Despite Having a Common P_1 Leu, Eglin C Inhibits α-Lytic Proteinase a Million-fold More Strongly than Does Turkey Ovomucoid Third Domain†

M. A. Qasim,*,‡ Robert L. Van Etten,‡ Tina Yeh,§ C. Saunders,|| P. J. Ganz,|| S. Qasim,‡ L. Wang,‡ and M. Laskowski, Jr.‡,⊥

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907-2084, Accelrys, Inc., 10188 Telesis Court, San Diego, California 92121, and Miami Valley Laboratories, The Procter and Gamble Company, Cincinnati, Ohio 45253

Received March 6, 2006; Revised Manuscript Received May 23, 2006

ABSTRACT: Results of the inhibition of α-lytic proteinase by two standard mechanism serine proteinase inhibitors, turkey ovomucoid third domain (OMTKY3) and eglin C, and many of their variants are presented. Despite similarities, including an identical P_1 residue (Leu) in their primary contact regions, OMTKY3 and eglin C have vastly different association equilibrium constants toward α-lytic proteinase, with K_a values of 1.8×10^3 and 1.2×10^9 M^{-1}, respectively. Although 12 of the 13 serine proteinases tested in our laboratory for inhibition by OMTKY3 and eglin C are more strongly inhibited by the latter, the million-fold difference observed here with α-lytic proteinase is the largest we have seen. The million-fold stronger inhibition by eglin C is retained when the K_a values of the P_1 Gly, Ala, Ser, and Ile variants of OMTKY3 and eglin C are compared. Despite the small size of the S_1 pocket in α-lytic proteinase, interscaffolding additivity for OMTKY3 and eglin C holds well for the four P_1 residues tested here. To better understand this difference, we measured K_a values for other OMTKY3 variants, including some that had residues elsewhere in their contact region that corresponded to those of eglin C. Assuming intrascaffolding additivity and using the K_a values obtained for OMTKY3 variants, we designed an OMTKY3-based inhibitor of α-lytic proteinase that was predicted to inhibit 10000-fold more strongly than wild-type OMTKY3. This variant ($K^{13}A/P^{14}E/L^{18}A/R^{21}T/N^{36}D$ OMTKY3) was prepared, and its K_a value was measured against α-lytic proteinase. The measured K_a value was in excellent agreement with the predicted one (1.1×10^7 and 2.0×10^7 M^{-1}, respectively). Computational protein docking results are consistent with the view that the backbone conformation of eglin C is not significantly altered in the complex with α-lytic proteinase. They also show that the strong binding for eglin C correlates well with more favorable atomic contact energy and desolvation energy contributions as compared to OMTKY3.

Standard mechanism inhibitors bind to their cognate serine proteinases like a substrate except for being hydrolyzed extremely slowly (*1*). At present, 18 families of standard mechanism serine proteinase inhibitors are known (*2*). The three-dimensional structure of at least one representative of 16 of the 18 families has been determined. They all have a canonical conformation in the reactive site binding loop (*3*), and they generally exhibit a high degree of superposition of all of the backbone atoms from P_3 to P_3' in this region (*4*, *5*). Other parts of the structures of these inhibitor families are very different.

We have focused our studies on the Kazal family of standard mechanism serine proteinase inhibitors, more specifically on a subset of this family, namely, the third domain of an avian egg white protein, ovomucoid. The X-ray crystallographic structure of ovomucoid third domain from turkey in complex with five enzymes, SGPB,[1] CHYM, HLE, CARL, and PPE, has been determined (*6*−*9*; M. N. G. James and M. Laskowski, Jr., unpublished results). X-ray structure determinations have provided us with a consensus set of 12 inhibitor residues that contact the enzyme in the inhibitor−enzyme complex (see Figure 1). Two of these, a Cys at position 16 (P_3) and an Asn at position 33 (P_{15}'), are structural and are nearly unvaried in Kazal inhibitors. In developing a sequence-to-reactivity algorithm, our laboratory has mutated all contact positions, except these two, to the other 19 amino acid residues (*10*, *11*). This provided us with a set of 190 single-change variants in addition to wild-type OMTKY3. Measurements of association equilibrium constants of all of these variants with our six selected enzymes produced a complete database of specificity variation at 10 of the 12 consensus contact positions. Assuming complete additivity at all of these contact positions, we can in principle calculate

† This research was supported by a grant from the National Institutes of Health (GM10831) and in part by the Howard Hughes Medical Institute through the Undergraduate Science Education Program.

* To whom correspondence should be addressed: Department of Chemistry, Xavier University of Louisiana, 1 Drexel Drive, New Orleans, LA 70125. E-mail: qasim.ma@gmail.com. Phone: (504) 520-5082. Fax: (504) 520-7942.

‡ Purdue University.

§ Accelrys, Inc.

|| The Procter and Gamble Co.

⊥ Deceased August 2, 2004.

[1] Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; CARL, subtilisin Carlsberg; CHYM, bovine chymotrypsin Aα; HLE, human leukocyte elastase; OMCNG3, OMGUI3, OMHPA3, OMLKD3, and OMTKY3, ovomucoid third domains from Canada goose, Helmet guinea fowl, Hungarian partridge, lake duck, and turkey, respectively; PPE, porcine pancreatic elastase; SGPA and SGPB, *Streptomyces griseus* protease A and B, respectively.

