Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P1 residues

ARTICLE *in* PROTEIN SCIENCE · MAY 1993

Impact Factor: 2.85 · DOI: 10.1002/pro.5560020509 · Source: PubMed

8 AUTHORS, INCLUDING:

Richard Wynn
Incyte Corporation
39 PUBLICATIONS **1,731** CITATIONS

SEE PROFILE

Protein Science (1993), *2*, 786–799. Cambridge University Press. Printed in the USA.

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

THERESA L. BIGLER, WUYUAN LU, SOON JAE PARK,[1] MISAO TASHIRO,[2] MACIEJ WIECZOREK,[3] RICHARD WYNN,[4] AND MICHAEL LASKOWSKI, JR.

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907-1393

(RECEIVED July 30, 1992; REVISED MANUSCRIPT RECEIVED December 23, 1992)

Abstract

In the association of serine proteinases with their cognate substrates and inhibitors an important interaction is the fitting of the P_1 side chain of the substrate or inhibitor into a preformed cavity of the enzyme called the S_1 pocket. In turkey ovomucoid third domain, which is a canonical protein proteinase inhibitor, the P_1 residue is Leu^{18}. Here we report the values of equilibrium constants, K_a, for turkey ovomucoid third domain and 13 additional Leu18X variants with six serine proteinases: bovine α chymotrypsin A, porcine pancreatic elastase, subtilisin Carlsberg, *Streptomyces griseus* proteinases A and B, and human leukocyte elastase. Eight of the Xs are coded amino acids: Ala, Ser, Val, Met, Gln, Glu, Lys, and Phe, and five are noncoded: Abu, Ape, Ahx, Ahp, and Hse. They were chosen to simplify the interamino acid comparisons. In the homologous series of straight-chain side chains Ala, Abu, Ape, Ahx, Ahp, free energy of binding decreases monotonically with the side-chain length for chymotrypsin with large binding pocket, but even for this enzyme shows curvature. For the two *S. griseus* enzymes a minimum appears to be reached at Ahp. A minimum is clearly evident for the two elastases, where increasing the side-chain length from Ahx to Ahp greatly weakens binding, but much more so for the apparently more rigid pancreatic enzyme than for the more flexible leukocyte enzyme. β-Branching (Ape/Val) is very deleterious for five of the six enzymes; it is only slightly deleterious for the more flexible human leukocyte elastase. The effect of γ-branching (Ahx/Leu), of introduction of heteroatoms (Abu/Ser), (Ape/Hse), and (Ahx/Met), and of introduction of charge (Gln/Glu) and (Ahp/Lys) are tabulated and discussed. An important component of the free energy of interaction is the distortion of the binding pocket by bulky or branched side chains.

Most of the variants studied were obtained by enzymatic semisynthesis. X^{18} variants of the 6–18 peptide $GlyNH_2$ were synthesized and combined with natural reduced peptide 19–56. Disulfide bridges were formed. The $GlyNH_2$ was removed and the reactive-site peptide bond X^{18}–Glu^{19} was synthesized by complex formation with proteinase K. The resultant complexes were dissociated by sudden pH drop. This kinetically controlled dissociation afforded virgin, reactive-site-intact inhibitor variants.

Keywords: binding of amino acid side chains; deformability of binding pockets; enzymatic semisynthesis; noncoded amino acid residues; ovomucoid third domains; serine proteinases

Reprint requests to: Michael Laskowski, Jr., Department of Chemistry, Purdue University, West Lafayette, Indiana 47907-1393.

[1] Present address: Lucky Central Research Institute, Science Town Dae Jeon, Korea.

[2] Present address: Kyoto Prefectural University, Shimogamo, Kyoto 606, Japan.

[3] Present address: CORTECH, Inc., Denver, Colorado 80301.

[4] Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511.

Abbreviations: P_1, the residue contributing the carbonyl portion to the reactive-site peptide bond; S_1, the pocket in the enzyme to which the P_1 residue binds; OMXXX3, ovomucoid third domain, with XXX designating the species; TKY, turkey; SVP, silver pheasant; SWN, black swan; WTD, West Indian tree duck; MNQ, Montezuma quail; X^{18}OMTKY3, X residue replaces Leu^{18} in OMTKY3 (Fig. 1); Abu α, aminobutyric acid; Ape α, aminopentanoic acid (norvaline); Ahx α, aminohexanoic acid (norleucine); Ahp α, aminoheptanoic acid; Hse α, amino, γ-hydroxybutyric acid (homoserine); CHYM, bovine chymotrypsin Aα; PPE, porcine pancreatic elastase; SUBT, subtilisin Carlsberg; SGPA, *Streptomyces griseus* proteinase A; SGPB, *S. griseus* proteinase B; HLE, human leukocyte elastase; BPTI, bovine pancreatic trypsin inhibitor (Kunitz); SSI, *Streptomyces* subtilisin inhibitor; SucAAPX[1] pNA, succinyl-alanyl-alanyl-prolyl-X-*p*-nitroanilide; BocAAXSBzl, butyloxycarbonyl-alanyl-alanyl-X-thiobenzylester; AcXOMe, acetyl-X-methylester.

The work of Nozaki and Tanford (1971) focused the attention of protein chemists on the importance of understanding the free energy of transfer of amino acid side chains from water to the organic phase. Recently, this topic was the subject of considerable controversy (e.g., Sharp et al., 1991; Holtzer, 1992). Experimental protein chemists encounter the side-chain transfer in two contexts. One is in the studies of protein stability where the intramolecular transfer of side chains from the surface to the interior is analyzed (e.g., Eriksson et al., 1992). The other is in the binding of peptide and protein ligands to other proteins containing surface cavities for the accommodation of amino acid side chains of the ligands. The binding of P_1 residues (Schechter & Berger, 1967) of substrates and inhibitors to the S_1 pockets of serine proteinases is probably the most widely known example of such a process and is the subject of this paper. Processes of this type were analyzed by Fersht (1977, 1985). In the ligand-binding situations the preformed cavities (Dill, 1990) are filled with water and some water may remain after a side chain is bound (Fujinaga et al., 1987). In the interior of proteins the cavities are often empty (Eriksson et al., 1992). Thus the two contexts may be thermodynamically somewhat different.

Many protein inhibitors of serine proteinases interact with their cognate enzymes according to the "standard mechanism" (Laskowski & Kato, 1980; Laskowski, 1986; Read & James, 1986; Bode & Huber, 1991, 1992). In these inhibitors the reactive site of the inhibitor, which in a Schechter and Berger (1967) notation is designated P_1–P_1', interacts with the enzyme in substrate-like manner, although it is intact in the enzyme–inhibitor complex. The P_1 residue fits into the S_1 pocket of the enzyme making many contacts with it (see Kinemage 1). This residue is often called the primary specificity residue of the inhibitor. When this residue is Lys or Arg we anticipate inhibition of trypsin; when it is Trp, Tyr, Phe, Leu, or Met, chymotrypsin inhibition is expected. The influence of the nature of this residue is often exaggerated by some workers, who assert that it is the only 1 of the about 12 residues of the inhibitor involved in enzyme inhibitor contact that seriously affects enzyme–inhibitor interaction. This is quite false; however, the P_1 residue is predominant for interaction with all the serine proteinases we have tested, except for subtilisin, which in fact has an S_1 cleft not an S_1 pocket.

Our laboratory is engaged in a detailed study of effects of amino acid residue replacements in the turkey ovomucoid third domain (Fig. 1) upon binding to various enzymes. It is our aim (Laskowski, 1980) to state a sequence to reactivity algorithm—a set of rules (and data) that, subject to stated approximations, will allow us to predict the binding constants, K_a, to various enzymes from the amino acid sequence of the inhibitor alone.

In this paper we deal with changes at the P_1 position in an effort to understand qualitatively and quantitatively

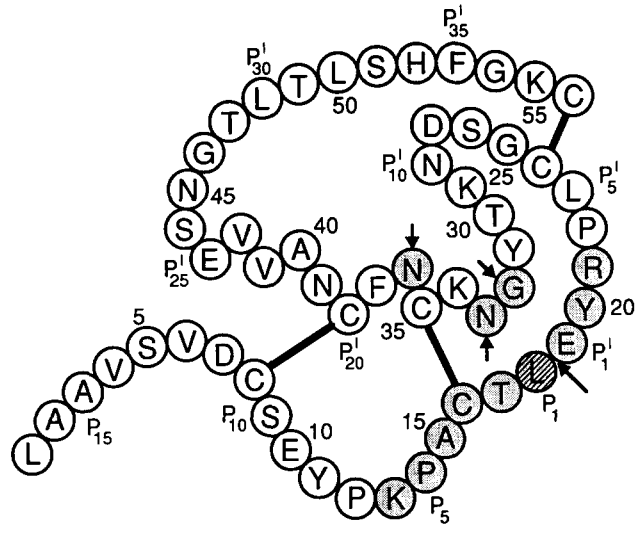

Fig. 1. The amino acid sequence of turkey ovomucoid third domain, OMTKY3 (−) (Laskowski et al., 1987). The numbering system is based on Weber et al. (1981). The reactive-site peptide bond between residues 18 and 19 or P_1 and P_1' in Schechter–Berger (1967) notation is indicated by an arrow. The residues comprising the consensus set of enzyme–inhibitor contact residues (Laskowski et al., 1990) are in gray. The P_1 residue Leu^{18} is hatched. All of the substitutions described in this paper are at that position. Most semisynthetic variants employed here are not full-length 1–56 but are NH_2-terminal truncated 6–56 (see text for details).

