J. Mol. Biol. (1997) **266**, 441–461

JMB

Binding of Amino Acid Side-chains to S_1 Cavities of Serine Proteinases

Wuyuan Lu[1], Izydor Apostol[1], M.A. Qasim[1], Nicholas Warne[1] Richard Wynn[1], Wen Lei Zhang[1], Stephen Anderson[2], Yi Wen Chiang[2] Eleanor Ogin[2], Irvin Rothberg[2], Kevin Ryan[2] and Michael Laskowski Jr[1]*

[1]*Department of Chemistry Purdue University, 1393 Brown, West Lafayette IN 47907, USA*

[2]*Center for Advanced Biotechnology and Medicine Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854 USA*

**Corresponding author*

The P_1 or primary specificity residue of standard mechanism canonical protein inhibitors of serine proteinases, inserts into the S_1 primary specificity cavity of the cognate enzyme upon enzyme-inhibitor complex formation. Both natural evolution and protein engineering often change the P_1 residue to greatly alter the specificity and the binding strength. To systematize such results we have obtained all 20 coded P_1 variants of one such inhibitor, turkey ovomucoid third domain, by recombinant DNA technology. The variants were extensively characterized. The association equilibrium constants were measured at pH 8.30, 21 (±2)°C, for interaction of these variants with six well characterized serine proteinases with hydrophobic S_1 cavities. The enzyme names are followed by the best, worst and most specific coded residue for each. Bovine chymotrypsin Aα (Tyr, Pro, Trp), porcine pancreatic elastase (Leu/Ala, Arg, Ala), subtilisin Carlsberg (Cys, Pro, Glu), *Streptomyces griseus* proteinase A (Cys, Pro, Leu) and B (Cys, Pro, Lys) and human leukocyte elastase (Ile, Asp, Ile). The data set was merged with K_a values for five non-coded variants at P_1 of turkey ovomucoid third domain obtained in our laboratory by enzymatic semisynthesis. The ratios of the highest to the lowest K_a for each of the six enzymes range from 10^6 to 10^8. The dominant force for binding to these pockets is the hydrophobic interaction. Excess steric bulk (too large for the pocket), awkward shape (Pro, Val and Ile), polarity (Ser) oppose interaction. Ionic charges, especially negative charges on Glu^- and Asp^- are strongly unfavorable. The Pearson pro-

Present addresses: W. Lu, Scripps Research Institute, 10666 N. Torrey Pines Road, La Jolla, CA 92037, USA; I. Apostol, Somatogen, Inc., 5797 Central Avenue, Boulder, CO 80301, USA; N. Warne, Genetics Institute, Drug Product Development, 1 Burtt Road, Andover, MA 01810, USA; R. Wynn, The DuPont Merck Pharmaceutical Company, Experimental Station, Rte. 141, Wilmington, DE 19880, USA; S. Anderson, Dept of Molecular Biology and Biochemistry, Rutgers University, New Brunswick, NJ 08903, USA; E. Ogin, Interferon Sciences, Inc., 783 Jersey Ave New Brunswick, NJ 08901, USA; I. Rothberg, Dept of Chemistry, Olson Laboratories, Rutgers University-Newark, 73 Warren St., Newark, NJ 07102, USA.

Abbreviations used: CHYM, bovine chymotrypsin Aα; PPE, porcine pancreatic elastase; CARL and BPN′, subtilisins Carlsberg and BPN′, respectively; SGPA and SGPB, *Streptomyces griseus* proteinases A and B, respectively; HLE, human leukocyte elastase; amino acids (L forms at all times), standard three-letter and one-letter abbreviations for the coded set; ''Cys''-SH, cysteine (as opposed to cystine); Abu, αaminobutyric acid; Ape, αaminopentanoic acid (norvaline); Ahx, αaminohexanoic acid (norleucine); Ahp, αaminoheptanoic acid; Hse, αaminoγhydroxybutyric acid (homoserine); Y, nitrotyrosine; Ac, acetyl; Boc-, butyloxycarbonyl; Abz-, *o*-aminobenzoyl; Suc-, succinyl; -OCH_3, methyl; -SBzl, thiobenzyl; pNA, *p*-nitroanilide; OM, ovomucoid; OMTKY3, turkey ovomucoid third domain; X^{18}OMTKY3, turkey ovomucoid third domain whose residue 18 (P_1 residue) was altered to X; OMSVP3, silver pheasant ovomucoid third domain; OMWTD3, West Indian tree duck ovomucoid third domain; OMSWN3, black-necked swan ovomucoid third domain; OMGOO3, greylag goose ovomucoid third domain; OMEMG3, emperor goose ovomucoid third domain; OMSWNGOO3, a chimera of OMSWN3 and OMGOO3; OMGUI3, Guinea fowl ovomucoid third domain; OMMNQ3, Montezuma quail ovomucoid third domain; OMMNQ3L = Leu^{18}OMMNQ3; OMMNQ3S = Ser^{18}OMMNQ3; OMKAP3, Kalij pheasant ovomucoid third domain; BPTI, bovine pancreatic trypsin inhibitor (Kunitz); SLPI, secretory leukocyte proteinase inhibitor; SSI, *Streptomyces* subtilisin inhibitor; amu, atomic mass unit; t_m, melting temperature; TFA, trifluoroacetic acid; IgG, immunoglobulin G; DTT, dithiothreitol; SPA, staphylococcal protein A.

0022–2836/97/070441–21 $25.00/0/mb960781

duct moment correlations for all the 15 enzyme pairs were calculated. We suggest that these may serve as a quantitative description of the specificity of the enzymes at P_1. The sets of *Streptomyces griseus* proteinases A and B and of the two elastases are strongly positively correlated. Strikingly, chymotrypsin and pancreatic elastase are negatively correlated (−0.10). Such correlations can be usefully extended to many other enzymes and to many other binding pockets to provide a general measure of pocket binding specificity.

Keywords: serine proteinases; proteinase inhibitors; protein–protein interaction; amino acid side-chains, binding of; equilibrium constants for association

Introduction

Substrate and inhibitor recognition by most serine proteinases involves *inter alia* the binding of the P_1 residue (see Figure 1 for explanation of Schechter & Berger, 1967 numbering) to the S_1 cavity of the enzyme. This is probably the most famous of protein-protein recognition motifs. Millions of students learn that trypsin is specific for Arg and Lys P_1 side-chains because of the presence of Asp^{189} (chymotrypsinogen numbering) at the bottom of its S_1 pocket. Similarly, chymotrypsin prefers Tyr, Trp, Phe, Leu and Met P_1 residues because of its commodious S_1 pocket with neutral Ser^{189} at the bottom of it.

Ovomucoid is a major glycoprotein in avian eggwhites whose single polypeptide chain consists of three tandem Kazal family† domains, each with a single, actual or putative, reactive site for inhibition of serine proteinases (Kato *et al.*, 1976, 1978, 1987). Rhodes *et al.* (1960) showed that ovomucoids from various, closely related species of birds exhibit striking variation in inhibitory activity and specificity. We have found that such variation persists when third domains (Figure 1) are examined alone (Empie & Laskowski, 1982; S. J. Park *et al.*, unpublished results) and that it is due to much more rapid fixation of mutations in the enzyme-inhibitor contact region compared to the rest of the inhibitor molecule (Laskowski *et al.*, 1987; 1990; Apostol *et al.*, 1993). We have attributed this hypervariability to positive Darwinian selection (Laskowski *et al.*, 1988).