10.1021/bi060445l CCC: $33.50
Published on Web 08/30/2006

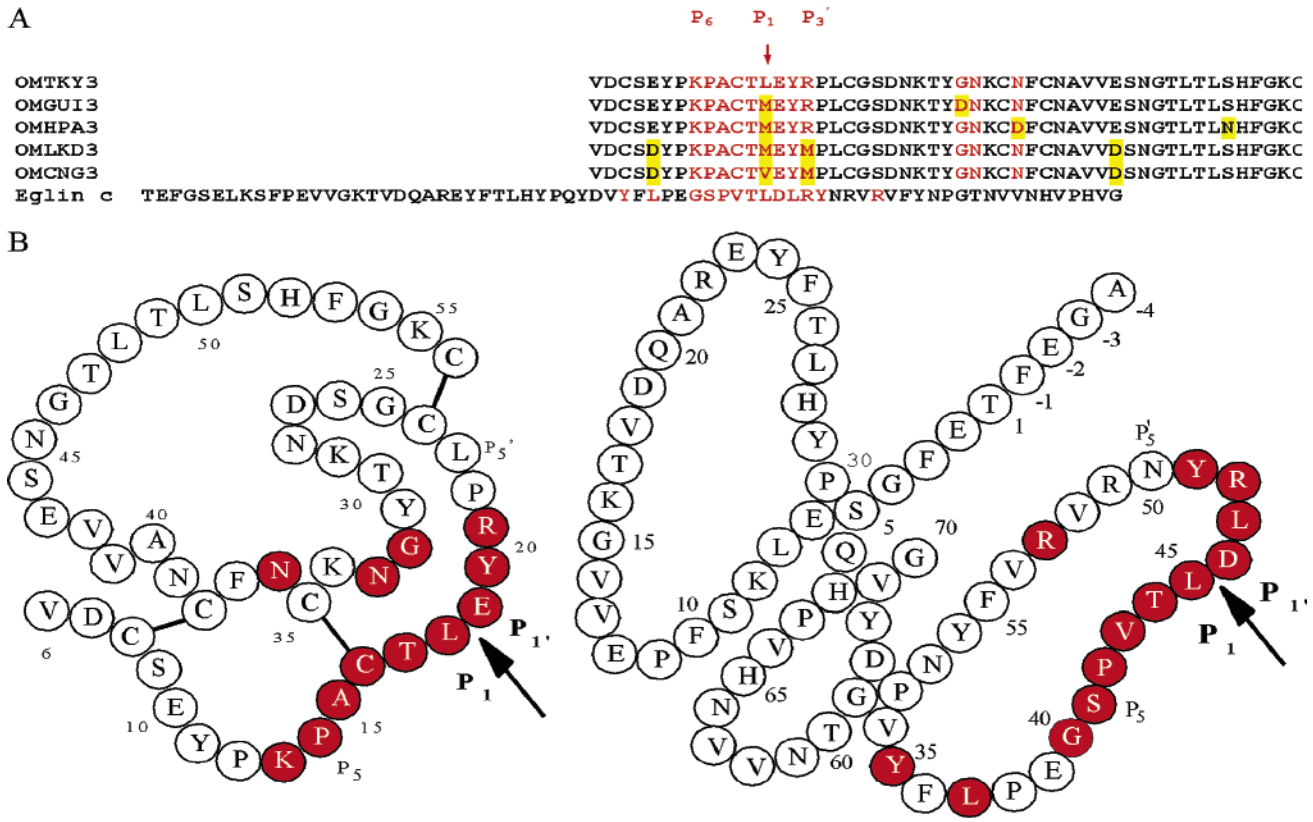

FIGURE 1: (A) Amino acid sequences of the five avian ovomucoid third domains and eglin C. The sequence of eglin C is aligned with respect to its reactive site binding loop only (residues P_6-P_3'). Consensus contact residues are colored red. For avian ovomucoid third domains, changes from the OMTKY3 sequence are highlighted in yellow. The reactive site residue is denoted with an arrow. (B) Covalent structures of turkey ovomucoid third domain, OMTKY3 (left) and eglin C (right). The recombinant OMTKY3 variants whose K_a values are utilized in this paper start at residue 6. The four NH_2-terminal residues in eglin C marked −4 to −1 are a result of recombinant DNA operations (*18*). The arrows indicate the reactive site peptide bonds, which serve as origins of the Schechter and Berger (*30*) P_n-P_n' notation. In each inhibitor, the consensus sets of residues that come in contact with the cognate enzyme are colored red. The disulfide linkages are represented as thick lines between cysteines. There are no disulfide bridges in eglin C.

the binding constant of any Kazal inhibitor against all of our six selected enzymes. This is the basis of our sequence-to-reactivity algorithm (*10*, *11*). With only a few restrictions, it holds very well (*10*, *12*). Our set of 190 variants is an ideal set for determining the complete specificity profile of hitherto uninvestigated serine proteinases, providing only that they can be inhibited by OMTKY3 and its variants.

α-Lytic proteinase is a serine proteinase that is secreted extracellularly by the soil bacterium *Lysobacter enzymogenes*. X-ray crystallographic structures have been determined for this enzyme alone (*13*), in complex with peptide boronic acid inhibitors (*14*, *15*), and in complex with its propeptide region (*16*) and a number of mutants in which the primary specificity pocket is potentially affected (*17*). Such structural studies together with measurements of the kinetic parameters of α-lytic proteinase with substrates differing in P_1 residues provide a molecular basis for interpreting the role of the primary specificity pocket (S_1) in this enzyme. However, little is known about the inhibition of this enzyme by protein inhibitors or about the specificity due to sites other than S_1.

Our initial inhibition studies of α-lytic proteinase with wild-type OMTKY3 ended in disappointment when OMTKY3 proved to be so weak as an inhibitor that the association equilibrium constant was barely measurable. Because previous studies from our laboratory (*18*) have shown that eglin C (a potato family I inhibitor) generally inhibits serine proteinases more strongly than OMTKY3, we also tried eglin C as an inhibitor of α-lytic proteinase. The switch to eglin C as an inhibitor proved to be highly successful. Here, we describe our results for the inhibition of α-lytic proteinase by OMTKY3, eglin C, and a number of their variants. The results of this study also enabled us to design an OMTKY3-based inhibitor that is a 10^4 times more powerful as an inhibitor than wild-type OMTKY3. Models of the α-lytic proteinase complexes with OMTKY3 and with eglin C were also compared using a computational protein−protein docking procedure, and a potentially useful working model of the eglin C complex was obtained.