what controls the binding to the S_1 pocket or cleft. As most of the variants we employed are semisynthetic (Sealock & Laskowski, 1969; Wieczorek & Laskowski, 1983; Wieczorek et al., 1987; Bigler, 1991), we did not choose the very important set of all the coded amino acids at P_1, as was done by Kojima et al. (1991), who studied the interaction of SSI variants, with subtilisin. Instead, we have chosen a mix of coded and of noncoded residues with the hope of best clarifying the effects of various changes upon binding. It is well known (e.g., Doolittle, 1979) that the set of all coded amino acids involves very complex relations between members and that increases in side-chain size are accompanied by changes in branching (Val, Leu), presence of heteroatoms (Ser, Gln, Met), or presence of heteroatoms leading to a change in charge (Lys, Glu). We have selected a homologous (in a chemical sense) series Ala, Abu, Ape, Ahx, and Ahp. We have also added Hse. These unnatural or noncoded members can be usefully compared not only to each other, but also to many coded residues (Fig. 3; Table 3). It should be noted that we are also not the first to introduce noncoded amino acids at the P_1 position of a protein inhibitor of serine proteinases. This honor belongs to Tschesche's group (Tschesche et al., 1987; Beckmann et al., 1988, 1989), who introduced many of the same noncoded residues into the P_1 position of BPTI. However, our set, where K_a values were obtained over a very broad range for six different enzymes, serves as a good basis for comparison with the Kojima

et al. (1991) and the Tschesche et al. (1987) data and with selected k_{cat}/K_m data on substrates reported by other workers.

Results

Table 1 is a list of K_a values for 14 P_1 variants (X^{18} OMTKY3). The superscripted X^{18} entries are based on direct measurements on semisynthetic X^{18}OMTKY3 variants consisting of 51 residues, i.e., residues 6–56 of Figure 1. The unsuperscripted Leu^{18}, Met^{18}, and Hse^{18} entries are based on measurements of variants consisting of 56 residues (residues 1–56 of Fig. 1). The Leu^{18} and Met^{18} are turkey ovomucoid third domain, OMTKY3, and silver pheasant ovomucoid third domain, OMSVP3, obtained by hydrolysis of the ovomucoids from these species by staphylococcal proteinase V8. Hse^{18} is a product of chemical modification of OMSVP3 and thus also consists of 56 residues. Wieczorek et al. (1987) have shown that K_a values obtained on 1–56 and 6–56 variants are indistinguishable. From now on, the difference in NH_2-terminal peptide length will be neglected.

Additivity-based calculations

The values for Ser^{18}, Val^{18}, and Phe^{18} were calculated from data on variants containing the Val^{18}, Phe^{18}, and Ser^{18} at their P_1 position but also differing from OMTKY3 at several other positions. To appreciate an additivity-based calculation, first consider cycle A in Figure 2. The four natural variants in this cycle are arranged in a rectangle so that the differences in sequence between adjacent variants lying on parallel sides of the rectangle are identical. For the horizontal sides the differences are $Leu^{18} \leftrightarrow Met^{18}$. If we assume by additivity that the relative change in K_a upon the $Leu^{18} \leftrightarrow Met^{18}$ substitution is independent of the rest of the sequence, then we can calculate K_a for any one of the four variants from the K_a values for the remaining three. These types of cycles were first introduced by Laskowski et al. (1983). Additivity is now widely used in assessing residue-residue interactions in proteins (Horovitz et al., 1991), in analyzing the properties of closely related proteins, and in protein design (see Wells [1990] for a review). It should be noted that whereas most substitutions in proteins seem to be additive, large non-additive effects are both observed and expected. Additivity of K_a values for ovomucoid third domains (Laskowski et al., 1989; Wells, 1990; Wynn, 1990) is greater than for other proteins, but again exceptions do occur. Thus, if possible, additivity should be avoided in calculating data such as given in Table 1.

However, the situation is not always so simple. Because of the enormous range of K_a values (see Table 1 and Discussion), using all of the direct variants is likely to impose great experimental difficulties on the K_a measurement. The Phe^{18}OMSWN3 variant was intentionally prepared in order not to force direct measurement of Phe^{18}OMTKY3 interaction with chymotrypsin and with porcine elastase. It can be seen in Table 2 that in the case of chymotrypsin, K_a for Phe^{18}OMSWN3 is much smaller than for Phe^{18}OMTKY3 (calculated), thus making the measurement easier. We had originally anticipated an even greater K_a for Phe^{18}OMTKY3. Similarly the Phe^{18} OMTKY3 K_a for PPE is quite low (we anticipated lower), and the use of Phe^{18}OMSWN3 allowed us to measure a larger value. Unless K_a measurement methods with K_a ranges even greater than our current 10^9 are developed,

Table 1. *Association equilibrium constants, K_a (M^{-1}) for X^{18}OMTKY3 interaction with six serine proteinases (21 ± 2 °C, pH 8.30)*[a]

X^{18}	CHYM	PPE	SUBT	SGPA	SGPB	HLE
Ala[b]	4.6×10^7	3.5×10^{10}	1.4×10^{10}	1.9×10^9	3.8×10^8	1.2×10^9
Abu[b]	1.1×10^9	$\underline{3.3 \times 10^{11}}$	1.0×10^{11}	1.0×10^{10}	2.3×10^9	$\underline{1.4 \times 10^{10}}$
Ape[b]	2.2×10^{10}	2.4×10^{11}	$\underline{1.4 \times 10^{11}}$	8.3×10^{10}	1.2×10^{10}	1.0×10^{10}
Ahx[b]	8.0×10^{10}	1.9×10^{11}	6.7×10^{10}	2.3×10^{11}	2.1×10^{10}	9.9×10^8
Ahp[b]	3.3×10^{11}	2.4×10^8	8.9×10^{10}	$\underline{4.7 \times 10^{11}}$	2.8×10^{10}	1.3×10^8
Phe[c]	$\underline{2.1 \times 10^{12}}$	$\underline{2.9 \times 10^4}$	1.0×10^{10}	1.2×10^{11}	1.6×10^{10}	1.1×10^6
Val[c]	1.8×10^8	3.1×10^9	1.3×10^9	2.3×10^9	3.6×10^8	2.3×10^9
Leu	1.8×10^{11}	4.1×10^{10}	4.4×10^{10}	2.9×10^{11}	$\underline{5.7 \times 10^{10}}$	6.2×10^9
Met	1.1×10^{11}	1.0×10^{10}	8.1×10^{10}	2.1×10^{11}	2.5×10^{10}	7.7×10^8
Ser[c]	3.5×10^7	1.3×10^9	1.8×10^9	2.0×10^8	3.9×10^7	ND
Hse	2.5×10^9	7.5×10^9	5.6×10^{10}	6.5×10^9	1.7×10^9	ND
Glu[b]	$\underline{1.8 \times 10^6}$	9.0×10^4	9.4×10^8	$\underline{6.7 \times 10^6}$	$\underline{1.2 \times 10^6}$	$\underline{4.0 \times 10^4}$
Lys[b]	1.4×10^8	ND	$\underline{2.8 \times 10^8}$	5.6×10^7	2.6×10^8	4.4×10^5
Gln[b]	4.7×10^8	8.0×10^7	1.2×10^{10}	5.8×10^8	8.4×10^8	2.0×10^7

[a] The largest K_a value for each enzyme is doubly underlined; the smallest is singly underlined. ND, not determined.
[b] Obtained by K_a measurements on semisynthetic 6–56 OMTKY3 variants.
[c] Obtained by calculations based on additivity assumptions. See Figure 2, Table 2, and text.

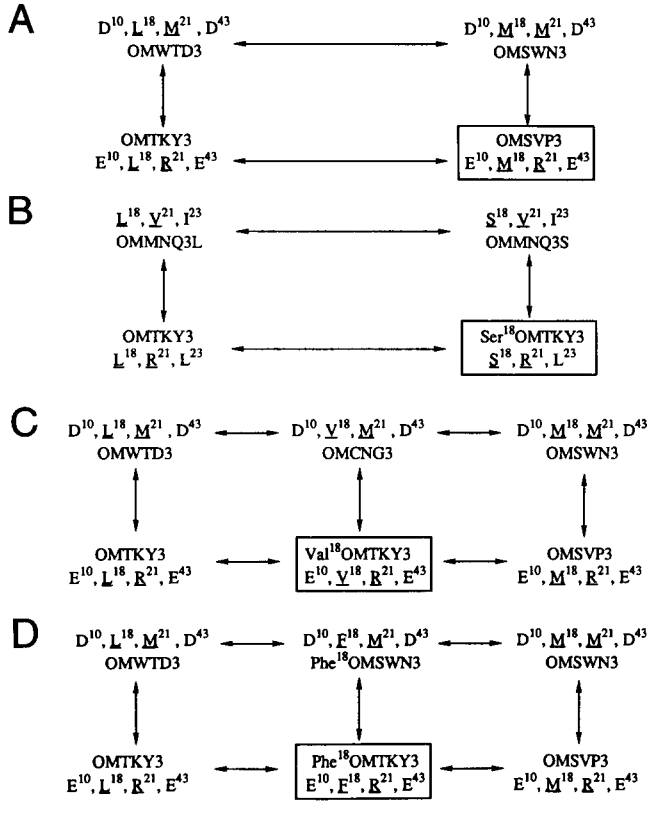

Fig. 2. Additivity cycles used in this work. **A:** OMWTD3 and OMSWN3 differ only by $L^{18} \leftrightarrow M^{18}$ substitution as do OMTKY3 and OMSVP3. The differences between OMWTD3 and OMTKY3 are also indicated, and the contact residue, which is likely to be the one responsible for differences in K_a values, is underlined. Equilibrium constants were measured for all four variants. However, K_a for any one of them, in this case the circled OMSVP3, can be calculated from the values for the other three. The results are shown in Table 2. **B:** The values for Ser^{18} OMTKY3 (not available) are calculated from the three available variants. **C:** The needed value for Val^{18}OMTKY3 can be calculated from either of two closely related additivity cycles. The average is actually used in Table 1. **D:** Analogous to C except that Phe^{18}OMSWN3 is an intentionally constructed semisynthetic variant (see text for details).

additivity may well have to be employed to make a complete analysis of the situation.