The reasons for the choice of turkey ovomucoid third domain as the wild type are the approximately equal inhibition of all six enzymes we study and the ready availability. Ovomucoid third domains are among the best studied of all small proteins and of protein inhibitors of serine proteinases. The three-dimensional structures of two free ovomucoid third domains, Japanese quail (Weber *et al.*, 1981; Papamokos *et al.*, 1982) and silver pheasant (Bode *et al.*, 1985) were determined both in virgin (reactive site intact) and in modified (reactive site hydrolyzed) forms (Musil *et al.*, 1991). The three-dimensional structures of complexes of turkey ovomucoid third domain with *Streptomyces griseus* proteinase B (Fujinaga *et al.*, 1982; Read *et al.*, 1983), human leukocyte elastase (Bode *et al.*, 1986) and bovine αchymotrypsin (Fujinaga *et al.*, 1987) were also determined. There are several

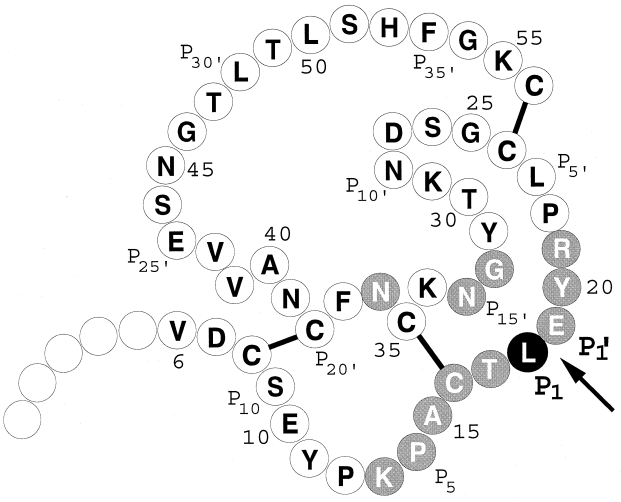

Figure 1. Covalent structure of turkey ovomucoid third domain (Laskowski *et al.*, 1987). The - indicate disulfide bridges. The numbering system, now widely employed for all avian ovomucoid third domains, starts from the hydrolysis site by *Staphyloccocus aureus* proteinase V8 (Weber *et al.*, 1981). The arrow indicates the reactive site peptide bond, which serves as the origin of Schecter & Berger (1967) P_n, P'_n notation. The consensus set of residues that contact the enzyme in various complexes (Apostol *et al.*, 1993) is darkened. The P_1 residue, Leu^{18} is in black. All 24 substitutions discussed in this paper are at that position. Most variants studied were third domains proper or the ''short forms'' 6-56. The natural ''long form'' has Leu^1 Ala Ala Val Ser^5 in the open cirlces; the recombinant long form has Ala^1 His Phe Ala Met^5 in these positions. Wieczorek *et al.* (1987) and this paper show that the presence or absence of either 1-5 peptide is without effect on K_a.

† For definitions of protein inhibitor families, see Laskowski & Kato (1980) and Laskowski (1986)

more structures that are either unpublished or now in progress, both in Martinsried and in Edmonton. The most relevant to this paper is that Huang, Bateman and James intend to obtain the structures of all 20 coded X^{18} OMTKY3 in complex with SGPB. Some of these structures e.g. Gly^{18}, Ala^{18} and Leu^{18} (short form) are already completed (Huang *et al.*, 1995), many others await publication (Huang, 1995; Bateman *et al.*, 1996), others are in refinement. These studies as well as the interpretation of the huge volume of thermodynamic data available at Purdue lead to the following conclusions (1) residue 18 always serves as the P_1 residue even when its binding to the S_1 cavity is locally deleterious, (2) the changes in the inhibitor and enzyme conformation on complex formation are very small, nearly lock and key, (3) with the exception of Pro^{18}, the nature of X^{18} has little effect on the interaction between the enzyme and the remaining residues of the inhibitor. Thus, for our purposes, the inhibitor-enzyme system serves as a molecular vise. It positions X^{18} into the S_1 cavity and by subtraction of Gly^{18} data (see below) it allows us to measure just the $X^{18} \ldots S_1$ interaction.

We have already reported K_a values for 14 variants of ovomucoid third domain at X^{18} (P_1 residue) interacting with six serine proteinases (Bigler *et al.*, 1993). Of these 14 variants, nine involved coded residues, while five were non-coded. Of the nine coded variants, six were obtained directly, while three were calculated from other data by making (reasonable) additivity assumptions. Most of the variants employed were generated by semisynthesis (Wieczorek & Laskowski, 1983).

Here, we describe a set of techniques announced by us earlier (Lu *et al.*, 1992) for the production of coded ovomucoid third domain variants. We apply these techniques to the production of all 20 coded

Table 1. Equilibrium constants for 25 X^{18}OMTKY3 P_1 variants; pH 8.30, $t = 21(\pm 2)^{\circ}C$, $T/2 = 0.10$ M (M^{-1})

Enzymes:	CHYM	PPE	CARL	SGPA	SGPB	HLE
P_1:						
0 Gly	6.6×10^6	9.0×10^8	6.4×10^8	4.5×10^7	1.2×10^7	2.2×10^7
1 Ala	6.1×10^7	4.2×10^{10}	2.0×10^{10}	1.9×10^9	3.6×10^8	1.0×10^9
2 Abu	1.1×10^9	3.3×10^{11}	1.0×10^{11}	1.0×10^{10}	2.3×10^9	1.4×10^{10}
Ser	4.2×10^7	8.6×10^8	2.4×10^9	1.6×10^8	5.0×10^7	3.3×10^7
Cys	2.2×10^9	2.5×10^{10}	3.3×10^{11}	3.3×10^{11}	5.7×10^{10}	7.1×10^9
3 Ape	2.2×10^{10}	2.4×10^{11}	1.4×10^{11}	8.3×10^{10}	1.2×10^{10}	1.0×10^{10}
Val	1.4×10^8	8.8×10^9	8.9×10^8	2.1×10^9	3.3×10^8	1.4×10^{10}
Hse	2.5×10^9	7.5×10^9	5.6×10^{10}	6.5×10^9	1.7×10^9	2.6×10^8
Thr	9.8×10^7	2.9×10^{10}	2.6×10^{10}	2.0×10^9	2.5×10^8	1.3×10^9
Pro	6.7×10^4	5.9×10^5	7.9×10^4	4.6×10^4	3.6×10^4	1.9×10^5
4 Ahx	8.0×10^{10}	1.9×10^{11}	6.7×10^{10}	2.3×10^{11}	2.1×10^{10}	9.9×10^8
Leu	1.9×10^{11}	4.2×10^{10}	3.4×10^{10}	3.0×10^{11}	5.6×10^{10}	6.1×10^9
Ile	9.9×10^7	5.9×10^9	1.4×10^8	1.5×10^8	2.9×10^7	2.1×10^{10}
Met	1.0×10^{11}	1.3×10^{10}	8.1×10^{10}	2.6×10^{11}	2.7×10^{10}	7.3×10^8
Asn	6.6×10^8	6.9×10^7	2.7×10^9	2.6×10^8	1.8×10^8	9.0×10^5
Asp	1.0×10^6	7.3×10^4	1.6×10^7	5.1×10^6	3.9×10^6	1.6×10^4
5 Ahp	3.3×10^{11}	2.4×10^8	8.9×10^{10}	4.7×10^{11}	2.8×10^{10}	1.3×10^8
Lys	1.0×10^8	4.6×10^4	1.2×10^8	5.6×10^7	2.6×10^8	3.8×10^5
Gln	1.3×10^9	4.3×10^7	1.3×10^{10}	1.2×10^9	6.9×10^8	2.3×10^7
Glu	2.5×10^6	8.4×10	8.6×10^8	7.9×10^6	2.3×10^6	4.1×10^4
6 His	1.2×10^9	3.7×10^4	1.3×10^{10}	2.6×10^9	3.1×10^9	1.1×10^5
7 Phe	2.4×10^{12}	4.9×10^4	1.2×10^{10}	1.8×10^{11}	5.5×10^9	9.9×10^5
Arg	2.2×10^8	4.8×10^3	2.4×10^8	8.2×10^7	1.9×10^8	3.3×10^4
8 Tyr	8.1×10^{12}	7.5×10^3	1.6×10^{10}	1.0×10^{11}	3.3×10^9	8.3×10^4
10 Trp	3.3×10^{12}	2.5×10^4	2.6×10^9	8.9×10^9	2.4×10^9	1.8×10^4

The largest K_a for each enzyme is double underlined, the smallest is single underlined and the most specific inhibitor is boxed. The data for the five non-coded residues Abu, Ape, Ahx, Ahp and Hse are from Bigler *et al.* (1993).