EXPERIMENTAL PROCEDURES

Expression of OMTKY3 and Eglin C Variants. OMTKY3 variants were expressed in *Escherichia coli* as a fusion protein with the modified B domain (Z domain) of protein A. The procedure has been described in detail previously (*19*). Briefly, the transformed cells were grown in 2×YT medium for 24 h at 30 °C. The cells were collected by centrifugation and subjected to osmotic shock to release the expressed protein in solution from periplasmic space. The supernatant was passed through an affinity column of IgG-Sepharose, and bound protein was released by washing with pH 3.3 buffer. The fusion protein contains two potential cleavage sites: one specific for the Glu-specific enzyme *Staphylococcus aureus* protease V8 and the other for cyanogen

bromide. With the exception of inhibitor variants that have an internal Cys or Met residue, the cleavage method of choice in our laboratory is cyanogen bromide. Both of the cleavage methods cut the fusion protein and produce two unequal sized fragments. The smaller of the two is the inhibitor. These fragments were separated on a P-10 size exclusion column. The inhibitor was further purified by ion exchange chromatography on S-Sepharose and Q-Sepharose columns. The identity of the inhibitor variants was confirmed by amino acid analysis and by ion spray mass spectrometry that yielded a mass that was within 0.5 amu of the calculated molecular weight of the variant.

Preparation of the $K^{13}A/P^{14}E/L^{18}A/R^{21}T/N^{36}D$ OMTKY3 Variant. The variant was prepared by site-directed mutagenesis using the pEZZ318.tky plasmid as the starting template (*19*). The variant was obtained in two steps. The first step involved the use of the $N^{36}D$ variant of OMTKY3 and introduction of the $K^{13}A/P^{14}E$ double mutation. The plasmid of this variant was then used to introduce the other two changes, $L^{18}A$ and $R^{21}T$. The PCR product was treated with *Dpn*I to digest the original plasmid, and then the product was transformed into *E. coli* strain XL-1 (Stratagene) for screening and selection. The plasmid was sequenced by the standard dideoxy chain termination method to confirm the mutation, and then it was transformed into *E. coli* strain RV-308 for protein expression. Expression and purification were carried out as described previously (*19*).

Eglin C variants were produced in *Bacillus* by using an engineered plasmid, where expression of the eglin C gene was driven by the *Bacillus amyloliquefaciens* subtilisin gene regulatory elements (*18*). The expression system uses the subtilisin signal sequence to direct the secretion of eglin C into the culture medium. The purification of eglin C involves a few simple steps starting with the removal of cells by centrifugation and acidification of the supernatant to pH 3.0. After the insoluble material is removed by centrifugation, the inhibitor is purified by ion exchange chromatography on an S-Sepharose column at pH 4.0. The identity of variants was confirmed by ion spray mass spectrometry.

α-Lytic protease was expressed from *Lysobacter enzymogenes* and purified to homogeneity as described previously (*20*).

K_a Measurements. The procedure involved for K_a measurements has been described in earlier publications from this laboratory (*19*, *21*). More detailed aspects can be found in some of the Ph.D. theses completed at Purdue University (*22*−*24*). The principle of K_a determination is based on mixing the enzyme and inhibitor at appropriate concentrations and determining the unbound enzyme concentration by using a suitable chromogenic or fluorogenic substrate. Generally, an enzyme at a fixed concentration is incubated in a set of cuvettes with increasing inhibitor concentrations such that the molar concentration of the inhibitor varies from 20 to 200% of the enzyme concentration. After the attainment of equilibrium (incubation time $\approx 10t_{1/2}$), a suitable substrate was added. In the case of chromogenic substrates, the kinetics of release of *p*-nitroanilide was followed between 380 and 410 nm on an HP8453 diode array spectrophotometer. Simultaneously, any disturbances due to light scattering were corrected by continuously subtracting the absorption contribution at 650−700 nm. This subtraction considerably improves the quality of the data. For fluorogenic substrates, excitation and emission wavelengths of 370 and 440 nm, respectively, were used. The K_a values were obtained by fitting the data to a two-parameter (for K_a values of $<10^7$) or three-parameter (for K_a values of $>10^7$) nonlinear equation (*23*).

Computational Protein−*Protein Docking*. Computational protein−protein docking studies were conducted using an initial-stage rigid-body docking algorithm, ZDOCK (*25*), and a refinement procedure, RDOCK (*26*). ZDOCK is based on the fast Fourier transform (FFT) (*27*) technique that efficiently explores the entire six degrees of freedom in the translational and rotational space. RDOCK refinement is based on CHARMm energy minimization, and a reranking of the predicted poses using a free energy scoring function composed of the electrostatic energy and the desolvation energy calculated by the atomic contact energy (ACE) method (*28*). We performed protein−protein docking of α-lytic proteinase with eglin C and OMTKY3, using the ZDOCKpro module in Insight II. In both studies, the structure of the proteinase was taken from PDB entry 1GBK. The structure of eglin C was taken from PDB entry 1ACB and the structure for OMTKY3 from PDB entry 1CHO.

ZDOCK was run with a 6° rotational sampling grid and the default 1.2 Å grid step size. A total of 54 000 poses were sampled, and the top 1000 poses were saved. Since we expected the catalytic site of α-lytic proteinase and the exposed loop region of the inhibitor to be involved in the binding interface, we filtered the ZDOCK run results. In the α-lytic proteinase−eglin C docking, the catalytic histidine (His57) of α-lytic proteinase and the P_1 residue (Leu45) were specified for filtering. In the α-lytic proteinase−OMTKY3 docking, the same catalytic histidine and the P_1 residue (Leu18) were specified for filtering. In each case, an 8 Å cutoff distance was used. The RDOCK refinement procedure performed CHARMm minimization on the top 50 poses of the filtered ZDOCK output, and the poses were reranked on the basis of the RDOCK empirical free energy scoring function.