Relative K_a values

In order to better deal with the data of Table 1, entries in each column were divided by the K_a value for the Ala^{18} variant generating relative K_a values. Within this data set the choice of Ala^{18} as the standard seems obvious. The relative equilibrium constants were then converted to $-\Delta\Delta G^0$ (A18X) values and plotted in Figure 3. We have spent considerable effort choosing the x-axis. A popular choice is the free energy of transfer, but as is clearly seen in Figure 3 for the two enzymes with small pockets, large residues fit poorly, and this is unrelated to the free energy of transfer. We feel that the very simple and noncontroversial number of nonhydrogen atoms in the side chain is the simplest property to employ. It is particularly useful for the homologous series $Ala^{18} \rightarrow Ahp^{18}$.

Discussion

The series Ala^{18}, Abu^{18}, Ape^{18}, Ahx^{18}, and Ahp^{18} is a homologous series of mostly noncoded amino acids. In order to highlight the data for this series the points corresponding to these residues are connected to each other by solid line segments in Figure 3. Inspection of the other points in Figure 3 indicated to us that Phe^{18} is located almost precisely where we would have extrapolated the homologous series; therefore we connected Ahp^{18} to Phe^{18} by a dotted line segment. (This may have been risky, please beware.)

We now make a qualitative examination of the behavior of this series in the six panels of Figure 3. For chymotrypsin increasing the number of carbons in the side chain leads to an increase in binding up to Phe^{18} and presumably beyond (Tyr^{18} and Trp^{18}). This is consistent with the improvement in binding with rising hydrophobicity of the chain. However, even here the negative of the relative binding free energy is not linear with the number of nonhydrogen atoms. A curvature for higher members is apparent.

The results for SGPA and especially SGPB show greater curvature so that the increase between Ahx^{18} and Ahp^{18} is very small and binding actually declines slightly for Phe^{18}. These results are consistent with smaller binding pockets of SGPA and SGPB compared to chymotrypsin.

It is well known (e.g., Bode et al., 1989) that the binding pockets of the elastases are much smaller than those of chymotrypsin, SGPA, and SGPB. The binding data confirm this. For both HLE and PPE, binding improves in the $Ala^{18} \rightarrow Abu^{18}$ step and is largely unaffected in the $Abu^{18} \rightarrow Ape^{18}$ step, where a maximum appears to have been reached. For $Ape^{18} \rightarrow Ahx^{18}$ there is only a slight decline for PPE, but a pronounced one for HLE. Thus PPE seems to exhibit a broader maximum than HLE. Beyond Ahx^{18}, binding declines for both enzymes. The residues are too large for the binding pocket, and binding appears to require distortion of the pocket. Note that residues that are much too large for the empty binding pocket can still be bound, but with a much weaker binding constant. It is also worth noting that $-\Delta\Delta G^0$ declines much more steeply for PPE than for HLE. We take it as evidence that the S_1 pocket of HLE is more easily distorted than the PPE pocket. Such a conclusion was also reached in X-ray crystallographic studies (Bode et al., 1989) and is confirmed by several other comparisons in this paper (see below).

In the case of subtilisin, the range of values is very small compared to that of the other five enzymes, presumably because of the presence of an S_1 cleft rather than an S_1 pocket.

Table 2. *K_a (M^{-1}) calculations based on additivity assumptions*[a]

	CHYM	PPE	SUBT	SGPA	SGPB	HLE
OMWTD3	2.6×10^9	3.9×10^{11}	1.7×10^{10}	3.9×10^{11}	4.9×10^{10}	2.2×10^{10}
OMTKY3	1.8×10^{11}	4.1×10^{10}	4.4×10^{10}	2.9×10^{11}	5.7×10^{10}	6.2×10^9
OMSWN3	1.9×10^9	1.2×10^{11}	5.1×10^{10}	2.5×10^{11}	2.5×10^{10}	1.5×10^9
OMMNQ3L	4.9×10^9	8.7×10^9	3.5×10^{10}	3.2×10^{11}	4.2×10^{10}	1.0×10^{10}
OMMNQ3S	9.6×10^5	2.8×10^8	1.4×10^9	2.2×10^8	2.9×10^7	ND
OMCNG3	2.7×10^6	3.2×10^{10}	6.1×10^8	2.8×10^9	3.3×10^8	5.7×10^9
F^{18}OMSWN3	3.3×10^{10}	3.0×10^5	5.0×10^9	1.5×10^{11}	1.5×10^{10}	2.8×10^6
OMSVP3						
Calculated	1.3×10^{11}	1.3×10^{10}	1.3×10^{11}	1.9×10^{11}	2.9×10^{10}	4.2×10^8
Determined	1.1×10^{11}	1.0×10^{10}	8.1×10^{10}	2.1×10^{11}	2.5×10^{10}	7.7×10^8
S^{18}OMTKY3						
Calculated	3.5×10^7	1.3×10^9	1.8×10^9	2.0×10^8	3.9×10^7	NA
V^{18}OMTKY3						
Calculated	1.9×10^8	3.4×10^9	1.6×10^9	2.1×10^9	3.8×10^8	1.6×10^9
Calculated	1.6×10^8	2.7×10^9	9.7×10^8	2.4×10^9	3.3×10^8	2.9×10^9
Average	1.8×10^8	3.1×10^9	1.3×10^9	2.3×10^9	3.6×10^8	2.3×10^9
F^{18}OMTKY3						
Calculated	2.3×10^{12}	3.2×10^4	1.3×10^{10}	1.1×10^{11}	1.7×10^{10}	7.9×10^5
Calculated	1.9×10^{12}	2.5×10^4	7.9×10^9	1.3×10^{11}	1.5×10^{10}	1.4×10^6
Average	2.1×10^{12}	2.9×10^4	1.0×10^{10}	1.2×10^{11}	1.6×10^{10}	1.1×10^6

[a] Please see Figure 2 to follow the comparisons involved. ND, not determined; NA, not available.

Turning to a quantitative analysis, the initial slope of the six panels in Figure 3, i.e., $-\Delta\Delta G^0$ (Ala 18 Abu) is 1.85 kcal/mol, 0.97, 1.05, 1.31, 1.43, and 1.14 kcal/mol for CHYM, SGPA, SGPB, PPE, HLE, and SUBT, respectively. These values can be compared to the free energy of transfer of a methylene group in a side chain from H_2O to an organic solvent. However, Fersht (1977, 1985) pointed out that in this type of a reaction not only the P_1 side chain of the inhibitor, but the sides of the S_1 binding pocket of the enzyme are removed from contact with water, and therefore the effect should be larger than just the side chain-to-organic solvent transfer. The situation is now complicated by the controversy about the proper method of evaluating the free energy of hydrophobic transfer (Sharp et al., 1991; Holtzer, 1992). Therefore, we do not go on with further calculations but simply provide model-free data. It is worth noting that the 1.85 kcal/mol for chymotrypsin is larger than even the Sharp et al. (1991) estimate.

The $Ahx^{18} \rightarrow Ahp^{18}$ changes in $-\Delta\Delta G^0$ for PPE and HLE are −3.9 kcal/mol and −1.1 kcal/mol, respectively. If the extension to Phe^{18} is allowed, then for Ahx^{18} (five nonhydrogen atoms) to Phe^{18} (seven nonhydrogen atoms) they are −2.6 kcal/mol per CH_2 group for PPE and −1.4 kcal/mol per CH_2 for HLE. These results point out that the free energy of pocket distortion by the binding side chains is a very large contributor to the total value of $-\Delta\Delta G^0$. Further, they show that the HLE pocket accepts distortions by paying a two to three times smaller energy price than the PPE pocket. Similar effects are seen below in the case of branching.

Range

Table 3 is a listing of various especially interesting comparisons between residues. Most of these are made possible by the availability of noncoded amino acid residues within the set. The first of these comparisons is the range, i.e., the ratio of the highest to the lowest K_a value in Table 1. The range is quite small (500-fold) for subtilisin, but very large (from 4.8×10^4 for SGPB to 1.1×10^7 for PPE) for the enzymes of the chymotrypsin superfamily. The low range for subtilisin is presumably explained by the S_1 binding cleft (Estell et al., 1986) rather than S_1 binding pocket in the other enzymes. The high range for the other enzymes arises from the combination of positive hydrophobic interactions and of negative interaction due to steric bulk, branching, and charge. The strong interactions of the remainder of the inhibitor molecule with the enzyme force the negative P_1–S_1 interactions, such as binding of the Glu side chain.