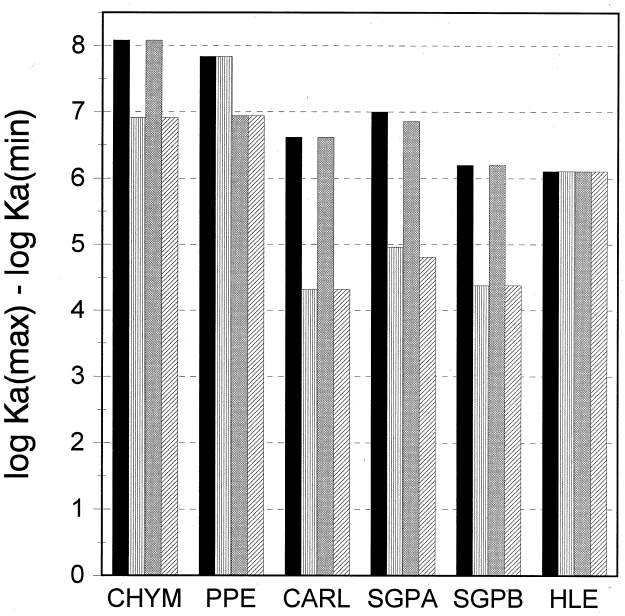

Figure 2. The K_a ranges expressed as the difference in logarithms. (■) All side-chains of Table 1. (▥) All but Pro. (▩) All 20 coded side-chains. (▨) All coded side-chains except Pro.

X^{18} ovomucoid third domain variants. We directly measure the K_a values for all these, interacting with the same six serine proteinases as employed by Bigler *et al*. (1993). We combine our complete set of 20 with the set of five non-coded residues (Abu, Ape, Ahx, Ahp and Hse) provided by Bigler *et al*. (1993). The non-coded set K_a values were also all obtained by direct measurements. The combined set consists of $25 \times 6 = 150$ K_a values. It allows for a great number of comparisons both internally and to literature values. Only some of these are provided below.

Results and Discussion

The K_a values

Table 1 lists the K_a values for all 20 coded variants of X^{18}OMTKY3. We were fortunate that in spite of the huge range of these data (see below) all the values fit into our dynamic range of measurements (roughly 10^3 M^{-1} to 10^{13} M^{-1}, the upper bound is lower for some enzymes e.g. HLE). As was already pointed out by Bigler *et al*. (1993) the set of all coded amino acids allows for only a few structurally simple pairwise comparisons. The homologous series e.g. Gly, Ala; Asp, Glu; Asn, Gln; Val, Ile are only two residues long. Inclusion of only a few well selected non-coded residues allows for many more good pairwise comparisons and for much longer series. The clear drawback is that one always wants to add more non-coded members. The set is never complete. Table 1 is augmented by the values for five non-coded variants obtained by semisynthesis. Their K_a values with the same six enzymes (Bigler *et al*., 1993) were measured by the same technique and under the same conditions as used for the coded set.

The ranges

We call the ratio of the highest to the lowest K_a for each enzyme, the range for that enzyme. The ratios, or more properly, their logarithms are shown in Figure 2. When converted to free energy differences, they range from 8.1 to 10.8 kcal/mol. These values are similar to the total stability of the native forms of many globular proteins.

The ranges given in the first columns of Figure 2 include all the 25 side-chains we have studied. However, (see below) Pro binds very weakly. As Pro is not an amino acid in a chemical sense, we provide a range for the 24 side-chains other than Pro (second column of each set). This affects the results for the four enzymes for which Pro is the smallest K_a entry. The set of 25 side-chains in Figure 2 is largely arbitrary. In sharp contrast, the set of all coded amino acid side-chains is complete. Therefore, we also provide the values of ranges for the set of all coded amino acid side-chains with (third columns) and without Pro (fourth columns). Obviously, decreasing the size of the set decreases the ranges in some cases but the important conclusion is that they remain very large.

Isofunctional residue clusters

In view of the huge ranges, plotting of the data in Table 1 requires some compression and the logarithmic compression seemed best. It was implemented by converting the K_a values to $-\Delta G^o(X^{18})_a = RT\ln K_a$, where $T = 294$ K. The results are plotted in Figure 3. Only coded residues are listed in order to reduce the crowding of the Figure. Note, however, that the entries in Figure 3 are not evenly spaced. Instead, for each enzyme there are clusters of isofunctional residues with nearly the same $-\Delta G^o(X^{18})_a$. One might expect that the side-chains in isofunctional clusters should be closely related. Frequently, they are related, but often they are strikingly different. As pointed out by Bigler *et al*. (1993), the value $-\Delta G^o(X^{18})_a$ is determined by opposing influences of hydrophobicity on one hand and of steric bulk, branching, polarity and charge on the other. The finding of isofunctional clusters is important for at least two reasons. One of these reasons is the study of positive Darwinian selection (Laskowski *et al*., 1988) of the contact residues in standard mechanism canonical inhibitors of serine proteinases. In most such inhibitors, the residues which are in contact with the enzyme in the enzyme–inhibitor complex are hypervariable. This phenomenon is especially pronounced in the Kazal inhibitor family and particularly so among the avian ovomucoid third domains (Laskowski & Kato, 1980; Laskowski *et al*.,