RESULTS AND DISCUSSION

Inhibition of α-Lytic Proteinase by OMTKY3 and Eglin C. Association equilibrium constants for inhibition of α-lytic proteinase by OMTKY3 and eglin C and several of their P_1 variants were measured at 22 °C and are listed in Table 1 (Table 1 of ref *29*). The third and fifth columns in this table list free energies of association ($\Delta G^\circ = -RT \ln K_a$) in kilocalories per mole. The inhibitor−proteinase association involves approximately a dozen residues of the inhibitor that are in contact with enzyme residues in the complex. A set of consensus contact residues has been defined on the basis of X-ray crystallographic structures of OMTKY3 and eglin C with several serine proteinases. These are shown in red letters in Figure 1 (Figure 1 of ref *30*). The two inhibitors share three identical residues in their primary contact region (P_2T, P_1L, and P_3'R) and chemically similar residues at P_1' (E and D). Despite these similarities, the K_a values of the two inhibitors with α-lytic proteinase are vastly different. Eglin C is a 10^6-fold stronger inhibitor of α-lytic proteinase than is OMTKY3 (Table 1). Indeed, a comparison of the K_a values of 14 different serine proteinases with OMTKY3 and eglin C shows that with the exception of PPE all the other enzymes are inhibited more strongly by eglin C than by OMTKY3. The million-fold difference observed for α-lytic

Table 1: Association Equilibrium Constants and $\Delta G°$ Values for the Association of P_1 Variants of OMTKY3 and Eglin C with α-Lytic Proteinase

	inhibitor				substrate[a]	
	OMTKY3		eglin C		Suc-AAPX-pna[b]	Ac-AAPX-NH_2[c]
inhibitor variant	K_a (M^{-1})	$\Delta G°$ (kcal/mol)	K_a (M^{-1})	$\Delta G°$ (kcal/mol)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
P_1 Gly	2.0×10^3	−4.46	2.5×10^9	−12.69	295	2.9
P_1 Ala	6.3×10^4	−6.48	7.5×10^{10}	−14.68	26800	48
P_1 Val	2.2×10^4	−5.86			1060	15
P_1 Ser	1.5×10^4	−5.64	1.2×10^{10}	−13.61		
P_1 Leu (wild type)	1.8×10^3	−4.39	1.2×10^9	−12.26	10.8	
P_1 Met	1.2×10^4	−5.51			1520	
P_1 Ile	9.2×10^2	−4.00	5.7×10^8	−11.82	2.1	
P_1 Asp			9.0×10^4	−6.69		
P_1 Glu			4.4×10^4	−6.27		
P_1 Pro			8.3×10^5	−7.99		

[a] Data taken from ref *17*. [b] Succinyl-Ala-Ala-Pro-Xxx-*p*-nitroanilide. [c] Data taken from ref *29*.

proteinase, however, is the largest among these. The two inhibitors also have different residues at four primary contact positions (Figure 1). To see if any of these residues (P_4−P_6 and P_2') could account for the large difference between eglin C and OMTKY3, we screened OMTKY3 variants that had the same residues that were found in eglin C for their inhibition of α-lytic proteinase. No significant improvement in inhibition was observed. Because inhibitors belonging to different families use an entirely different set of secondary contact residues, it is not possible to evaluate their contribution by interscaffold residue substitutions.

The importance of the P_1 residue is also seen in studies of the kinetics of hydrolysis of substrates that possess different residues at the scissile peptide bond. For comparison, the k_{cat}/K_M values obtained by two research groups (*17*, *29*) using slightly different substrates are also listed in Table 1. Qualitatively, the k_{cat}/K_M values agree with the K_a values of the P_1 variants of OMTKY3 and eglin C. For instance, among the five comparisons that are possible, the Ala at P_1 is best both for substrate hydrolysis and for inhibitor binding, while Ile is the worst. A plot of log K_a of OMTKY3 or eglin C versus log k_{cat}/K_M for these substrates yields a correlation coefficient of 0.93 and a slope of 1.92.

Table 1 also shows K_a results obtained with Gly, Ala, and Ser at P_1 of OMTKY3 and eglin C. These three P_1 variants of eglin C were also approximately 10^6-fold stronger inhibitors for α-lytic proteinase than were the corresponding OMTKY3 variants. This is consistent with earlier results showing extensive free energy additivity between P_1 variants of OMTKY3 and eglin C for six serine proteinases (*18*). Since OMTKY3 and eglin C belong to different inhibitor families (*2*) and have a very different scaffold, we often refer to this additivity as interscaffolding additivity. For example, the difference in the free energy of association of P_1 Ala and P_1 Gly variants (see Table 1) of OMTKY3 ($\Delta\Delta G° = -2.02$ kcal/mol) is identical to the difference between the P_1 Ala and P_1 Gly variants of eglin C ($\Delta\Delta G° = -1.99$ kcal/mol). The $\Delta\Delta G°$ values of other pairs of P_1 variants were also, within experimental error, very similar and demonstrate interscaffolding additivity between P_1 of OMTKY3 and eglin C in their interaction with α-lytic proteinase. Since inhibitors belonging to different inhibitor families in general have a canonical reactive site loop that involves residues P_3−P_3', one might naively expect that interscaffolding additivity should hold for these residues as well. However, except for P_1, this has not yet been demonstrated. It is unclear if there is any interscaffolding additivity involving other primary contact residues, but it is highly unlikely that there is for secondary contact residues, since inhibitors belonging to different families have very different sets of secondary contact residues.