Branching

We have known for some time (Laskowski et al., 1983) that replacement of Leu^{18} by Val^{18} (see Fig. 2B and appropriate entries in Table 2) decreases K_a values for chymotrypsin by a factor of about 10^3. This dramatic change (~4 kcal/mol) is far too large to be due to an addition of a single methylene group. It was apparent that the huge effect was a consequence of β-branching (Val^{18}) being far more deleterious to binding than γ-branching (Leu^{18}). More careful examination of this effect was an additional

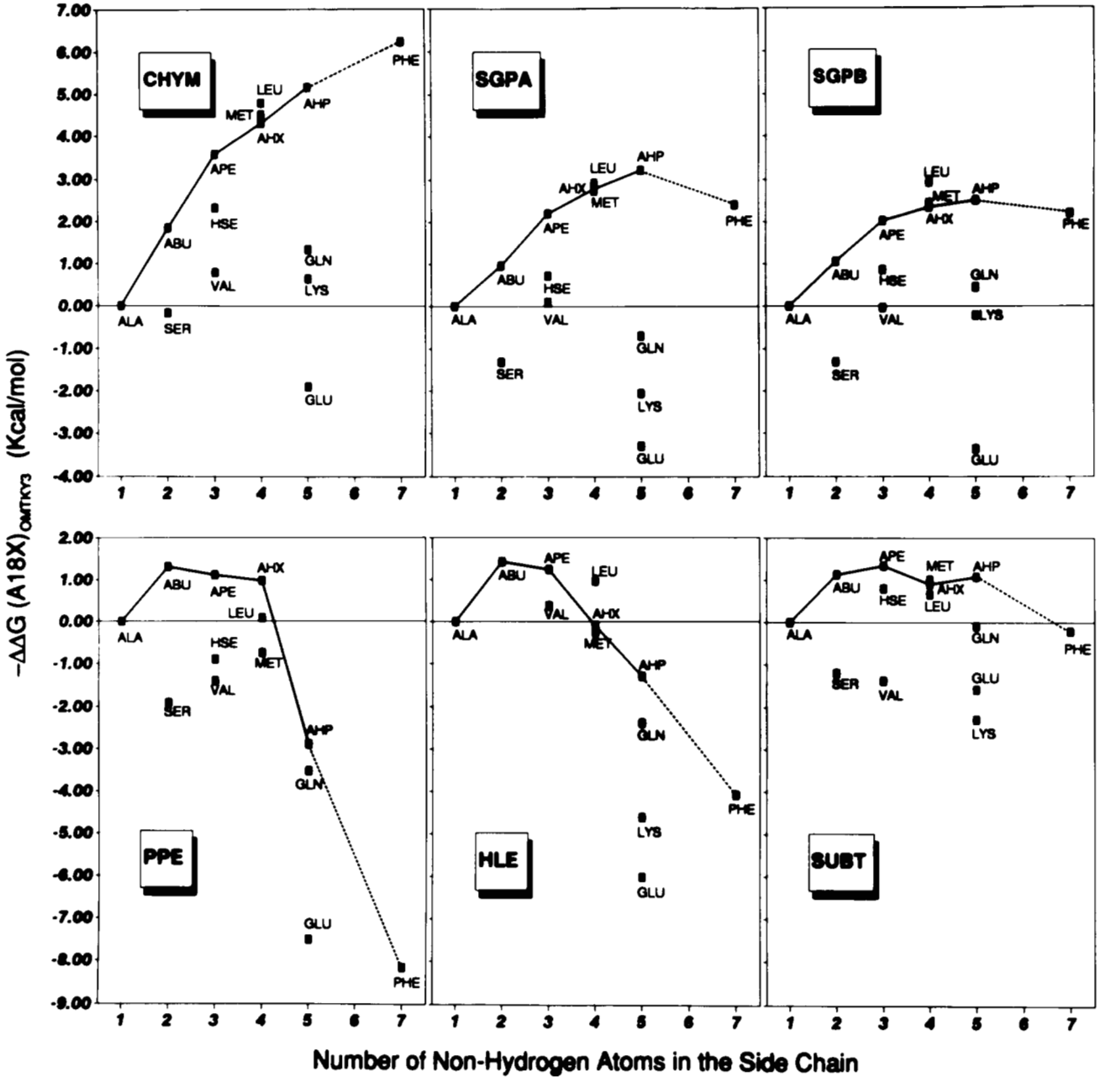

Fig. 3. The values of $-\Delta\Delta G^0$ (Ala18X) at 21 °C, pH 8.30, for the various side chains. All data were calculated from Table 1 (see text). CHYM, bovine pancreatic chymotrypsin Aα; PPE, porcine pancreatic elastase; HLE, human leukocyte elastase; SGPA and SGPB, *Streptomyces griseus* proteinases A and B, respectively; SUBT, subtilisin Carlsberg.

motivation for preparing the series of noncoded variants reported here. Table 3 contains ratios of K_a values for Ape^{18} and Val^{18} and of Ahx^{18} and Leu^{18}. The inverses of these numbers can be viewed as "the price of β-branching" and "the price of γ-branching," respectively. It can be seen that β-branching is unfavorable for all enzymes we have studied; however, it is far less unfavorable for HLE. In contrast, γ-branching effects are very small, but generally they lead to an increase rather than a decrease in binding. The Ahx^{18}/Leu^{18} ratio is larger than unity for PPE (branching is bad), but this comparison involves Ahx^{18}, which is involved in the anomalous Ahx^{18}/Met^{18} comparison in the case of PPE. It is thus clear that most of the decrease in K_a for chymotrypsin upon Leu^{18} → Val^{18} substitution is due to the change between the slightly favorable γ-branching and the very unfavorable β-branching. Only a small part of the effect is due to the chain extension (compare Ahx^{18} with Ape^{18}).

HLE is very tolerant of β-branching, and it has been known for some time that P_1 Val and Ile are the best sub-

Table 3. *Comparison of effects of side chains of some residues at P_1*[a]

	CHYM	PPE	SUBT	SGPA	SGPB	HLE
Range	1.2×10^6	1.1×10^7	5.0×10^2	7.0×10^4	4.8×10^4	3.5×10^5
Ape/Val	1.2×10^2	7.7×10^1	1.1×10^2	3.6×10^1	3.3×10^1	4.3
Ahx/Leu	4.4×10^{-1}	4.6	1.5	7.9×10^{-1}	3.7×10^{-1}	1.6×10^{-1}
Ahx/Met	7.3×10^{-1}	1.9×10^1[b]	8.3×10^{-1}	1.1	8.4×10^{-1}	1.3
Abu/Ser	3.1×10^1	2.5×10^2	5.6×10^1	5.0×10^1	5.9×10^1	ND
Ape/Hse	8.8	3.2×10^1	2.5	1.3×10^1	7.0	ND
Ahp/Lys	2.4×10^3	ND	3.2×10^2	8.4×10^3	1.1×10^2	3.0×10^2
Gln/Glu	2.6×10^2	8.9×10^2	1.3×10^1	8.7×10^1	7.0×10^2	5.0×10^2

[a] ND, not determined.
[b] 2.9×10^1 For BPTI (Beckmann et al., 1989).

stitutions that can be made at P_1 to change protein inhibitors of other enzymes into the strongest or most specific (Laskowski et al., 1989) HLE inhibitors. The especially large ability of the HLE S_1 pocket to accommodate β-branching is further evidence of much greater deformability of this pocket (see above and Bode et al., 1989) than those of other enzymes.

Heteroatoms: The Ahx/Met comparison

Most protein chemists (e.g., Anfinsen & Corley, 1969) have long assumed that, except in highly unusual situations, the $-CH_2-$ group and the -S- group should be viewed as isofunctional. Therefore the effect of introduction at P_1 of Ahx and of Met should be identical. As is seen in Table 3 this is strikingly confirmed for five of the six enzymes, where the ratios are unity within experimental error. Whenever available, all of the data from the work of others quoted below confirm this conclusion.

However, in sharp contrast the results for PPE show that Ahx is much better than Met. We found this surprising enough to repeat the Ahx[18]OMTKY3 semisynthesis and to redetermine its K_a values, but no significant differences were observed. We have also used the competitive displacement method (Empie & Laskowski, 1982) to compare the K_a ratio of Ahx[18]OMTKY3 to that of OMSVP3 (natural variant with P_1 Met[18]). We have found that the K_a ratio was greater than 10. Strikingly, Beckmann et al. (1989) found that in the bovine pancreatic trypsin inhibitor (BPTI) system, K_a for Ahx[15]BPTI inhibiting PPE is 29 times greater than that for Met[15]BPTI inhibiting the same enzyme. On the other hand k_{cat}/K_m values for BocAAXSBzl show that Ahx is only twice better for PPE than Met (Harper et al., 1984). At present we have no explanation for the Ahx/Met anomaly for PPE (see also Table 5).

The $-CH_3$ by -OH replacement

The $-CH_3$ by -OH replacement would be expected to weaken binding as the -OH group is much less hydrophobic than the $-CH_3$ group. Indeed, for all five enzymes for which we have data, the Abu/Ser ratio is somewhat larger than we would have anticipated. The Ape/Hse ratio is also greater than unity for all five enzymes we have studied, but in every case it is significantly smaller than the Abu/Ser ratio. We suspect that this is a real effect, and that replacements at γ positions affect K_a more than replacements at δ position.

Effect of side-chain charge

All six enzymes prefer neutral side chains, therefore, we expected and found that charged side chains will be avoided. This is revealed by the isosteric comparisons Ahp/Lys and Gln/Glu, all of which show that charged side chains are strongly avoided by all the enzymes.