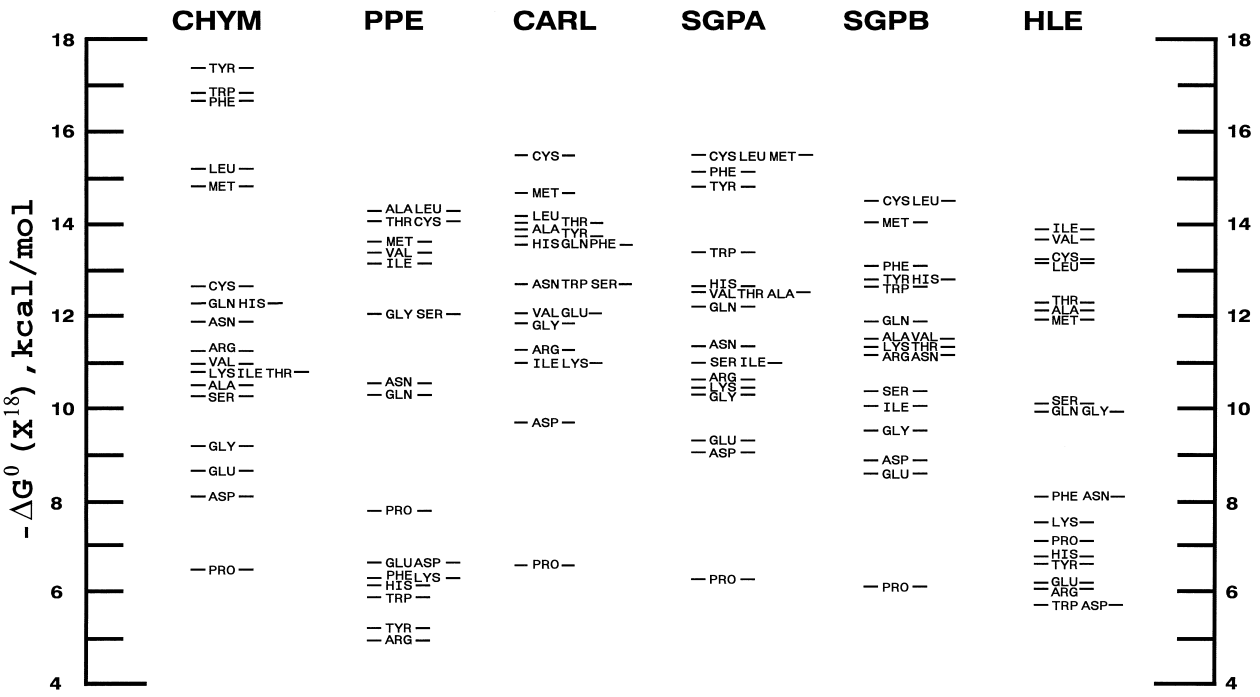

Figure 3. Data for the 20 coded side-chains plotted as $-\Delta G_a^o(X^{18})$ at pH 8.30. The strong binders are at the top, the weak binders are at the bottom. Crowded regions are isofunctional clusters discussed in the text. The relative values for the non-ionizable residues are expected to be unaffected by pH. For ionizable residues they change dramatically.

1987, 1988; Creighton & Darby, 1989). Avian ovomucoid third domains from 153 species were sequenced. At P_1 position, 11 different residues, more than at any other position in the third domain, were found. The composition is $Leu_{65.5}$, Met_{60}, $Ala_{6.5}$, Gln_5, $Val_{4.5}$, Thr_3, Pro_3, Lys_2, $Ser_{1.5}$, Gly_1 and Ile_1 (Apostol *et al.*, 1993)† Inspection of Figure 3 shows that some of the best (Leu, Met for PPE, CARL, SGPA and SGPB) and worst (Pro) residues are represented.‡ Thus, either the target enzyme for the ovomucoid third domains changes among the species examined or the equilibrium constant for inhibition of a constant target enzyme undergoes a very large change.

It is worth noting that for almost all of the enzymes, Leu and Met form an almost isofunctional pair. Leu and Met are the two most highly populated (43% and 39%, respectively) members of the P_1 residue set. It is probable that the evolutionary changes between Leu and Met at P_1 are not selective but neutral. A more familiar example is the frequent variation between P_1 Lys and P_1 Arg in inhibitors whose target enzyme is clearly trypsin, for example, pancreatic secretory trypsin inhibitors (Kazal). In these inhibitors the Arg ⇌ Lys mutations are selectively neutral. However, while isofunctional clusters explain a part of the observed hypervariability, they do not explain it all. K_a values of ovomucoid third domains from closely related species show striking variations (Empie & Laskowski, 1982; Park, 1985; Wynn 1990; Lu, 1994). The implications of the existance of isofunctional clusters upon scanning techniques of protein engineering such as the alanine scan (Cunningham & Wells, 1989; Clackson & Wells, 1995) have already been discussed by Bigler *et al.* (1993).

Gly^{18} as a standard

The set of $-\Delta G_a^o$ (X^{18}) in Figure 3 reflects the variation of binding free energy with the nature of the P_1 residue. However, the set lacks a zero value. It seems natural to choose Gly^{18} as a reference and to assert that side-chains, whose $-\Delta G_a^o$ (X^{18}) is greater than $-\Delta G_a^o$ (Gly^{18}), make a positive contribution to binding while the others are deleterious. There are two possible objections to this choice. The first is that inhibitors with P_1 Gly may not bind to the S_1 subsite of some enzymes but instead place a more favorable residue in that subsite. Huang *et al.* (1995) determined the three-dimensional structure of Gly^{18} OMTKY3-SGPB complex. Gly^{18} clearly binds to the S_1 subsite. The second objection is

† The subscripts denote the number of species in which the P_1 residue was found. The fractions arise because for species in which the P_1 residue was found to be dimorphic each of the two dimorphic residues was given an arbitrary weight of 0.5.

‡ In these considerations we disregard Cys as it is not present in its reduced form in any of the contact positions in standard mechanism canonical inhibitors. The disulfide bridged homodimer of Cys^{18} OMTKY3 is inactive.

Table 2. Ratios of K_a for pairs of isomeric or isosteric residues

	CHYM	PPE	CARL	SGPA	SGPB	HLE
			Isomers			
			Proline (pseudoisomers)			
Ape/Pro	3.3×10^5	4.1×10^5	1.8×10^6	1.8×10^6	3.3×10^5	5.3×10^4
			Branching			
Ahx/Leu	0.42	4.5	2.0	0.77	0.38	0.16
Ape/Val	160	27	160	40	36	0.71
Ahx/Ile	810	32	480	1500	720	0.047
Leu/Ile	1900	7.1	240	2000	1900	0.29
Hse/Thr	26	0.25	2.2	3.3	6.8	0.20
			Isosteres			
			Neutral, heteroatom containing			
Abu/Ser	26	380	42	63	46	420
Val/Thr	1.4	0.30	0.034	1.1	1.3	11
Ape/Hse	8.8	32	2.5	13	7.1	38
Leu/Asn	290	610	13	1200	310	6800
Leu/Asp^o					2.8	
Abu/Cys	0.46	13	0.30	0.030	0.040	2.0
Cys/Ser	57	29	140	2100	1100	220
Ahx/Met	0.80	15	0.83	0.88	0.78	1.4
Asp^o/Asn					120	
Glu^o/Gln					14	
Ahp/Lys^o					4.0	
			Charged			
Asp^o/Asp$^-$					72,000	
Glu^o/Glu$^-$					19,000	
Lys^o/Lys$^+$					25	
His^o/His$^+$					210	
			Ionizable residues, pH 8.30			
			(Charged state in complex not known except for SGPB)			
Asn/Asp	660	950	170	51	46	56
Gln/Glu	520	510	15	150	300	560
Ahp/Lys	3300	5200	740	8400	110	340

based on greater flexibility of glycyl residues compared to all other residues, as Gly lacks a β-carbon. There is no available three-dimensional structure of a free, canonical protein inhibitor with Gly at P_1 position. One might speculate that the main-chain of such free inhibitors might be more flexible at the P_1 Gly position than the main-chains of inhibitors with other residues at P_1. Upon complex formation, the main-chain would rigidify (it is known to be normal in Gly[18] OMTKY3-SGPB complex, Huang *et al*. 1995). This would entail a free energy cost and therefore the use of Gly[18] as a standard would not be fair. Huang *et al*. (1995) showed that the observed $-\Delta\Delta G_a^o$ (Gly18X) values are linearly dependent upon the buried hydrophobic surface area and therefore, excluded this objection to Gly[18] as a standard.

