surface depressions that accommodate those bulky hydrophobic residues become interesting when improved specificity and/or enhanced affinity are concerned. Such surface depressions bear a strong physical resemblance to the S_1 binding pocket of SGPB. Because functional epitopes are the primary targets for inhibitor design, knowledge obtained from OMTKY3 may be applicable to designing ligands targeting hormone receptor proteins, MHC molecules, and other functionally important proteins. We have demonstrated that the P_1 substitution-associated changes in buried hydrophobic surface area are the direct cause of the variations in the binding energy of OMTKY3 with SGPB, α-chymotrypsin, and SGPA. It is reasonable to believe that the buried hydrophobic surface area-binding free energy correlation is general to other molecular complexes in which large hydrophobic depressions are involved. It would be very interesting to see what values the constant of proportionality would be in other systems, to identify the factors that influence the hydrophobic strength. It will help to resolve the controversy over the values of the unit contribution of hydrophobic effect and further our understanding about the nature of the hydrophobic effect to protein stability.

Materials and methods

Crystallization

SGPB, purified according to the procedure of Jurasek et al. (1971), was kindly provided by Dr. L.B. Smillie. The cloning, overexpression system, and the purification procedures for OMTKY3 have been described (Lu et al., 1992; Lu, 1994). In addition to the P_1 substitutions, the recombinant variants are five amino acids shorter at the N-terminus than native OMTKY3. The 1:1 stoichiometric complexes of SGPB–OMTKY3 variants were prepared by mixing SGPB with each OMTKY3 variant in a 1:1.05 molar ratio of enzyme to inhibitor. Crystals of the complex were grown by hanging drop methods, under conditions similar to those reported previously (Read et al., 1983). Each drop contained 3 μL of protein solution (10 mg/mL) and an equal volume of reservoir solution (2% of PEG 6000, 50 mM sodium/potassium phosphate at pH 6.5). Seeding was usually required to get good crystals. The crystals reach their final size (1.0 × 0.3 × 0.1 mm³) in about a week. These crystals are isomorphous to those of the complex of SGPB with native OMTKY3. Unit cell parameters of the crystals of the three complexes (Leu[18], Ala[18], and Gly[18]) are listed in Table 6.

Data collection

X-ray diffraction data for the Leu[18] variant and the Gly[18] variant complexes were collected on San Diego Multiwire System twin area detectors. Graphite monochromated Cu K_α radiation was generated with a Rigaku rotating anode generator RU-200BH operating at 40 kV and 150 mA. The 1.8-Å data set was collected from a single crystal of the Leu[18] variant complex, the 1.9-Å data set was collected from a single crystal of the Gly[18] variant complex. Data collections were carried out at 24 °C, each lasting about 2 days. The 1.4-Å data set of the Ala[18] variant complex was collected at beam-line 18B of the Photon Factory in Japan (room temperature 18 °C), using Weissenberg geometry with a crystal-to-image-plate distance of 286.50 mm (Sakabe, 1991). The recorded reflections were indexed and integrated by profile fitting with the program WEIS (Higashi,

Table 6. *Statistics of data collection*

Space group:	$P2_1$			
Cell dimensions:	a (Å)	b (Å)	c (Å)	β (°)
Leu[18] complex	45.53	54.66	45.59	119.19
Ala[18] complex	45.38	54.62	45.47	119.20
Gly[18] complex	45.48	54.68	45.55	118.99

Data collection	Leu[18] complex	Ala[18] complex	Gly[18] complex
Maximum resolution (Å)	1.8	1.4	1.9
No. of crystals used	1	1	1
No. of total observations	41,124	83,250	36,307
No. of unique reflections	17,183	30,783	13,125
Average redundancy	2.4	2.7	2.8
Completeness of data			
Overall (%)	94.3	80.3	84.9
Highest resolution shell (%)	51.1 (2.0–1.8 Å)	56.0 (1.5–1.4 Å)	36.5 (2.0–1.9 Å)
R_{merge}[a] (%)	5.29	7.33	4.29

[a] $R_{merge} = \sum_{hkl} [(\sum_i |I_i - \langle I \rangle|) \sum_i I_i]$.

1989). Data-collection statistics for all these complexes are given in Table 6.