The Ahp/Lys ratio and especially the Gln/Glu ratio are quite small for subtilisin probably because subtilisin has a more open S_1 cleft (Estell et al., 1986) rather than the closed S_1 pocket of the other five enzymes. The more open structure should interfere less with hydration of charged P_1 side chains. Bode et al. (1989) remarked on Asp^{226}, which is buried in the interior of HLE near the bottom of the HLE pocket. As the charge of Asp^{226} is not balanced by any counterions, it might predispose to strong binding of positively charged P_1. This is not the case, presumably because Asp^{226} of HLE is shielded from the pocket by the side chains of Val^{190} and Val^{216}. However, the Ahp/Lys ratio for HLE is an order of magnitude smaller than for chymotrypsin.

The various ratios in Table 3 are quite similar for the two closely related *S. griseus* proteinases. Therefore, the large differences in Gln/Glu ratio and especially in the Ahp/Lys ratio deserve comment. We have already noted the relatively strong preference of SGPB over SGPA for Lys (Park, 1985; Wynn, 1990).

The above analysis of Table 3 appears to provide some reasonably good rationalizations for the trends in the K_a values. However, we wish to caution both ourselves and the readers that they are mainly a posteriori rationalizations and that they are not likely to carry large predictive

powers for additional P_1 residues. Whether detailed X-ray crystallographic analysis of many more cases would improve the predictive power remains to be seen.

Isofunctional residues

Ovomucoid third domains are in contact with their cognate enzymes at about 12 consensus contact positions (marked in Fig. 1). These positions are tabulated in Laskowski et al. (1990) based on three-dimensional structures of complexes of turkey ovomucoid third domain with SGPB (Fujinaga et al., 1982; Read et al., 1983), chymotrypsin (Fujinaga et al., 1987), and HLE (Bode et al., 1986). A reasonably good argument can be made that these contact positions remain essentially the same for all cognate serine proteinases and for all natural avian ovomucoid third domains. We have now isolated from 152 species (Laskowski et al., 1987, 1990, unpubl.) ovomucoid third domains with 90 unique sequences. We have measured the K_a values for interaction of most of these with the six enzymes studied here (Empie & Laskowski, 1982; Park, 1985; Wynn, 1990). Analysis of these data leads to the somewhat anticipated conclusion that changing of the noncontact residues generally has little effect on K_a values, whereas changing the contact residues generally results in very large changes in K_a values. What is more surprising, however, is that in 20–30% of the cases, changes in contact residues produce only very small changes in K_a (Laskowski et al., 1988, 1989; Wynn, 1990, unpubl.). We call such residue sets isofunctional. Sometimes the members of such residue sets are very similar, but at other times they are dramatically dissimilar. For example, Ala, Val, and Lys are isofunctional of SGPB (Table 1; Fig. 3). This paper provides an excellent explanation for dissimilar residues being isofunctional by emphasizing and illustrating the opposing effects of hydrophobicity on one hand and steric bulk, branching, presence of heteroatoms, and charge on the other.

It is clear that sets of residues may be isofunctional at one contact position and not at another and that they can be isofunctional for one enzyme and not for another. Our interest in isofunctional residues is predominantly based on our concern with the hypervariability of contact position in most protein inhibitors of serine proteinases (see Laskowski & Kato, 1980; Laskowski, 1986; Creighton & Darby, 1989) and especially in ovomucoid third domains (Laskowski et al., 1987, 1988, 1990). Among various possible explanations the observed hypervariability may be either a result of neutral drift among isofunctional residues or of positive Darwinian selection. For ovomucoid third domains, a predominant role of drift among isofunctional residues is unlikely as the K_a values for the six enzymes we study vary by many orders of magnitude among third domains from natural species. However, drift among isofunctional residues may well be an explanation for active or reactive-site hypervariability, where the analysis is solely based on measuring variability and not on its effect on biological properties (e.g., Creighton & Darby [1989] and many other papers).

Another concern about the existence of isofunctional residues arises for workers who try to find the residues comprising the epitope—the combining region of the antigen involved in antigen–antibody association—by comparing the association of a series of homologous protein antigens (especially if the three-dimensional structure of one member is known) and inferring the epitope from the observed changes in binding (for reviews see Benjamin et al. [1984] and Davies et al. [1988]). This method has been greatly extended to protein–protein association by the widely used alanine scan of Cunningham and Wells (1989) where a large set of mutants of one of the partner proteins is prepared, with systematic one-at-a-time replacements of existing residues by Ala. If an Ala replacement causes a big change in binding the position is judged to be a part of the combining region. In a majority of cases these are superb procedures giving excellent results. However, in some cases, isofunctional residue pairs may be encountered, and a position may be mistakenly excluded from the epitope even though it belongs there.

Possible upper limit on the best P_1 residue

As protein engineers, we are quite interested in knowing what would be the largest $-\Delta\Delta G^0$ (Ala 18X) for any possible residue, X. If the set of all residues is the coded set, the answer becomes available after an exhaustive search of all possible residues. This was already done in SSI by Kojima et al. (1991) and the effects were measured for subtilisin BPN′. They report Ala^{73}, Leu^{73}, and Phe^{73} as the three best residues.

However, the more interesting question is which is the "best" residue in the set of both coded and noncoded residues. For such a large and poorly defined set an exhaustive search becomes impractical. However, it appears from Figure 3 and from the analysis of the data presented above that all branched or substituted side chains bind to our enzymes more weakly than the straight-chain side chains of appropriate length. This is not a very firm conclusion. γ-Branching is sometimes slightly beneficial. The highly deformable pocket of HLE may surprise us. However, it seems likely that we cannot do much better than the homologous aliphatic straight chains. Note that in Figure 3, almost all of the points lie below the homologous straight chain results.

Other enzymes

Thus far we discussed the binding of the 14 P_1 variants to six enzymes with a strong preference for neutral P_1 residues. However, we have also measured the interaction of X^{18}OMTKY3 with glutamic acid-specific *S. griseus* proteinase, GluSGP (Komiyama et al., 1991). The K_a

values for X^{18} Glu, Gln, Met, and Leu are $5.5 \times 10^{10}\ M^{-1}$, $7.2 \times 10^{6}\ M^{-1}$, $7.2 \times 10^{4}\ M^{-1}$, and $7.7 \times 10^{3}\ M^{-1}$, respectively. In this case Glu is the best, rather than the worst of the X^{18} residues. Over the limited sample of X^{18} residues, the range ($\sim 10^7$) is similar to that for the other enzymes. We also measured the K_a value for bovine β-trypsin for X^{18} Lys, Met, and Leu and they are $3.2 \times 10^{8}\ M^{-1}$, $1.0 \times 10^{5}\ M^{-1}$, and $6.4 \times 10^{3}\ M^{-1}$, respectively (the last two values are as yet unpublished results by Komiyama). Because of lack of material we did not try Lys^{18}OMTKY3 as an inhibitor of GluSGPB nor Glu^{18} OMTKY3 as an inhibitor of bovine β-trypsin. These might have greatly increased the range for both enzymes, but massive amounts of material are needed to measure very low K_a values.

Comparison with other results

The objective of this section is to ask whether the *relative* results of this paper (i.e., data of Fig. 3) are moderately general for all cases where the amino side chains interact with the S_1 pockets of the six enzymes we study. Even from the study of various avian ovomucoid third domains we already know that the changes at P_1 position often are not strictly additive with changes at some other position, e.g., P'_{14}. (Note: In all the comparisons involved in this paper P'_{14} is Gly and these particular nonadditivities were avoided.) In view of the nonadditivities within the ovomucoid third domain set, it is quite gratifying and surprising that there is a good deal of agreement with *relative* K_a changes in other inhibitors when the P_1 residue is varied, and that the *relative* k_{cat}/K_m values of substrates are often quite similar to the results given here. At first it may appear that the literature is replete with needed data. There are, however, serious difficulties in making the comparisons. First, the same enzymes had to be employed. Second, the amino acid side-chain sets must significantly overlap to make comparisons worthwhile and must contain Ala to make the comparison easy. Because our data show a very large range, it would be convenient if the other sets did too. This is generally not the case. Many authors determine K_a or k_{cat}/K_m values only over a 10^3 range and point out that other side chains are out of range. There is the usual and important caveat that the conditions of various experiments are not comparable and another that exact determination of K_a values is a difficult undertaking. For that reason, of the many comparisons we could have made, we chose only a few.

The two inhibitor sets we compare are: the X^{73}SSI variants of Kojima et al. (1991) who prepared all 20 variants and measured their K_a values for subtilisin BPN′. We measured our set with SUBT whose K_a values with ovomucoid third domains are generally higher than of subtilisin BPN′. However, the relative values for Carlsberg and BPN′ appear to be the same for many natural avian ovomucoid third domains.

The second set we employ are the X^{15}BPTI variants of Tschesche and coworkers (Tschesche et al., 1987; Beckmann et al., 1988, 1989). This is a semisynthetic set containing several P_1 residues that overlap our set. However, for only two of our six enzymes do they report enough data to be useful. We exclude their scattered data for chymotrypsin, but we point out that they report that Leu^{15} BPTI inhibits chymotrypsin less well than the wild-type Lys^{15}BPTI. In contrast we find that the wild-type Leu^{18} OMTKY3 is a 1.3×10^3 times better inhibitor of chymotrypsin than Lys^{18}OMTKY3. This seems to be a major discrepancy.

A slogan has emerged from the work on protein inhibitors of serine proteinases—"inhibitors are substrates." This suggests that *relative* K_a values should be compared to *relative* values characterizing the enzyme-substrate interaction. We are now faced with a choice. Should we choose to compare K_a with at least formally analogous $1/K_m$ or with k_{cat}/K_m, which generally shows a greater range and is a measure of specificity (Fersht, 1977, 1985)? In this paper we choose k_{cat}/K_m. It is a successful choice especially for the Dorovska et al. (1972) data on chymotrypsin, but a relatively poor choice for subtilisin.