We often make a simple assumption that scaffolding does not matter in inhibitor−enzyme association. Although this is true for many cases, there are some clear exceptions. For instance, BPTI shows easily measurable K_a values (range from 10^5 to 10^8 M^{-1}) for CHYM, SGPA, and SGPB, but it does not inhibit CARL at all (M. A. Qasim and M. Laskowski, Jr., unpublished results). X-ray crystallographic and modeling studies show serious clashes between the enzyme residues and the scaffolding region of BPTI (*31*). Unlike the majority of other standard mechanism inhibitors, eglin C may have greater flexibility in its reactive site binding loop (*32*−*34*). If so, the flexibility of the eglin C reactive site might enable it to adapt to the active sites of a variety of different serine proteinases. In this case, the demonstration of interscaffolding additivity for interaction of α-lytic proteinase with P_1 variants of OMTKY3 and eglin C is an important result in that it strongly implies the identical orientations of P_1 residues of OMTKY3 and eglin C in the S_1 pocket of α-lytic proteinase. Furthermore, it implies that the large difference in the association constant of the two inhibitors is probably not due to a qualitative difference in their mode of binding. Currently, only limited structural information is available. The interactions of α-lytic proteinase with peptide boronic acid inhibitors and structure determinations of these inhibitors with α-lytic proteinase have been examined by Bone et al. (*14*). However, the complexity of the adduct formed in the interaction of boronic acid with serine proteinases makes it difficult to compare those results with the ones presented here. The only X-ray crystallographic structures of the complex of OMTKY3 and eglin C with the same serine proteinase are with bovine chymotrypsin Aα. In these structures, all the backbone atoms of the two inhibitors from P_3 to P_3' and all the side chain atoms of Leu show complete superimposibility (*18*).

Computational Modeling of the Complex. In the absence of a crystal structure of a complex of α-lytic proteinase with eglin C that would reveal detailed contact residues, we utilized a computational docking procedure to compare the two inhibitors. The protein complex structures that resulted from the docking studies were analyzed in the Insight II environment. In each case, the top-ranked poses form clusters

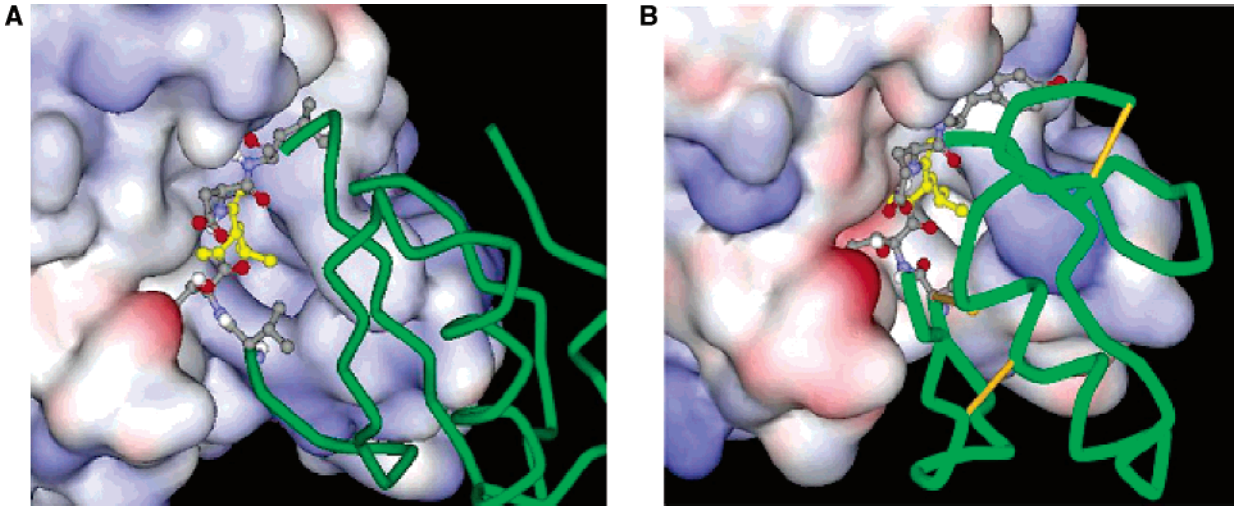

FIGURE 2: (A) Top-ranked α-lytic protease−eglin C complex (ZDOCK pose 43). α-Lytic protease is rendered as a solvent accessible surface and colored according to electrostatic potential. For eglin C, loop residues Val43, Thr44, Leu45 (yellow), Asp46, and Leu47 are in ball-and-stick format and the remainder is rendered as a tube. (B) Top-ranked α-lytic protease−OMTKY complex (ZDOCK pose 5). α-Lytic protease is rendered as a solvent accessible surface. For OMTKY3, loop residues Cys16, Thr17, Leu18 (yellow), Glu19, and Tyr20 are in ball-and-stick format and the rest is rendered as a tube.

that reveal the expected binding mode (Figure 2). Among the docked poses in the cluster, the residues in the binding loop region of the inhibitor tend to exhibit much smaller backbone rmsd values than regions in the rest of the molecule. Although protein side chain movements among the binding interface residues are observed after the RDOCK refinement, the backbone conformation changes are minor. Several intermolecular hydrogen bonds consistently form between the binding loop region and the active site region of the proteinase in both complexes.

Remarkably, for the α-lytic proteinase−eglin C complex, the top 25 ranked poses of the 50 RDOCK predicted poses are clustered around the common binding mode. For the α-lytic proteinase−OMTKY3 complex, such clustering is only observed with the top eight ranked poses of the 50 RDOCK predicted poses (data not shown). For all the top-ranking RDOCK predicted poses, the desolvation energy scores calculated by the ACE method for the α-lytic proteinase−eglin C complexes were consistently more favorable (by approximately 6−8 kcal/mol) than those of the α-lytic proteinase−OMTKY3 complexes. The electrostatic energy was comparable in magnitude in both systems (see the tables of the Supporting Information). This finding for α-lytic protease complexes is consistent with computational free energy calculations conducted for α-chymotrypsin−inhibitor complexes, where it was found that desolvation interactions were the major contributors to the overall stability of the complexes (*35*).