Structure determination and refinement

The phases for the complexes of SGPB with the three OMTKY3 variants were taken from those of the complex of SGPB with native OMTKY3 (Read et al., 1983). A $(2|F_o| - |F_c|)\exp(i\alpha_{cal})$ map, where $|F_o|$ is the observed structure–factor amplitudes of the variant complex, $|F_c|$ and α_{cal} are the structure–factor amplitudes and phases calculated from the structure of Read et al. (1983), was calculated with the program SIGMAA (Read, 1986) and displayed with FRODO (Jones, 1985). The P_1 residue Leu 18 was replaced by the appropriate P_1 amino acid side chains to fit best the observed electron density. The starting R-factor for the Leu[18], Ala[18], and Gly[18] variant complexes were 36.4%, 42.5%, and 35.9%, respectively, and were improved to 20.6%, 22.8%, and 20.8%, respectively, after energy minimization and simulated annealing with XPLOR (Brünger, 1986). The model was then subjected to TNT refinement (Tronrud, 1992). All reflections with $d_{min} < 20.0$ Å were used for the refinements. After 10 cycles of TNT refinement, a new $2|F_o| - |F_c|$ map was made for visual inspection. Special attention was given to regions where there were noticeable conformational disparities with the previously reported structure and where atoms displayed high thermal factors. Up to this stage, no water molecules were built into the complex model. A $(|F_o| - |F_c|)\exp(i\alpha_{cal})$ difference map was made to assign water molecules. Water molecules that were present in the previously reported structure were built into the variant complex model if there were associated densities for it above the 2.5σ contour level in the difference map. The manually adjusted model was subjected to another 10 cycles of TNT refinement. Afterward, water molecules with B-factors higher than 50 Å^2 were examined closely. A water molecule would be deleted if there was no associated density at the 1σ contour level on a $2|F_o| - |F_c|$ map and if the B-factor was greater than 60 Å^2. New water molecules were added if there was electron density above the 2.5σ level on a $|F_o| - |F_c|$ difference map. The geometry of potential hydrogen bond donor or acceptor atoms was also checked to assure that it was meaningful to introduce a water molecule in that region. Refinement was judged to reach convergence typically after a total of 30 cycles of TNT refinement.

Structure comparison

Superpositions of the several structures were done in a pairwise manner using the lsq commands in the program "O" (Jones et al., 1991). The RMSDs for common atoms were obtained subsequently from the program. Crystallographically determined water molecules were grouped into three shells on the basis of their thermal factors. The three B-factor shells include all water molecules, water molecules with B-factor lower than the $B_{average}$, and those lower than $1.5 B_{average}$, respectively. Water molecules within the same B-factor shell are compared.

Modeling of the noncoded amino acid variants of OMTKY3

The currently available quantities of the noncoded amino acid variants (Abu, Ape, Ahx, and Ahp) were not sufficient for con-

ducting crystallization trials. Accordingly, we have built models of the four complexes on the basis of the crystal structure of the complex between SGPB and the Ala[18] OMTKY3 variant. In doing so, the side chains of the P_1 residues were rotated to most favorable rotamer conformations while maintaining favorable van der Waals contacts. The modeling was very straight forward, no energy minimization was required or applied.

Calculation of solvent-accessible surface area

Solvent-accessible surface area was calculated using the program MS (Connolly, 1983). The probe radius for water molecules was 1.4 Å. Van der Waals radii for protein atoms are the default of the program. Sulfur is treated as hydrophobic. Solvent-accessible surface areas for the free SGPB and the free OMTKY3 variants were calculated with the coordinates extracted from the corresponding complexes, under the valid assumption that the binding process is close to a rigid-body association.

Coordinates

The atomic coordinates have been deposited with the Protein Data Bank. The ID codes are 1SGP (Leu[18]OMTKY3-SGPB complex), 1SGQ (Ala[18]OMTKY3-SGPB complex), and 1SGR (Gly[18]OMTKY3-SGPB complex).

References

Arnez JG, Steitz TA. 1994. Crystal structure of unmodified tRNA[Gln] complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry 33*:7560–7567.

Aveyard R, Haydon DA. 1965. Thermodynamic properties of aliphatic hydrocarbon/water interfaces. *Trans Faraday Soc 61*:2255–2261.

Bigler TL, Lu W, Park SJ, Tashiro M, Wieczorek WR, Laskowski M Jr. 1993. Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P_1 residues. *Protein Sci 2*:786–799.

Bode W, Epp O, Huber R, Laskowski M Jr, Ardelt W. 1985. The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur J Biochem 147*:387–395.

Bode W, Huber R. 1992. Natural protein proteinase inhibitors and their interactions with proteinases. *Eur J Biochem 204*:433–451.

Bode W, Wei AZ, Huber R, Meyer E, Travis J, Neumann S. 1986. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J 5*:2453–2458.

Chothia C. 1974. Hydrophobic bonding and accessible surface area in proteins. *Nature 248*:338–339.

Clarkson T, Wells JA. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science 267*:383–386.

Codding PW, Delaere LT, Hayakawa K, Hutcheon WL, James MNG, Jurasek L. 1974. The 4.5 Ångstrom resolution structure of a bacterial serine proteinase from *Streptomyces griseus*. *Can J Biochem 52*:208.

Connolly ML. 1983. Analytical molecular surface calculation. *J Appl Crystallogr 16*:548–558.

De Young LR, Dill KA. 1990. Partitioning of nonpolar solutes into bilayers and amorphous *n*-alkanes. *J Phys Chem 94*:801–809.

Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and stability. *Nature 319*:199-203.

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science 255*:178-183.

Fersht NR, Serrano L. 1993. Principles of protein stability derived from protein engineering experiments. *Curr Opin Struc Biol 3*:75-83.

Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M Jr, James MNG. 1987. Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J Mol Biol 195*:397-418.

Higashi T. 1989. The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J Appl Crystallogr 22*:9-18

Jones TA. 1985. Interactive computer graphics: FRODO. *Methods Enzymol 115*:157-171. [With modifications for TOM UQV/CIT, v.2.8.0, by C. Cambillau, M. Israel, G. Griffin, & A. Chirino.]

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A 47*:110-119.

Jurasek L, Johnson P, Olafson RW, Smillie LB. 1971. An improved fractionation system for pronase on CM-sephadex. *Can J Biochem 49*:1195-1201.

Kellis JT Jr, Nyberg K, Fersht AR. 1989. Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry 28*:4914-4922.

Kim JL, Burley SK. 1994. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct Biol 1*:638-651.

Kollman PA. 1994. Theory of macromolecule-ligand interactions. *Curr Opin Struct Biol 4*:240-245.

Komiyama T, Bigler TL, Yoshida N, Noda K, Laskowski M Jr. 1991. Replacement of P₁ Leu[18] by Glu[18] in the reactive-site loop of turkey ovomucoid third domain converts it into a strong inhibitor of Glu-specific *Streptomyces griseus* proteinase (GluSGP). *J Biol Chem 266*:10727-10730.

Laskowski M Jr, Kato I. 1980. Protein inhibitors of proteinases. *Biochemistry 49*:593-626.

Lu W. 1994. Energetics of the interactions of ovomucoid third domain variants with different serine proteinases [thesis]. West Lafayette, Indiana: Purdue University.

Lu W, Zhang W, Molly SS, Thomas G, Ryan K, Chiang Y, Anderson S, Laskowski M Jr. 1993. Arg[15]-Lys[17]-Arg[18] turkey ovomucoid third domain inhibits human furin. *J Biol Chem 268*:14583-14585.

Madden DR, Gorga JC, Strominger JL, Wiley DC. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight binding to MHC. *Cell 70*:1035-1048.

Matthews BW. 1993. Structural and genetic analysis of protein folding and stability. *Curr Opin Struct Biol 3*:589-593.

Musil D, Bode W, Huber R, Laskowski M Jr, Lin TY, Ardelt W. 1991. Refined X-ray crystal structures of the reactive site modified ovomucoid

inhibitor third domains from silver pheasant (OMSVP3*) and from Japanese quail (OMJPQ3*). *J Mol Biol 220*:739-755.

Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct Funct Genet 11*:282-296.

Papamokos E, Weber E, Bode W, Empie MW, Kato I, Laskowski M Jr. 1982. Crystallographic refinement of Japanese quail ovomucoid, a Kazal type inhibitor, and model building studies of complexes with serine proteases. *J Mol Biol 158*:515-537.

Quiocho FA, Wilson DK, Vyas NK. 1989. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature 340*:404-406.

Read RJ. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A 42*:140-149.

Read RJ, Fujinaga M, Sielecki AR, James MNG. 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry 22*:4420-4433.

Read RJ, James MNG. 1986. Introduction to the protein inhibitors: X-ray crystallography. In: Barrett AJ, Salvesen G, eds. *Proteinase inhibitors.* Amsterdam: Elsevier. pp 301-336.

Sakabe N. 1991. X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl Methods Phys Res Sect A 303*:448-463.

Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun 27*:157-162.

Shakked Z, Guzikevich-Guerstein G, Frolow F, Rabinovich D, Joachimiak A, Sigler PB. 1994. Determinants of repressor/operator recognition from the structure of the trp operation binding site. *Nature 368*:469-473.

Shapiro L, Fannon AM, Kwong PD, Thompson A, Lehmann MS, Grubel G, Legrand J, Als-nielsen J, Colman DR, Hendrickson WA. 1995. Structural basis of cell-cell adhesion by cadherins. *Nature 374*:327-336.

Sharp KA, Nicholls A, Friedman R, Honig B. 1991. Extracting hydrophobic free energies from experimental data: Relationship to protein folding and theoretical models. *Biochemistry 30*:9686-9697.

Shortle D, Stites WE, Meeker AK. 1990. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry 29*:8033-8041.

Tame JRH, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, Higgins CF, Wilkinson AJ. 1994. The structural basis of sequence-independent peptide binding by OppA protein. *Science 264*:1578-1581.

Tronrud DE. 1992. Conjugate-direction minimization: An improved method for the refinement of macromolecules. *Acta Crystallogr A 48*:912-916.

Weber E, Papamokos E, Bode W, Huber R, Kato I, Laskowski M Jr. 1981. Crystallization, crystal structure analysis and molecular model of the third domain of Japanese quail ovomucoid, a Kazal type inhibitor. *J Mol Biol 149*:109-123.

Wieczorek M, Park SJ, Laskowski M Jr. 1987. Covalent hybrids of ovomucoid third domains made from one synthetic and one natural peptide chain. *Biochem Biophys Res Commun 144*:499-504.