The large values of K_a for Gly[18] variants (Table 1) and of $-\Delta G_a^o$ (Gly[18]) (Figure 3) are an indication that binding of the P_1 side-chain to S_1 cavity is not the only important energetic contribution to complex formation. The interaction of the remaining 11 consensus contact side-chains (Apostol *et al*., 1993) and of the main-chain make a very significant contribution. It is this contribution that is clearly responsible for directing the P_1 side-chains to the S_1 subsite even when their individual interaction is neutral (Gly[18]) or deleterious. The finding that the K_a and the $-\Delta G_a^o$ (Gly[18]) values are similar for the six enzymes may be a coincidence arising from the choice of turkey ovomucoid third domain as the wild-type. Of the six enzymes we study, Gly[18] variant is weakest for chymotrypsin. Yet, turkey ovomucoid third domain contains Arg[21] (P_3') which is exceptionally favorable for chymotrypsin (Lu, 1994). Were Arg[21] changed to another residue, K_a values for the Gly[18] variants would no longer be approximately equal. Chymotrypsin's K_a would be much weaker than those of the rest.

It is interesting to note that one natural ovomucoid third domain (tawny frogmouth, Australian bird belonging to the goatsucker family) has Gly at P_1 position (Apostol *et al*., 1993). It is a moderately good inhibitor of PPE, CARL and SGPB. It is likely that its true target enzyme has a very small or nonexistent S_1 pocket.

Binding of Pro[18]

Pro[18] is very poor and deleterious compared to Gly[18] for all the enzymes we study. It is the worst member of the set for all but the two elastases. Elastases have small S_1 cavities. Therefore, large and charged residues are also very deleterious. Modeling suggests that Pro[18] side-chain is difficult to wedge into the S_1 pocket. Furthermore, the N

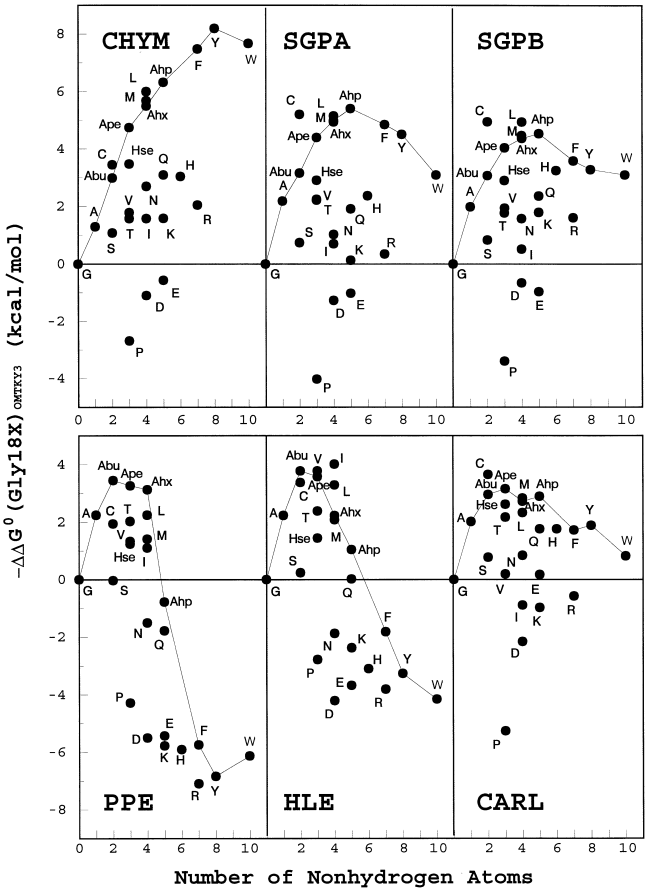

Figure 4. The values of $-\Delta\Delta G^o$ (Gly18X) for 20 coded and five non-coded amino acid residues obtained at 21°C, pH 8.3. The homologous series of residues Gly, Ala, Abu, Ape, Ahx and Ahp is extended by hydrophobic Phe, Tyr and Trp and plotted against the number of non-hydrogen atoms in the side-chains. These are a rough measure of both hydrophobicity and of steric bulk.

atom of Pro[18] cannot approach the enzyme as closely as the atoms of other P_1 residues. A complex of Pro[18] OMTKY3 with SGPB has been crystallized and a high resolution structure was obtained (K. Huang, W. L., M. L., S. A. & M. N. G. James, unpublished results). This complex differs from the complexes with all other residues at P_1. In the other complexes, the positions of all contact residues, other than P_1, are superimposable. In the Pro[18] complex some of these residues make small motions to aid in the wedging of P_1 Pro into the S_1 pocket. Since Pro is the only member of our set that is not an amino acid in the chemical sense, neither a strictly isosteric nor a strictly isomeric comparison is possible. However, Ape is nearly a Pro isomer. Comparisons of K_a values for all six enzymes for Ape[18] and Pro[18] is given in the first row of Table 2.

Since Pro[18] is so deleterious, it is surprising that it was found three times among the 153 species of birds whose ovomucoid third domains were sequenced (Apostol *et al*., 1993). The Pro[18] containing species are black night heron and the two species of cassowaries we examined. Furthermore, in alligator ovomucoid, which consists of four Kazal domains (March, 1980), the third domain has Pro at P_1 position. Over the last 15 years we have tested alligator ovomucoid and the P_1 Pro containing avian third domains for inhibition of likely targets such as prolyl endopeptidases. None was effective.

Effect of residue size on binding

Consider a homologous (in a chemical sense) series of amino acid side-chains Gly, Ala, Abu, Ape, Ahx and Ahp. The last four entries are from Bigler *et al*. (1993), who also had Ala but not Gly, an important residue. Technical problems also prevented Bigler *et al*. (1993) from going beyond Ahp but they chose to add Phe to the series. We further add Tyr and Trp and plot the $-\Delta\Delta G^o$(Gly18X) values *versus* the number of non-hydrogen atoms in the side-chain.

For all six enzymes $-\Delta\Delta G^o$ increases (Figure 4) with the number of atoms in the side-chain, rises to a maximum and then declines. The qualitative explanation is straightforward. All six pockets are hydrophobic. Increasing the size increases the hydrophobicity of the residues in this series. Thus, binding improves with increasing size. However, all pockets are finite. As the pocket size is exceeded, energy consuming changes must be made to accommodate the oversized side-chain.

This qualitative explanation is sufficient to emphasize the point made by Bigler *et al*. (1993) that the $-\Delta\Delta G^o$(Gly18X) values are frequently a result of two large and opposing terms, exemplified by hydrophobicity and inelasticity of the too small pocket. As a result, a small and a large residue may turn out to have the same $-\Delta\Delta G^o$(Gly18X) values. As will be seen below, branching, polarity and charge also can oppose hydrophobic bonding, thus leading to some truly unexpected isofunctional sets (Figure 3).

A good start was made in quantitative treatment of the ascending sides of Figure 4. Huang *et al*. (1995) measured the buried hydrophobic surface area for the Gly[18]-, Ala[18]-, Abu[18]-, Ape[18]-, Ahx[18]-, and Ahp[18] OMTKY3 complexes† with SGPB. Our $-\Delta\Delta G^o$(Gly18X) values were then plotted *versus* Δ (the buried hydrophobic surface area) yielding an impressive straight line (Figure 8 of Huang *et al*., 1995) with a slope of 34.1 cal/mol per Å^2. As discussed by Huang *et al*. (1995), this value for hydrophobic strength is intermediate among the

† The structures of complexes with Gly[18] and Ala[18] were experimentally determined. Those with the non-coded amino acids were only modeled as the Purdue group was unable to supply sufficient quantity of these variants (Bigler *et al*. 1993) for X-ray crystallography of complexes.