The ACE method determines the desolvation free energy by estimating the effective energy changes for the transfer of protein atoms of different types from water to a protein environment. The approach is based on knowledge-based potentials for residue−residue contact energies, derived from the structural information for molecular packing in the known protein crystal structures. All protein heavy atoms are classified into 18 atom types on the basis of considerations of chemical properties and cooperative side chain interactions, and the coordination numbers and contact energies for the protein atoms are determined by considering statistical information on atom packing and observed pairing frequencies of 20 amino acids in the database of protein structures. In this approach, the desolvation energy calculated with the ACE method would include both the polar (electrostatic) and nonpolar (hydrophobic) contributions. However, the direct pairwise electrostatic interactions are not included in the ACE calculation. Therefore, we use the ACE energy plus the CHARMm electrostatic interaction energy with a distance-dependent dielectric constant as the RDOCK scoring function to rank the final docking poses. This scoring function can be considered as a complete free energy function except for an entropy term.

The docking studies indicated that the RDOCK scoring function successfully discriminated the expected enzyme−inhibitor binding mode with P_1 Leu as the leading anchor residue in both systems. This correct binding mode is identified as the top-ranked predicted pose in both systems. Moreover, for the α-lytic proteinase−eglin C complex, the top-ranked poses formed a very large cluster all with the correct binding mode and all with favorable ACE and electrostatic energy scores. Importantly, the docking study predictions provide a quantitative energetic measure that supports the experimentally determined much stronger binding of eglin C versus OMTKY3 for α-lytic proteinase. Moreover, these docking results suggest that the backbone geometry is not significantly changed in the complex. Although the docking results must be regarded as suggestive in the absence of a specific high-resolution crystal structure, they appear to be fully consistent with the results of Camacho and colleagues, who have stressed the role of anchor residues and latch residues, acting in conjunction with a weakly bound, nativelike conformation, in the formation of high-affinity complexes (*36*).

Influence of Other Contact Positions on α-Lytic Protease Binding. We have prepared all 190 single variants of OMTKY3 at the 10 consensus contact positions (*10*, *11*). The measurement of association equilibrium constants with six serine proteinases and the assumption of complete intra-scaffolding additivity at each of the 10 contact position are the bases of our sequence-to-reactivity algorithm. Measurements of association equilibrium constants of these variants

Table 2: Association Equilibrium Constants and $\Delta G°$ Values for the Association of Different Variants of OMTKY3 with α-Lytic Proteinase

	K_a (M^{-1})	$\Delta G°$ (kcal/mol)	$\Delta\Delta G°$ (kcal/mol) ($\Delta G°_{mut} - \Delta G°_{wt}$)
wild type	1.8×10^3	−4.39	0.00
P_6 Gly	3.5×10^2	−3.43	0.96
P_6 Ala	7.0×10^3	−5.19	−0.80
P_6 Ser	5.5×10^3	−5.05	−0.66
P_5 Ala	2.7×10^3	−4.63	−0.24
P_5 Asp	6.8×10^3	−5.17	−0.78
P_5 Glu	6.9×10^3	−5.18	−0.79
P_5 Phe	6.7×10^3	−5.17	−0.78
P_5 Tyr	7.2×10^3	−5.21	−0.82
P_3' Gly	1.7×10^3	−4.36	0.07
P_3' Ala	1.7×10^3	−4.36	0.07
P_3' Thr	4.5×10^3	−4.93	−0.54
P_3' Met	4.4×10^3	−4.92	−0.53
P_3' Phe	3.7×10^3	−4.82	−0.43
P_{14}' Asp	1.1×10^3	−4.11	0.28
P_{18}' Asp	1.5×10^4	−5.64	−1.25

Table 3: Measured and Predicted Free Energies of Association of Different Ovomucoid Third Domains with α-Lytic Proteinase

	measured		predicted[a]		
	K_a (M^{-1})	$\Delta G°$ (kcal/mol)	K_a (M^{-1})	$\Delta G°$ (kcal/mol)	change from OMTKY3[b]
OMTKY3	1.8×10^3	−4.39			
OMGUI3	1.0×10^4	−5.40	7.3×10^3	−5.22	LP_1M, $GP_{14}'D$
OMHPA3	2.1×10^5	−7.18	1.0×10^5	−6.75	LP_1M, $NP_{18}'D$
OMLKD3	2.9×10^4	−6.02	3.9×10^4	−6.20	LP_1M, $RP_3'M$
OMCNG3	7.3×10^4	−6.57	5.4×10^4	−6.39	LP_1V, $RP_3'M$

[a] These values were calculated as described in the text and in earlier publications (*10*, *11*). [b] See Figure 1 for the locations of these changes in the sequence.

with other serine proteinases could in principle provide a wealth of information about the specificity and ultimately the design of a specific inhibitor for such serine proteinases. There is, however, a caveat. If the wild-type inhibitor is as weak as an inhibitor of the enzyme as it is for the inhibition of α-lytic proteinase by OMTKY3, then the measurement of association equilibrium constants of variants weaker than the wild type becomes nearly impossible. However, variants which improve the binding over the wild type can still be very useful for improving the inhibitory activity in inhibitor design. In Table 2, we give K_a measurements for a few selected variants at positions other than P_1. The improved binding of several such variants can be seen in the negative $\Delta\Delta G°$ values (Table 2). The cumulative effect of these changes in OMTKY3 coupled with the best P_1 residue should produce a moderately strong inhibitor of α-lytic proteinase. Conversely, the comparable set of mutations introduced into eglin C can be predicted to yield an inhibitor with a K_a of $\sim 1 \times 10^{13}$ M^{-1}, a value that is approximately 2 orders of magnitude beyond our limit of experimental measurement.