Table 4 shows the rather good agreement between our data and those of Kojima et al. (1991). Both groups agree on the particularly low range of K_a values for the subtilisins. This agreement is not shared by the k_{cat}/K_m data of Estell (pers. comm.) where the range is much greater than in K_a. The $1/K_m$ comparison works out rather better in this case (Wynn, 1990).

Table 5 shows that the behavior for PPE of X^{18} OMTKY3, of X^{15}BPTI, and of the synthetic substrate is qualitatively the same. β-Branching—the Ape/Val ratio (see Table 3)—is very large for all three cases; Ahx/Met is larger than one for all three cases, but only a factor of two for the substrate. The general shape of the binding

Table 4. *Comparison of relative values for subtilisins*

	Carlsberg[a] K_a X^{18}OMTKY3	BPN′[b] K_a X^{73}SSI	Carlsberg[c] k_{cat}/K_m SucAAPX′pNA
Ala	1.0	1.0	1.0
Ape	1.0×10^{1}	ND[d]	1.3×10^{1}
Leu	3.1	1.0	2.6×10^{1}
Met	5.8	8.0×10^{-1}	2.3×10^{1}
Val	9.3×10^{-2}	5.8×10^{-2}	2.1×10^{-3}
Phe	7.1×10^{-1}	1.0	3.9×10^{1}
Ser	1.3×10^{-1}	2.9×10^{-1}	7.7×10^{-1}
Lys	2.0×10^{-2}	6.9×10^{-2}	8.3
Glu	6.7×10^{-2}	1.3×10^{-2}	3.7×10^{-1}
Gln	8.6×10^{-1}	3.3×10^{-1}	2.2

[a] Calculated from Table 1.
[b] Kojima et al. (1991).
[c] D. Estell (pers. comm.).
[d] ND, not determined.

Table 5. *Comparison of relative values for PPE*[a]

	K_a[b] X^{18}OMTKY3	K_a[c] X^{15}BPTI	k_{cat}/K_m[d] BocAAX^{1}SBzl
Ala	1.0	1.0	1.0
Abu	9.4	2.50×10^{1}	ND
Ape	6.9	6.4×10^{1}	9.2
Ahx	5.4	1.1×10^{1}	4.1
Ahp	6.9×10^{-3}	ND	ND
Leu	1.2	1.5	5.9
Met	2.9×10^{-1}	3.6×10^{-1}	2.3
Val	8.9×10^{-2}	5.0×10^{-1}	8.9×10^{-1}
Phe	8.3×10^{-7}	ND	NR
Ser	3.7×10^{-2}	ND	5.8×10^{-1}
Glu	2.6×10^{-6}	ND	NR

[a] ND, not determined; NR, no reaction was observed.
[b] Calculated from Table 1.
[c] Tschesche et al. (1987); Beckmann et al. (1988, 1989).
[d] Harper et al. (1984).

Table 7. *Comparison of relative values for chymotrypsin*[a]

	K_a[b] X^{18}OMTKY3	k_{cat}/K_m[c] AcX^{1}OMe	k_{cat}/K_m[d] BocAAX^{1}SBzl
Gly	ND	7.1×10^{-2}	ND
Ala	1.0	1.0	1.0
Abu	2.4×10^{1}	1.2×10^{1}	ND
Ape	4.8×10^{2}	2.0×10^{2}	4.6×10^{2}
Ahx	1.7×10^{3}	1.7×10^{3}	1.6×10^{3}
Leu	3.9×10^{3}	ND	1.4×10^{3}
Met	2.4×10^{3}	ND	1.1×10^{3}
Val	3.9	1.1	NR
Phe	4.6×10^{4}	5.8×10^{4}	4.3×10^{3}
Ser	7.6×10^{-1}	ND	1.7×10^{1}
Glu	3.9×10^{-2}	ND	1.2

[a] ND, not determined; NR, no reaction was observed.
[b] Calculated from Table 1.
[c] Dorovska et al. (1972).
[d] Harper et al. (1984).

vs. size curve is similar with Abu, Ape, and Ahx binding almost equally well. Unfortunately, there are no data by others to show the low deformability of the PPE pocket.

Table 6 is where the HLE values are tabulated. The Ala → Ahx series is in striking numerical agreement with X^{15}BPTI. So are the P_1 Met and Lys values. On the other hand the relative Leu and Val values differ strongly. The synthetic substrate is intermediate. In X^{15}BPTI Val ≫ Leu, in X^{18}OMTKY3 Leu > Val. (This was already explicitly discussed in Laskowski et al. [1989].) This may well be another reflection of the deformability of the HLE pocket.

Table 7 is the comparison for chymotrypsin involving only k_{cat}/K_m. However, the comparison to the data of Dorovska et al. (1972) is astonishingly good. The quantitative comparison is excellent. Qualitatively the Dorovska et al. (1972) data are linear with size, whereas the ovomucoid K_a data show a pronounced curvature. The Dorovska Gly point is included. It would probably be quite valid to extrapolate our chymotrypsin data to Gly by the use of this point. Comparison with the very rich Harper et al. (1984) set is also rather good, although the discord on the Phe, Ser, and Glu values seems larger than we would have expected.

In global conclusion we are somewhat surprised that our results agree well with other results and therefore the conclusion may have considerable generality. Had the comparison with simple substrates proved perfect throughout we might have regretted not using them. We have no such regrets. The relatively large protein proteinase inhibitors with a highly complementary contact area provide a much better assurance of a common binding mode than simple substrates. It is clear that the introduction of noncoded amino acids and especially of long series of noncoded amino acids into proteins is a useful way to improve our understanding of protein chemistry. Enzymatic semisynthesis seems to be an excellent method of doing that, but considerable success was also achieved by genetic technology (Mendel et al., 1992).

Table 6. *Comparison of relative values for HLE*[a]

	K_a[b] X^{18}OMTKY3	K_a[c] X^{15}BPTI	k_{cat}/K_m[d] BocAAX^{1}SBzl
Ala	1.0	1.0	1.0
Abu	1.2×10^{1}	1.1×10^{1}	ND
Ape	8.3	1.6×10^{1}	1.3×10^{1}
Ahx	8.3×10^{-1}	9.3×10^{-1}	2.5
Leu	5.2	8.5×10^{-1}	3.0
Met	6.4×10^{-1}	8.5×10^{-1}	1.2
Val	1.9	2.3×10^{1}	4.1
Phe	9.2×10^{-4}	ND	NR
Lys	3.7×10^{-4}	7.3×10^{-4}	ND
Glu	3.3×10^{-6}	ND	NR

[a] ND, not determined; NR, no reaction was observed.
[b] Calculated from Table 1.
[c] Tschesche et al. (1987); Beckmann et al. (1988, 1989).
[d] Harper et al. (1984).

Materials and methods

Natural ovomucoid third domains

The experimental K_a values in Table 1 are based on direct measurements for the Leu^{18} variant (OMTKY3) and the Met^{18} variant (OMSVP3). The Ser^{18} and Val^{18} entries are obtained from additivity calculations on OMMNQ3S and OMCNG3, respectively (see Fig. 2, Table 2, and text). In order to carry out these calculations OMMNQ3L, OMWTD3, and OMSWN3 were also needed. All of these

natural variants were prepared as described in Laskowski et al. (1987). Furthermore, large amounts of OMTKY3 of OMSWN3 and of OMSVP3 were needed for semisynthetic work. These were prepared as above with scale-up modifications described by Bigler (1991).

Semisynthetic OMTKY3 variants

Ala^{18}, Abu^{18}, Ape^{18}, Ahx^{18}, Ahp^{18}, Lys^{18}, Glu^{18}, and Gln^{18} were prepared by enzymatic semisynthesis as described by Wieczorek and Laskowski (1983), Wieczorek et al. (1987), and Bigler (1991). The reactive-site peptide bond in OMTKY3 (Fig. 1) was hydrolyzed by SGBP at pH 1.5. The modified inhibitor (with the Leu^{18}-Glu^{19} bond hydrolyzed) was isolated by anion exchange chromatography. Reduction of the disulfide bridges yields two fragments: NH_2-terminal (1–18) and COOH-terminal (19–56). These were separated by size exclusion under N_2 on a Biogel P-10 column. The NH_2-terminal fragment was of no further interest in this study. The COOH-terminal fragment was converted to glutathione mixed disulfide at its four CysH residues (positions 24, 35, 38, 56).

Synthesis of the NH_2-terminal peptide

In order to make the replacement described here, the entire NH_2-terminal peptide is synthesized. However, Wieczorek et al. (1987) proved that the presence of the first five NH_2-terminal residues has no effect on the K_a values and on the synthetic procedure. Therefore, in all of the semisynthetic variants only residues 6–18 were introduced. On the other hand the synthetic peptides were all COOH-terminally extended by $-Gly^{18A}NH_2$. This expedient (for the first production of modified inhibitor $-GlyNH_2$ see Kowalski et al. [1974]) has two advantages. It eliminates the need for resins to which the noncoded amino acids are coupled and it aids in the formation of the enzyme inhibitor complex with proteinase K. The synthesis was carried out in a DuPont Rapid Manual Peptide Synthesizer (RaMPS), following the procedures suggested by the manufacturer (DuPont, 1987). The overall procedure is based on the fluorenylmethyloxycarbonyl (FMOC) approach (Carpino & Han, 1972; Atherton & Sheppard, 1989). The synthesized peptide was converted to glutathione mixed disulfide at its two Cys residues (positions 6, 16).