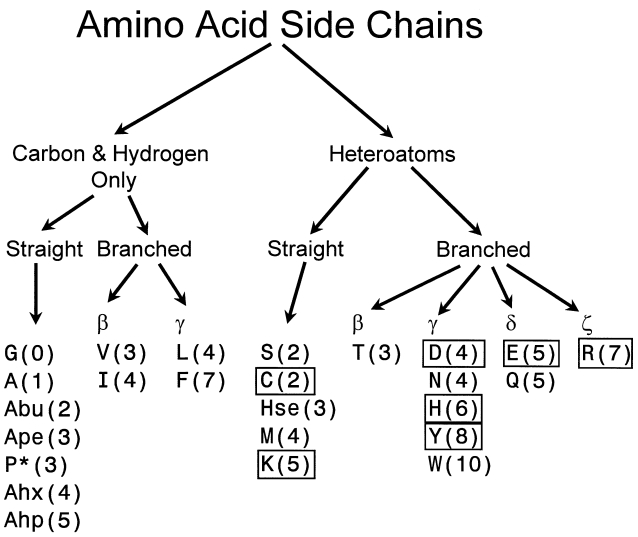

Figure 5. A classification of the 25 amino acid side-chains studied in this paper. The coded residues are given their one letter symbols, the non-coded ones three letters. The ionizable residues are boxed. The number after each side-chain indicates the number of non-hydrogen atoms. *P, Proline is not an amino acid in a chemical sense, therefore, it poses difficulties with classification.

currently proposed values in the literature. Huang *et al.* (1995) modeled the same series of complexes for SGPA and CHYM. Excellent straight $-\Delta\Delta G^{o}$(Gly18X) *versus* Δ (buried hydrophobic surface area) lines with slopes ranging from 34 to 40 cal/mol per Å^2 were obtained. The Huang *et al.* (1995) straight lines arise from the Δ(hydrophobic buried surface area) results increasing less than linearly with the number of non-hydrogen atoms of the side-chains. This non-linearity matches the non-linearity of the ascending sides in the plots of Figure 4 to produce the straight lines. Therefore, the hydrophobic contribution to binding cannot be simply calculated from a formula with a linear dependence on the number of methylene groups in the side-chain.

The P_1 side-chains of free inhibitors are highly exposed and very mobile. Upon insertion in the S_1 cavity of the enzyme they become highly restricted. This was recently demonstrated in the SSI-subtilisin system by deuterium NMR studies (Tamura *et al.*, 1996). This restriction is associated with an entropy loss (Doig & Sternberg, 1995). Clearly, such an entropy loss must arise in our homologous series of side-chains. Presumably, it is one of the many effects that are incorporated into the $\Delta\Delta G^{o}$(Gly18X) *versus* Δ (buried surface area) correlation.

We know much less about the descending side of the $-\Delta\Delta G^{o}$(Gly18X) plots. In the case of SGPB, Phe^{18}-, Tyr^{18} and Trp^{18}- bind less well than Leu^{18}- in spite of greater buried hydrophobic surface area. Bateman *et al.* (1996) obtained the structures of these complexes and showed that the aromatic side-chains fit without distorting the S_1 pocket. Why then do they bind more poorly than the less hydrophobic Leu^{18}? In order to fit in without distorting the pocket, they adopt an unfavorable χ_2 angle. Bateman *et al.* (1996) calculate that this uncommon angle costs about 3kcal/mol, somewhat more than is needed to explain the decline in binding of the aromatic residues compared to Leu^{18}.

The most dramatic descents occur for the two elastases, PPE and HLE. The P_1 preference of both enzymes is very similar, as judged by Pearson product moment correlation (see below). Their maximal $-\Delta\Delta G^{o}$(Gly18X) residues are essentially the same; however, the descending sides differ strongly. The descent in PPE is much steeper than in HLE. This extends to the more complex residues as well. Our conclusion is that HLE pocket is much more flexible than that of PPE. The conclusion that the S_1 pocket of HLE is more flexible than those of the five other enzymes is reached here on three different bases. The other two are the preference for Val and Ile, of which the other enzymes are intolerant and the non-additivity seen only in HLE. These two effects are discussed below. Furthermore, Bode *et al.* (1989) have already concluded on the basis of X-ray crystallographic studies that HLE's S_1 pocket is more flexible than PPE's. It is clear that more structures involving both PPE and HLE will be needed to resolve these problems. The sharp decline, especially for PPE, does not continue for the largest residues. Instead, there is a levelling out toward a relatively constant low value. It's cause is unknown.

Other amino acid side-chains

Figure 5 provides an elementary classification of amino acid side-chains. These are first divided into two groups. The first contains only carbon and hydrogen atoms. The second contains in addition heteroatoms (oxygen, sulfur and nitrogen). They are next divided into straight and branched chains. In analyzing the various groups above we resort whenever possible to binary comparisons between isomers and isosteres.

As pointed out by Bigler *et al.* (1993), most points in Figure 4 fall below rather than above the solid line consisting of hydrophobic, straight chain side-chains. This observation, however, is now somewhat less striking than it was in Bigler *et al.* (1993), since Cys is often better than Abu, and Ile is better than Ahx for HLE. Neither side-chain was available in the earlier study. The second interesting aspect is that the large range for CHYM is achieved mainly above the Gly zero line, i.e. it is the strong improvement in binding from Gly to Tyr that accounts for most of the range. The very large range for PPE is mostly achieved below the line as some substituents are very deleterious.

Branching

Of the 20 coded amino acid side-chains, 13 are branched (Figure 5). For aliphatic, hydrophobic residues, the presence or absence of branching such as Ape/Val, Ahx/Leu and Ahx/Ile comparisons or the position of branching (Leu/Ile) have only small effects on the transfer free energy from water to organic solvent (e.g. Radzicka & Wolfenden, 1988 and references therein). Such results are simple to rationalize. The surface areas of unbranched and variously branched isomers are closely similar and the cavities in the solvents readily adjust to different shapes of the side-chains. The branching problem was also addressed for residues buried in the protein interior. Sturtevant, Richards and their co-workers are engaged in a major study of binding of X^{13} S-peptide variants to ribonuclease S-protein to yield X^{13} ribonuclease S complexes. They determine three-dimensional structures and ΔG^{o}, ΔH^{o} and ΔCp^{o} of binding of the variants. In a recent paper, they write: ''substitution of wild-type methionine at position 13, with either leucine or isoleucine, unsurprisingly produces little or no change in the free energy of association'' (Thomson *et al*., 1994). While we are unable to find equally strongly worded quotes in the other papers, we believe that this quote fairly summarizes the work of Kellis *et al*. (1988, 1989), Matsumura *et al*. (1988), Shortle *et al*. (1990), Sandberg & Terwilliger (1991), Eriksson *et al*. (1992), Serrano *et al*. (1992) and Matthews (1993). These data are rather compactly organized and discussed by Lee (1993). Interchanges of Ahx (when available), Met, Leu and Ile have no effect upon the free energy of stabilization of the interiors of globular proteins. These results, in turn, are consistent with the considerable liquid-like adaptability of the protein interiors.

In spite of considerable similarities between the effects of side-chains on the stabilization of proteins and of binding to the S_1 cavity of serine proteinases, the effects of branching are sharply different. Grøn *et al*. (1992) summarize their wide reading of the proteolytic enzyme literature by ''...substrates with β branched amino acids Val and Ile at P_1 are very poor. This inefficiency in cleaving substrates with a β-branched P_1 residue is common to many proteolytic enzymes...''. Note that they properly extend the field beyond that of serine proteinases. They cite numerous references and we could add many more. We are, however, aware of at least two exceptions. One is human leukocyte elastase, discussed below. The other is the interaction between carboxypeptidase A and potato chymotrypsin inhibitor, where β-branched P_1 residues are preferred to their isomers (Molina *et al*., 1994).