Interaction of Avian Ovomucoid Third Domains with α-Lytic Proteinase. The association equilibrium constants for the interaction of recombinant OMTKY3 with α-lytic proteinase are listed in Tables 1 and 3. Natural OMTKY3 differs from recombinant OMTKY3 in having five extra amino acids at the N-terminus (*19*). There is overwhelming evidence from our laboratory demonstrating that these N-terminal residues have no influence on the association equilibrium constant. They do, however, increase the stability of third domains by ~9 °C (as measured by T_m determinations). The association equilibrium constants were also measured for four additional ovomucoid third domains which have a favorable residue at P_1 (see Figure 1). The measured values are listed in the second and third columns of Table 3. The fourth and fifth columns of this table list the predicted values. The predicted values were calculated by using the following equation (*10*)

$$\Delta G°_{predicted} = \Delta G°_{wt} + \sum_{i=1}^{i=10} \Delta\Delta G°(X_{wt}iX)$$

where $\Delta\Delta G°(X_{wt}iX)$ is the free energy increment as a result of mutation at one of the 10 consensus contact positions. Predicted free energies of association for the four ovomucoid third domains listed in Table 3 were calculated by using the data from Table 2. The agreement between the predicted and measured values is very good, suggesting that changes in these ovomucoid third domains produce additive intra-scaffolding effects in the free energy of interaction with α-lytic proteinase. These results are consistent with the observation that ~90% of the 450 cases we have studied for the six enzymes are either fully or partially additive (*10*, *12*). The algorithm has also been successfully used to predict the association equilibrium constant of nonclassical Kazal inhibitors for subtilisin Carlsberg (*37*). We decided to test this further by using our limited data set at each of the contact positions for which we had K_a data. By selecting the best residue at each of the contact positions, we prepared a variant of OMTKY3 that involved five changes. These were $K^{13}A$, $P^{14}E$, $L^{18}A$, $R^{21}T$, and $N^{36}D$ (Figure 1 and Table 2). A simple calculation predicts that such a variant should have an association equilibrium constant value of 2.0×10^7 M^{-1} at pH 8.3. The actual measured value is 1.1×10^7 M^{-1} which is again in excellent agreement with the predicted number. This represents an ~10000-fold improvement over that of OMTKY3. The aim in this design was not to make the strongest possible OMTKY3 inhibitor for α-lytic proteinase. That would require K_a data for all 10 consensus contact positions and for all 19 amino acids at each of these positions. The successful design of a moderately strong inhibitor based on a limited data set serves to prove the point.

Three important points emerge from these results. First, although OMTKY3 has served us as an ideal wild-type inhibitor of the six enzymes on which we have focused over the last several years, it may not be ideal for other enzymes. This is clearly true for α-lytic proteinase. We would have been far better off to have selected OMHPA3 as the wild type here. Second, the excellent agreement between the measured and predicted values clearly shows that the substitutions in the contact region of OMTKY3 are intra-scaffolding additive. Finally, the complete set of 190 variants of OMTKY3 is a powerful tool for deciphering the specificity of little known serine proteinases that are inhibited by Kazal family inhibitors.

ACKNOWLEDGMENT

We thank Prof. W. W. Bachovchin and Elissa L. Ash of Tufts University (Medford, MA) for their generous gift of purified α-lytic protease.