Synthesis of noncoded FMOC amino acids

Whereas FMOC-L-coded amino acids are readily available from various suppliers, many FMOC-L-noncoded amino acids are not. Therefore, FMOC was added to L-Abu, L-Ape, and L-Ahx (Sigma) by the procedure of Carpino and Han (1972). This procedure was also used to add FMOC to DL-Ahp (Dixon Fine Chemicals) and to DL-Leu (as a control). We could not find a commercial supplier of L-Ahp or of higher homologs. Attempts to prepare FMOC derivatives of DL-higher homologs failed (Bigler, 1991).

Formation of disulfide bridges

The glutathione mixed disulfides of the NH_2-terminal peptide variant X^{18} (6–18) Gly^{18A} (0.5 mM) and the glutathione mixed disulfide of the natural COOH-terminal peptide (19–56) (0.09 mM) were mixed in 0.2 M Tris, 0.2 M KCl, 1 mM EDTA, pH 8.7, with 5 mM cysteine HCl under N_2 and allowed to undergo disulfide interchange overnight at 4 °C. This mixture was desalted and separated on S-Sepharose at pH 3.5. The peak corresponding to modified inhibitor was found as it inhibits subtilisin. This product is by a number of criteria the $Gly^{18A}NH_2$ of (6–56) modified inhibitor variant.

Closure of the reactive-site peptide bond

In all complexes of "standard mechanism" (Laskowski & Kato, 1980) inhibitors with their cognate serine proteinases (Read & James, 1986; Bode & Huber, 1991, 1992), the reactive-site peptide bond is intact. It is the essence of the "standard mechanism" (Finkenstadt & Laskowski, 1967; Laskowski & Kato, 1980) that the complexes made from virgin and from modified inhibitors are the same. Thus forming a complex from modified inhibitor suffices to reform the reactive-site peptide bond in complex. This was accomplished by combining 1.0 μmol of the modified inhibitor variant in 1 mL with 0.9 μmol of proteinase K at pH 6.5 in 0.1 M bis-Tris buffer and 0.02 M KCl. The complex once formed must be dissociated to obtain the desired virgin inhibitor. If the dissociation is slow, the predominant product is likely to be the modified inhibitor (Ardelt & Laskowski, 1983, 1991). However, at low pH the rate of dissociation to virgin inhibitor is much more rapid than to modified (Ardelt & Laskowski, 1985), therefore kinetically controlled dissociation at very low pH yields virgin inhibitor.

Therefore, 3 min after the complex was formed, the pH value was lowered by adding with stirring an amount of 1 M HCl required to lower the pH value to 1.0. The enzyme and inhibitor were separated by size exclusion chromatography (Biogel P-10), and the inhibitor was further purified by both cation (S-Sepharose, pH 3.5) and anion (Q-Sepharose, pH 9.0) exchange chromatography.

Stereomeric resolution

For Ahp (and all higher homologs) we could purchase only a racemic mixture (Dixon Fine Chemicals). We suspected that our enzymatic synthesis suffices to resolve this mixture, and therefore we have prepared $L\text{-}Ahp^{18}$ homolog from the NH_2-terminal peptide $DL\text{-}Ahp^{18}$ (6–18) $Gly^{18A}NH_2$. To test this, we prepared the Leu^{18} variant from $DL\text{-}Leu^{18}$ (see above). We have then measured the

K_a values for this variant and found them indistinguishable from those of natural OMTKY3 or from semisynthetic L-Leu^{18} OMTKY3 (Wieczorek et al., 1987). Furthermore, the inhibitor made from DL-Leu^{18} was ~100% active (not 50% active). Thus, it appears that at least at the P_1 position our procedure yields resolved L-X^{18} variants from peptides that are DL at X^{18}.

Phe^{18}OMSWN3

Because we were concerned that Phe^{18}OMTKY3 may be too strong an inhibitor of chymotrypsin and too weak an inhibitor of PPE (see text), we chose to synthesize Phe^{18} OMSWN3. The procedures employed were essentially the same as above except that OMSWN3 COOH-terminal peptide (19–56) was used. This differs from OMTKY3 by R21M, E43D. The NH_2-terminal peptide was also altered E10D, L18F.

Hse^{18}OMTKY3

Silver pheasant ovomucoid third domain (OMSVP3) differs from OMTKY3 only by L18M replacement. Note that OMSVP3 contains only one Met residue. OMSVP3 was subjected to CNBr cleavage, and the product with the Met^{18}-Glu^{19} bond cleaved and Hse^{18} was isolated. The modified inhibitor with Hse^{18} was then used as a starting material for closure of the reactive-site peptide bond. It is worth noting that this procedure is likely to be general for all standard mechanism inhibitors that have a single Met at the P_1 position.

Characterization of variants

The NH_2-terminal (synthetic) and the COOH-terminal (natural) peptides and their glutathione mixed disulfides were extensively tested by amino acid analysis. The completed variants (virgin form) were subjected to hydrolysis and amino acid analysis in a Beckman 7300 Analyzer. The virgin inhibitors were also subjected to NH_2-terminal sequencing in a Porton 2020 Sequencer. This was carried at least to Glu^{19} (the P_1' residue). The sequencing showed not only that the synthetic peptide that was inserted had the anticipated sequence, but also that all of the inhibitor had the reactive-site peptide bond closed and that the Gly$^{18A}NH_2$ was quantitatively eliminated in the enzymatic bond closure step.

Enzymes

The enzymes used in this work for determination of equilibrium constants were as follows: CHYM was purchased from Worthington Biochemical Co.; PPE, prepared to eliminate contamination by other pancreatic proteinases according to Satoh et al. (1979), was a gift from the late Prof. M. Laskowski, Sr.; SUBT (from *Bacillus subtilis*) was purchased from Sigma. It was compared to a sample donated by D. Estell of Genencor and found indistinguishable; SGPA and SGPB were initially gifts from Prof. L. Smillie of the University of Alberta and later were isolated in our laboratory from pronase. HLE was originally a gift from Prof. J. Travis, University of Georgia. It was later purchased from Elastin Products. In all cases the K_a values for the enzyme that was used as a replacement for a well-characterized gift with OMTKY3 were indistinguishable (within ± 20% experimental error) from the original values.

Determinations of equilibrium constants

Determinations of equilibrium constants were carried out by an extensively modified procedure of Green and Work (1953) as described by Empie and Laskowski (1982) and for later modifications by Park (1985) and Wynn (1990). The detailed procedure will be a subject of a separate publication (Tashiro et al., in prep.). The procedure allows for the determination of K_a values over nine orders of magnitude (K_a from about 10^3 M^{-1} to 10^{12} M^{-1}) for some enzymes. The accuracy is 10–20%. The data were obtained at 21 ± 2 °C at pH 8.30 in 0.1 M Tris-HCl containing 0.005% Triton X-100 and 0.02 M $CaCl_2$. The free enzyme concentrations in the incubation mixture were measured by addition of an appropriate tetrapeptide *p*-nitroanilide substrate (selected from a large library of such substrates). The rate of change of absorbance was measured with a Diode Array HP8450A Spectrophotometer by displaying the difference between the average absorbance between 380 nm and 410 nm and subtracting the difference in the average absorbance from 650 nm to 700 nm. The subtraction of the longer wavelengths helps to correct for spurious effects of dust and of bubbles and allows for accurate reading of smaller rates.

Acknowledgments

We thank I. Apostol for the sequencing and amino acid analyses of the semisynthetic variants; D. Estell, the late M. Laskowski, Sr., L. Smillie, and J. Travis for gifts of enzymes; and N. Warne for introducing RaMPS synthesis in our laboratory.

References

Anfinsen, C.B. & Corley, L.G. (1969). An active variant of staphylococcal nuclease containing norleucine in place of methionine. *J. Biol. Chem. 224*, 5149–5152.

Ardelt, W. & Laskowski, M., Jr. (1983). Thermodynamics and kinetics of the hydrolysis and resynthesis of the reactive site peptide bond in turkey ovomucoid third domain by aspergillopeptidase B. *Acta Biochem. Polon. 30*, 115–126.

Ardelt, W. & Laskowski, M., Jr. (1985). Turkey ovomucoid third domain inhibits eight different serine proteinases of varied specificity on the same ... Leu^{18}-Glu^{19} ... reactive site. *Biochemistry 24*, 5313–5320.

Ardelt, W. & Laskowski, M., Jr. (1991). Effect of single amino acid replacements on the thermodynamics of the reactive site peptide bond hydrolysis in ovomucoid third domain. *J. Mol. Biol. 220*, 1041–1053.

Atherton, E. & Sheppard, R.C. (1989). *Solid Phase Peptide Synthesis.* IRL Press, Oxford, England.

Beckmann, J., Mehlich, A., Schröder, W., Wenzel, H.R., & Tschesche, H. (1988). Preparation of chemically 'mutated' aprotin homologues by semisynthesis. *Eur. J. Biochem. 176,* 675–682.

Beckmann, J., Mehlich, A., Schröder, W., Wenzel, H.R., & Tschesche, H. (1989). Semisynthesis of Arg^{15}, Glu^{15}, Met^{15}, and Nle^{15}-aprotinin involving enzymatic peptide bond resynthesis. *J. Protein Chem. 8,* 101–113.

Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R.N., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M., Reichlin, M., Sercarz, E.E., Smith-Gill, S.J., Todd, P.E., & Wilson, A.C. (1984). The antigenic structure of proteins: A reappraisal. *Annu. Rev. Immunol. 2,* 67–101.