The contrast between Leu and Ile being isofunctional (protein thermodynamicists) and Ile being greatly inferior (enzymologists of proteinases) gives a strong feeling of two cultures. We assume that the explanation lies in the far greater rigidity of preformed cavities than of protein interiors. The Ahx/Leu ratios in Table 2 are all relatively small. The greatest deviations from unity (but still less than an order of magnitude) are for the two elastase PPE and HLE, where packing of these side-chains into a small cavity is the overriding concern. The relative Ahx/Leu equivalence aids those who wish to discuss the S_1 cavity phenomenon using only coded variants. They can talk about the Leu/Ile ratio almost as if it were Ahx/Ile.

Now let us turn to the Ape/Val ratio, where Val is highly deleterious for the first five enzymes. It is slightly favorable for HLE. The deleterious effects range from 2 to 3 kcal/mol. The comparison between Ahx/Ile is a repeat; except, now the unfavorable effect is an order of magnitude higher as is the favorable effect for HLE. The Leu/Ile comparison restates the results; except, now in three cases the ratio reaches 2000, a large effect indeed. The reason seems simple. It is much easier to pack in the Ahx and the γ-branched Leu residue than the β-branched Val and Ile.

It is easy to conclude from these data that all β branching is highly deleterious, except for HLE. However, this conclusion becomes tempered by the next entry, the Hse/Thr comparison. While Thr is deleterious for four enzymes and advantageous for the two elastases: PPE and HLE, the effects are quite modest. The relevant X-ray structures are not yet analyzed but an attractive explanation is that the environment into which the additional β-branched group is forced is somewhat polar, such as a peptide bond, thus explaining the relative preference of -OH *versus* $-CH_3$ at this position. The direct comparison of isosteric Val/Thr below confirms this speculation, but clearly only the structures can decide. The contrast between the strong aversion to Ile by five of the enzymes and a weak preference for Ile by HLE is the second of our three arguments for flexibility of HLE S_1 pocket.

Neutral, heteroatom containing side-chains

Whereas, some heteroatom containing side-chains are neutral (Ser, Thr, Hse, Met, Asn, Gln) over the whole accessible pH range, others are ionizable. Such side-chains are involved in a pH-dependent equilibrium between their protonated and deprotonated forms. For these amino acids, one of the two forms is always neutral. We designate this form Asp^{o}, Glu^{o}, Cys^{o}, Tyr^{o}, His^{o}, Lys^{o} and Arg^{o}. Equilibrium constants K_a for such forms describe the binding of the inhibitor, in which the P_1 residue is in X^{o} form yielding a complex in which the P_1 residue remains in the X^{o} form. Such constants can be readily compared to those of neutral side-chains. A preliminary analysis suggests that K_a values at pH 8.30 for His, Cys, and Tyr are for the His^{o}, Cys^{o} and Try^{o} forms as defined above. Tyr was already treated as Tyr^{o} in the discussion of size

effects. In this section Cys is treated as Cys°. Unfortunately, we have no appropriate isostere for His comparison.

The Abu/Ser ratio is quite large for all six enzymes. As might have been expected, the $-CH_3 \rightharpoonup -OH$ replacement is deleterious. The large effect is somewhat surprising. For example, Ser is a poorer binder than Ala for all six enzymes. It is also striking that Ser is particularly deleterious for the two elastases, PPE and HLE. We offer no explanation. The very deleterious effect of the $-CH_3 \rightharpoonup -OH$ replacement does not persist for the Val/Thr substitution. In fact, Thr is better than Val for PPE and much better than Val for CARL. It is virtually equivalent for CHYM, SGPA and SGPB. Only in the case of HLE is Thr significantly deleterious when compared to Val. In the discussion of branching, we noted that Val was quite deleterious compared to Ape for all of our enzymes except HLE (see Table 2). As Thr is smaller than Val, its steric hindrance problems might be smaller. It may also be that β branching forces one of the γ-methyl groups in Val into a slightly polar environment. Such an environment might be better tolerated by the γ OH group of Thr. The Val/Thr comparison is another contrast between our rigid pocket results and the stabilization of flexible interiors of globular proteins where Val is much better than Thr (Matthews, 1993).

Another case where the very large $-CH_3 \rightharpoonup -OH$ effect is not observed is the Ape/Hse comparison. In this case, Ape is appreciably better than Hse, but about an order of magnitude less good than the Abu/Ser ratio. As might have been anticipated, Leu is much better for the six hydrophobic pockets than Asn. Note that there one oxygen and one nitrogen atom substitute for the two carbon atoms. It is startling to us that Leu is very closely similar to Asp° in the binding to SGPB. The three-dimensional structures of both members of this pair (Huang *et al.*, 1995; Huang, 1995) are available. Before offering an interpretation, we would like to have Leu/Asp° ratios for the remaining five enzymes. Such work is now in progress.

In strong contrast to the $-CH_3 \rightarrow -OH$ replacement in Abu, the $-CH_3 \rightharpoonup -SH$ replacement in this side-chain generally favors Cys over Abu, by relatively large factors for SGPA and SGPB. Only the two elastases PPE and HLE favor Abu. The strong preference for Cys over Abu makes Cys the best P_1 residue for CARL and SGPB and the best coded residue for SGPA (non-coded Ahp is better). This finding may be of physiological interest. It is possible that hydrolysis of P_1 Cys bonds in their substrates is an important role of some bacterial enzymes provided that they act in reducing environments. The substitution of $-CH_3 \rightharpoonup -SH$ yields ratios that differ from the $-CH_2- \rightarrow -S-$substitutions. The comparison between Ahx and Met was addressed by Bigler *et al.* (1993). The current paper does not offer any additional relevant data. Therefore, only the results are given in Tables 1 and 2 and in Figures 3 and 4.

The remaining isosteric comparisons deal with the replacement by a nitrogen atom. Both Asp° and Glu° bind significantly more strongly to the SGPB pocket than do their amides. The Asp°/Asn ratio (2.8 kcal) is particularly large. In this case, the structural comparisons based on X-ray structures (Huang, 1995) are already available. We await the Asp°/Asn ratios for several more S_1 pockets as the pH dependence analysis was only completed for SGPB (Qasim *et al.* 1995). The Ahx/Lys° ratio reflecting the $-CH_3 \rightharpoonup -NH_2$ substitution seems very small to us. This may well be true as (see below) SGPB has a very large preference for Lys, compared to the other five enzymes.

pH dependence

Our data were obtained at pH 8.30. The K_a values for OMTKY3 associating with any of the six enzymes we study are pH-dependent. An important question is whether the relative K_a values for the various P_1 variants are pH-independent and therefore, general. For the non-ionizable P_1 residues the answer seems to be that they are pH-independent for the six enzymes we study. What about ionizable residues? For some of these, the relative K_a values are clearly highly pH-dependent. Let us consider the simplest possible situation. Because the P_1 residue in the free inhibitor is known to be highly exposed, it ionizes independently of the other residues of the inhibitor. Its ionization is characterized by a simple ionization constant, K_f, where f stands for free inhibitor. Similarly, in complex with one of the six enzymes we study here, the P_1 residue is imbedded in a hydrophobic cavity and it does not interact with other ionizable groups of the enzyme or of the inhibitor. Therefore, its ionization is characterized by a simple constant, K_c, where c stands for complex. Given these assumptions, we can write (Qasim *et al.*, 1995).

$$R = \frac{K_a^{\text{ionizable}}}{K_a^{\text{nonionizable}}} = R^0 \frac{1 + \frac{K_c}{[H^+]}}{1 + \frac{K_f}{[H^+]}} \qquad (1)$$

where R is the ratio of K_a with an ionizable P_1 residue to the K_a with a non-ionizable P_1 residue. At very low pH ($[H^+] \gg K_c$, $[H^+] \gg K_f$) it clearly follows that $R = R^o$. R^o is the ratio between association constant K_a for the ionizable P_1 in its fully protonated form both in the free inhibitor and in complex and K_a for the nonionizable P_1 residue. For the uncharged amino acids Asp, Glu, Cys and Tyr this is the Asp°, Glu°, Cys° and Tyr° form we have already discussed. However, for the cationic amino acid side-chains, His, Lys and Arg, this is the His^+, Lys^+ and Arg^+ form. At very high pH, where $[H^+]$ is very small, the limit of R in equation (1) is $R^o \times (K_c/K_f)$. At the high pH limit, we are measuring the ratio of the equilibrium constant for binding of the deprotonated P_1 side-chain, both in the complex and in

the free inhibitor, to that of the non-ionizable reference side-chain. Specifically, we deal here with the Asp^-, Glu^-, Cys^- and Tyr^- constants for uncharged side-chains and with His^o, Lys^o and Arg^o for the cationic side-chains.