SUPPORTING INFORMATION AVAILABLE

Tables 1 and 2 present desolvation and electrostatic energy contributions for the 24 top-ranked ZDOCK poses obtained in the docking of α-lytic protease with eglin C and with OMTKY3, respectively. Corresponding to each set, the structure of the highest-ranked docked complex is presented as a PDB file (ALP_Eglinc_RDOCK_1.pdb and ALP_OMTKY_RDOCK_1.pdb, respectively). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Laskowski, M., Jr., and Kato, I. (1980) Protein inhibitors of proteinase, *Annu. Rev. Biochem. 49*, 593−626.
2. Laskowski, M., Jr., and Qasim, M. A. (2000) What can the structures of enzyme−inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta 1477*, 324−337.
3. Bode, W., and Huber, R. (1992) Natural protein proteinase-inhibitors and their interaction with proteinases, *Eur. J. Biochem. 204*, 433−451.
4. Apostoluk, W., and Otlewski, J. (1998) Variability of the canonical loop conformations in serine proteinases inhibitors and other proteins, *Proteins 32*, 459−474.
5. Krowarsch, D., Cierpicki, T., Jelen, F., and Otlewski, J. (2003) Canonical protein inhibitors of serine proteases, *Cell. Mol. Life Sci. 60*, 2427−2444.
6. Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M., Jr., and James, M. N. G. (1982) Refined crystal structure of the molecular complex of *Streptomyces griseus* protease-b, a serine protease, with the 3rd domain of the ovomucoid inhibitor from turkey, *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868−4872.
7. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M., Jr., and James, M. N. G. (1987) Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid 3rd domain at 1.8 Å resolution, *J. Mol. Biol. 195*, 397−418.
8. Bode, W., Wei, A.-Z., Huber, R., Meyer, E., Travis, J., and Neumann, S. (1986) X-ray crystal-structure of the complex of human-leukocyte elastase (pmn elastase) and the 3rd domain of the turkey ovomucoid inhibitor, *EMBO J. 5*, 2453−2458.
9. Maynes, J. T., Cherney, M. M., Qasim, M. A., Laskowski, M., Jr., and James, M. N. G. (2005) Structure of the subtilisin Carlsberg-OMTKY3 complex reveals two different ovomucoid conformations, *Acta Crystallogr. D61*, 580−588.
10. Lu, S. M., Lu, W., Qasim, M. A., Anderson, S., Apostol, I., Ardelt, W., Bigler, T., Chiang, Y.-W., Cook, J., James, M. N. G., Kato, I., Kelly, C., Kohr, W., Komiyama, T., Lin, T.-Y., Ogawa, M., Otlewski, J., Park, S. J., Qasim, S., Ranjbar, M., Tashiro, M., Warne, N., Whatley, H., Wieczorek, A., Wieczorek, M., Wilusz, T., Wynn, R., Zhang, W., and Laskowski, M., Jr. (2001) Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases, *Proc. Natl. Acad. Sci. U.S.A. 98*, 1410−1415.
11. Laskowski, M., Jr., Qasim, M. A., and Yi, Z.-P. (2003) Additivity-based prediction of equilibrium constants for some protein−protein associations, *Curr. Opin. Struct. Biol. 13*, 130−139.
12. Qasim, M. A., Lu, W., Lu, S. M., Ranjbar, M., Yi, Z.-P., Chiang, Y.-W., Ryan, K., Anderson, S., Zhang, W., Qasim, S., and Laskowski, M., Jr. (2003) Testing of the additivity-based protein sequence to reactivity algorithm, *Biochemistry 42*, 6460−6466.
13. Fujinaga, M., Delbaere, L. T., Brayer, G. D., and James, M. N. G. (1985) Refined structure of α-lytic protease at 1.7-Å resolution: Analysis of hydrogen-bonding and solvent structure, *J. Mol. Biol. 184*, 479−502.
14. Bone, R., Frank, D., Kettner, C. A., and Agard, D. A. (1989) Structural-analysis of specificity: α-Lytic protease complexes with analogs of reaction intermediates, *Biochemistry 28*, 7600−7609.
15. Bone, R., Fujishige, A., Kettner, C. A., and Agard, D. A. (1991) Structural basis for broad specificity in α-lytic protease mutants, *Biochemistry 30*, 10388−10398.
16. Sauter, N. K., Mau, T., Rader, S. D., and Agard, D. A. (1998) Structure of α-lytic protease complexed with its pro region, *Nat. Struct. Biol. 5*, 945−950.
17. Mace, J. E., and Agard, D. A. (1995) Kinetic and structural characterization of mutations of glycine-216 in α-lytic protease: A new target for engineering substrate specificity, *J. Mol. Biol. 254*, 720−736.
18. Qasim, M. A., Ganz, P. J., Saunders, C. W., Bateman, K. S., James, M. N. G., and Laskowski, M., Jr. (1997) Interscaffolding additivity. Association of P-1 variants of eglin C and of turkey ovomucoid third domain with serine proteinases, *Biochemistry 36*, 1598−1607.
19. Lu, W., Izydor, A., Qasim, M. A., Warne, N., Wynn, R., Zhang, W.-L., Anderson, S., Chiang, Y.-W., Ogin, E., Rothberg, I., Ryan, K., and Laskowski, M., Jr. (1997) Binding of amino acid side-chains to S-1 cavities of serine proteinases, *J. Mol. Biol. 266*, 441−461.
20. Ash, E. L., Sudmeiser, J. L., De Fabo, E. C., and Bachovchin, W. W. (1997) A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment, *Science 278*, 1128−1132.
21. Empie, M. W., and Laskowski, M., Jr. (1982) Thermodynamics and kinetics of single residue replacements in avian ovomucoid 3rd domains: Effect on inhibitor interactions with serine proteinases, *Biochemistry 21*, 2274−2284.
22. Park, S. J. (1985) Effect of amino acid replacements in ovomucoid third domains upon their association with serine proteinases, Ph.D. Thesis, Purdue University, West Lafayette, IN.
23. Wynn, R. (1990) Design of a specifc human leukocyte elastase inhibitor based on vomucoid third domains, Ph.D. Thesis, Purdue University, West Lafayette, IN.
24. Lu, W. (1994) Energetics of the interactions of ovomucoid third domain variants with different serine proteinases, Ph.D. Thesis, Purdue University, West Lafayette, IN.
25. Chen, R., Li, L., and Weng, Z. (2003) ZDOCK: An initial-stage protein-docking algorithm, *Proteins 52*, 80−87.
26. Li, L., Chen, R., and Weng, Z. (2003) RDOCK: Refinement of rigid-body protein docking predictions, *Proteins 53*, 693−707.
27. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., and Vakser, I. A. (1992) Molecular-surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques, *Proc. Natl. Acad. Sci. U.S.A. 89*, 2195−2199.
28. Zhang, C., Vasmatzis, G., Cornette, J. L., and DeLisi, C. (1997) Determination of atomic desolvation energies from the structures of crystallized proteins, *J. Mol. Biol. 267*, 707−726.
29. Bauer, C. A., Brayer, G. D., Sielecki, A. R., and James, M. N. (1981) Active site of α-lytic protease: Enzyme−substrate interactions, *Eur. J. Biochem. 120*, 289−294.
30. Schechter, I., and Berger, A. (1967) On size of active site in proteases. I. Papain, *Biochem. Biophys. Res. Commun. 27*, 157−162.
31. Mitsui, Y., Satow, Y., Watanabe, Y., Hirono, S., and Iitaka, Y. (1979) Crystal structures of *Streptomyces* subtilisin inhibitor and its complex with subtilisin BPN, *Nature 277*, 447−452.
32. Hyberts, S. G., and Wagner, G. (1990) Sequence-specific H-1-NMR assignments and secondary structure of eglin C, *Biochemistry 29*, 1465−1474.
33. Hipler, K., Priestle, J. P., Rahuel, J., and Grutter, M. G. (1992) X-ray crystal structure of the serine proteinase inhibitor eglin C at 1.95 Å, *FEBS Lett. 309*, 139−145.
34. Heinz, D. W., Hyberts, S. G., Peng, J. W., Priestle, J. P., Wagner, G., and Grutter, M. G. (1992) Changing the inhibitory specificity and function of the proteinase inhibitor eglin C by site-directed mutagenesis: Functional and structural investigation, *Biochemistry 31*, 8755−8766.
35. Camacho, C. J., Weng, Z., Vajda, S., and DeLisi, C. (1999) Free energy landscapes of encounter complexes in protein−protein association, *Biophys. J. 76*, 1166−1178.
36. Rajamani, D., Thiel, S., Vajda, S., and Camacho, C. J. (2004) Anchor residues in protein−protein interactions, *Proc. Natl. Acad. Sci. U.S.A. 101*, 11287−11292.
37. Tian, M., Benedetti, B., and Kamoun, S. (2005) A second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato, *Plant Physiol. 138*, 1785−1793.

BI060445L