Bigler, T.L. (1991). Semisynthesis of ovomucoid third domain variants: Structure–function studies. Ph.D. Thesis, Purdue University, West Lafayette, Indiana.

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. (1991). Ligand binding: Proteinase–protein inhibitor interactions. *Curr. Opin. Struct. Biol. 1,* 45–52.

Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem. 204,* 433–451.

Bode, W., Meyer, E., & Powers, J.C. (1989). Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry 28,* 1951–1963.

Bode, W., Wei, A.-Z., 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.

Carpino, L.A. & Han, G.Y. (1972). The 9-fluorenylmethoxycarbonyl amino-protecting group. *J. Organ. Chem. 37,* 3404–3409.

Creighton, T.E. & Darby, N.J. (1989). Functional evolutionary divergence of proteolytic enzymes and their inhibitors. *Trends Biochem. Sci. 14,* 319–324.

Cunningham, B.C. & Wells, J.A. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science 244,* 1081–1085.

Davies, D.R., Sheriff, S., & Padlan, E.A. (1988). Antibody–antigen complexes. *J. Biol. Chem. 263,* 10541–10544.

Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry 29,* 7133–7155.

Doolittle, R.F. (1979). Protein evolution. In *The Proteins,* Vol. IV (Neurath, H. & Hill, R.L., Eds.), pp. 2–110. Academic Press, New York.

Dorovska, V.N., Varfolomeyev, S.D., Kazanskaya, N.F., Klyosov, A.A., & Martinek, K. (1972). The influence of the geometric properties of the active centre on the specificity of α-chymotrypsin catalysis. *FEBS Lett. 23,* 122–124.

Empie, M.W. & Laskowski, M., Jr. (1982). Thermodynamics and kinetics of single residue replacements in avian ovomucoid third domains: Effect on inhibitor interactions with serine proteinases. *Biochemistry 21,* 2274–2284.

Eriksson, A.E., Baase, W.A., Zhang, X.-J., Heinz, D.W., Blaber, M., Baldwin, E.P., & Matthews, B.W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science 255,* 178–183.

Estell, D.A., Graycar, T.P., Muller, J.V., Powers, D.B., Burnier, J.P., Ng, P.G., & Wells, J.A. (1986). Probing steric and hydrophobic effects on enzyme–substrate interactions by protein engineering. *Science 233,* 659–663.

Fersht, A. (1977). *Enzyme Structure and Mechanism.* W.H. Freeman, Reading and San Francisco.

Fersht, A. (1985). *Enzyme Structure and Mechanism,* 2nd Ed. W.H. Freeman, New York.

Finkenstadt, W.R. & Laskowski, M., Jr. (1967). Resynthesis by trypsin of the cleaved peptide bond in modified soybean trypsin inhibitor. *J. Biol. Chem. 212,* 771–773.

Fujinaga, M., Read, R.J., Sielecki, A., Ardelt, W., Laskowski, M., Jr., & James, M.N.G. (1982). Refined crystal structure of the molecular complex of *Streptomyces griseus* protease B, a serine protease, with the third domain of the ovomucoid inhibitor from turkey. *Proc. Natl. Acad. Sci. USA 79,* 4868–4772.

Fujinaga, M., Sielecki, A.R., Read, R.J., Ardelt, W., Laskowski, M., Jr., & James, M.N.G. (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.

Green, N.M. & Work, E. (1953). Pancreatic trypsin inhibitor. *Biochem. J. 54,* 347–352.

Harper, J.W., Cook, R.R., Roberts, C.J., McLaughlin, B.J., & Powers, J.C. (1984). Active site mapping of the serine proteases human leukocyte elastase, cathepsin G, porcine pancreatic elastase, rat mast cell proteases I and II, bovine chymotrypsin A_α, and *Staphylococcus aureus* protease V-8 using tripeptidethiobenzyl ester substrates. *Biochemistry 23,* 2995–3002.

Holtzer, A. (1992). The use of Flory–Huggins theory in interpreting partitioning of solutes between organic liquids and water. *Biopolymers 32,* 711–715.

Horovitz, A., Serrano, L., & Fersht, A.R. (1991). COSMIC analysis of the major α-helix of barnase during folding. *J. Mol. Biol. 219,* 5–9.

Kojima, S., Nishiyama, Y., Kumagai, I., & Miura, K. (1991). Inhibition of subtilisin BPN′ by reaction site P1 mutants of *Streptomyces* subtilisin inhibitor. *J. Biochem. 109,* 377–382.

Komiyama, T., Bigler, T.L., Yoshida, N., Noda, K., & Laskowski, M., Jr. (1991). Replacement of P_1 Leu^{18} by Glu^{18} in the reactive site of turkey ovomucoid third domain converts it into a strong inhibitor of Glu-specific *Streptomyces griseus* proteinase (GluSGP). *J. Biol. Chem. 266,* 10727–10730.

Kowalski, D., Leary, T.R., McKee, R.F., Sealock, R.W., Wang, D., & Laskowski, M., Jr. (1974). Replacements, insertions, and modifications of amino acid residues in the reactive site of soybean trypsin inhibitor (Kunitz). In *Bayer-Symposium V "Proteinase Inhibitors" 324,* pp. 311–323. Springer-Verlag, Berlin.

Laskowski, M., Jr. (1980). An algorithmic approach to sequence → reactivity of proteins. *Biochem. Pharmacol. 29,* 2089–2094.

Laskowski, M., Jr. (1986). Protein inhibitors of serine proteinases—Mechanism and classification. In *Nutritional and Toxicological Significance of Enzyme Inhibitors in Food* (Friedman, M., Ed.), pp. 1–17. Plenum, New York.

Laskowski, M., Jr., Apostol, I., Ardelt, W., Cook, J., Giletto, A., Kelly, C.A., Lu, W., Park, S.J., Qasim, M.A., Whatley, H.E., Wieczorek, A., & Wynn, R. (1990). Amino acid sequences of ovomucoid third domain from 25 additional species of birds. *J. Protein Chem. 9,* 715–725.

Laskowski, M., Jr. & Kato, I. (1980). Protein inhibitors of proteinases. *Biochemistry 49,* 593–626.

Laskowski, M., Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M.W., Kohr, W.J., Park, S.J., Parks, K., Schatzley, B.L., Schoenberger, O.L., Tashiro, M., Vichot, G., Whatley, H.E., Wieczorek, A., & Wieczorek, M. (1987). Ovomucoid third domains from 100 avian species: Isolation, sequences, and hypervariability of enzyme–inhibitor contact residues. *Biochemistry 26,* 202–221.

Laskowski, M., Jr., Kato, I., Kohr, W.J., Park, S.J., Tashiro, M., & Whatley, H.E. (1988). Positive Darwinian selection in evolution of protein inhibitors of serine proteinases. *Cold Spring Harbor Symp. Quant. Biol. 52,* 545–553.

Laskowski, M., Jr., Park, S.J., Tashiro, M., & Wynn, R. (1989). Design of highly specific inhibitors of serine proteinases. In *Protein Recognition of Immobilized Ligands, UCLA Symposia on Molecular and Cellular Biology, New Series, 80* (Hutchens, T.W., Ed.), pp. 149–168. Alan R. Liss, Inc., New York.

Laskowski, M., Jr., Tashiro, M., Empie, M.W., Park, S.J., Kato, I., Ardelt, W., & Wieczorek, M. (1983). Relationship between the amino acid sequence and inhibitory activity of protein inhibitors of proteinases. In *Proteinase Inhibitors: Medical and Biological Aspects* (Katunama, N., et al., Eds.), pp. 55–68. Japan Scientific Society Press, Tokyo/Springer-Verlag, Berlin.

Mendel, D., Ellman, J.A., Chang, Z., Veenstra, D.L., Kollman, P.A., & Schultz, P.G. (1992). Probing protein stability with unnatural amino acids. *Science 256,* 1798–1802.

Nozaki, Y. & Tanford, C. (1971). The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem. 246,* 2211–2217.

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, Indiana.

Read, R.J., Fujinaga, M., Sielecki, A.R., & James, M.N.G. (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, R.J. & James, M.N.G. (1986). Introduction to the protein inhibitors: X-ray crystallography. In *Proteinase Inhibitors* (Barrett, A.J. & Salvesen, G., Eds.), pp. 301–336. Elsevier, Amsterdam.

Satoh, S., Kurecki, T., Kress, L.F., & Laskowski, M., Sr. (1979). The dual nature of the reaction between porcine elastase and human plasma α_1 proteinase inhibitor. *Biochem. Biophys. Res. Commun. 86*, 130–137.

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

Sealock, R.W. & Laskowski, M., Jr. (1969). Enzymatic replacement of the arginyl by a lysyl residue in the reactive site of soybean trypsin inhibitor. *Biochemistry 8*, 3703–3710.

Sharp, K.A., 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.

Tschesche, H., Beckmann, J., Mehlich, A., Schnabel, E., Truscheit, E., & Wenzel, H.R. (1987). Semisynthetic engineering of proteinase inhibitor homologues. *Biochim. Biophys. Acta 913*, 97–101.

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.

Wells, J.A. (1990). Additivity of mutational effects in proteins. *Biochemistry 29*, 8509–8517.

Wieczorek, M. & Laskowski, M., Jr. (1983). Formation of covalent hybrids from amino-terminal and carboxy-terminal fragments of two ovomucoid third domains. *Biochemistry 22*, 2630–2636.

Wieczorek, M., Park, S.J., & 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.

Wynn, R. (1990). Design of a specific human leukocyte elastase inhibitor based on ovomucoid third domains. Ph.D. Thesis, Purdue University, West Lafayette, Indiana.