The effect of charge

For each of the seven ionizable amino acid residues (Figure 5), we can make an isosteric comparison such as Asp^o/Asp^- and Lys^o/Lys^+. However, values are now available only for the four cases (all in SGPB) for which the full pH dependence of K_a was determined (Qasim *et al.*, 1995 and unpublished; Table 2). It is seen there that the uncharged form of each of these four side-chains binds to the S_1 cavity of SGPB much more strongly than the charged form. This is true for both uncharged (Asp, Glu) and cationic (His, Lys) members of the set. We strongly expect that it would also be true if the data were available for Cys, Tyr and Arg. Similarly, we expect that it also would be true for all seven ionizable side-chain binding not just to SGPB but to all S_1 cavities that are studied here. Indeed, we have a significant number of preliminary data confirming these conclusions.

The P_1 side-chains in all standard mechanism canonical protein inhibitors of serine proteinases are highly exposed to solvent (Read & James, 1986; Bode & Huber, 1992). Therefore, when the P_1 side-chain is charged, the charge resides in a high dielectric medium. Upon transfer to the hydrophobic S_1 cavity of one of the six enzymes we study, it is forced to reside in a much lower dielectric environment. Transfers of a charge from high to low dielectric are strongly disfavored (Born, 1920). This provides a satisfying qualitative explanation for the preference for neutral over both positive and negative charges.

However, the neutral forms of both types of ionizable residues will not be favored by all S_1 cavities in all serine proteinases. We expect trypsin (with a negative S_1 pocket due to the presence of Asp^{189}) to strongly favor Lys^+ over Lys^o and to disfavor Asp^- compared to Asp^o. Similarly, we expect positive S_1 cavities of enzymes such as (Glu Sgp) the glutamic acid specific *Streptomyces griseus* proteinase (Yoshida *et al.*, 1988; Komiyama *et al.*, 1991; Nienaber *et al.*, 1993) to strongly favor Glu^- over Glu^o and to disfavor Lys^+ over Lys^o.

Returning to Table 2, we note that the Asp^o/Asp^- and Glu^o/Glu^- ratios are truly huge, exceeding four and approaching five orders of magnitude. On the other hand, the His^o/His^+ and especially Lys^o/Lys^+ ratios, while still quite large, are much more modest. SGPB appears to favor the cationic residues His, Lys and Arg far more than the other five enzymes (see below). However, it still appears that the Asp^o/Asp^- and Glu^o/Glu^- ratios for CHYM are much larger than the Lys^o/Lys^+ ratio. We tentatively conclude that while the primary effect we are observing is the change in the dielectric constant, there is a secondary effect in all six enzymes we study, relatively favoring positive charges and disfavoring negative charges on the buried P_1 side-chains. Clearly, more data are required.

Preliminary measurements suggest that for all six enzymes we study, the K_a values at pH 8.3 are for His^o, Cys^o, Tyr^o and Arg^+. On the other hand, we are still awaiting pK_c measurements for Asp, Glu and Lys for enzymes other than SGPB. We were clearly hasty to suggest (Bigler *et al.*, 1993) that these residues were in their charged form at pH 8.3.

Specificity of serine proteinases

We have thus far focused on the comparisons within columns of Table 1, addressing such questions as the best and worst residues, isofunctional clusters, Leu/Ile ratio, uncharged/charged ratio, all for a particular enzyme. In this section, we address questions involving within row comparison and a related column to other columns comparison. We have (Laskowski *et al.*, 1989) defined the concept of the most specific residue for each enzyme. This most specific residue is not just a function of the target enzyme but also of all the other members of the enzyme set under consideration. In view of that, our set of six enzymes is so highly arbitrary that it can almost be regarded as capricious. Sets of all human serine proteinases or of all serine proteinases from organisms likely to infect eggs are probably much more reasonable sets to investigate. Proceeding in the manner recommended by Laskowski *et al.* (1989), we have selected for each of the six enzymes a P_1 residue that is the most specific within our set. These are boxed in Table 1. For CHYM, it is Trp, not the best Tyr. The reason is that CHYM has the largest S_1 pocket of the six enzymes. Therefore, while changing Tyr^{18} to Trp^{18} is only slightly poorer for CHYM, it is appreciably poorer for the nearest competitor, SGPA. While Trp^{18} is actually better than Tyr^{18} for PPE, this is without significance because both Tyr^{18} and Trp^{18} are much too large for PPE and HLE. For PPE, with a very small pocket, our choice was Ala^{18} (Abu^{18} is a close competitor). The reasoning is essentially the same as for CHYM, except in reverse; small residue for a small pocket. We found the choice of Glu (at pH 8.3, please remember that relative binding constants of ionizable residues depend upon pH) for CARL somewhat startling. We did not expect it. However, Glu is much better for CARL than for any other enzyme in the set. While SGPA and SGPB are sister enzymes with closely similar specificities, for most P_1 residues SGPA has higher K_a values by about a factor of 10 or more. Therefore, the problem of finding the most specific residue for SGPA is largely reduced to finding a moderately large residue that was better for SGPA than for CHYM. Ahp^{18} fits the bill. Note that if a coded residue was required Leu would also work. Because SGPA is typically more strongly inhibited than SGPB, difficulty in finding a most specific P_1

Table 3. The P_1 specificity of six serine proteinases

	Pearson product-moment correlations (r) Coded residues only					
	CHYM	PPE	CARL	SGPA	SGPB	HLE
CHYM	1.00					
PPE	−0.10	1.00				
CARL	0.66	0.39	1.00			
SGPA	0.87	0.30	0.89	1.00		
SGPB	0.83	0.23	0.87	0.96	1.00	
HLE	0.03	0.94	0.42	0.41	0.35	1.00

residue for SGPB was expected. However, we note in Table 1 that at pH 8.30, the three ionizable, cationic variants Lys^{18}, His^{18}, and Arg^{18} inhibit SGPB better than SGPA. Lys^{18} was chosen as most specific for SGPB. The finding that all cationic side-chains prefer SGPB to SGPA raises a question. Is the effect due to positive charges†, to nitrogenous substituents or to both. It seems clear that in binding to SGPB at pH 8.30 we deal with Arg^{+} and His^{o}. Lys ($pK_c = 8.7$) is an intermediate case. Comparisons will become much easier when more pK_c values for the other S_1 pockets are determined. The most specific and also the best P_1 residue for HLE is Ile. Much of this paper was already concerned with the abhorrence of β-branched aliphatic residues by the first five enzymes in Table 1. In contrast, HLE finds such residues better than their straight chain and γ-branched counterparts.

Aside from picking out the most specific residue for each enzyme, we can instead think of each column in Table 1 as a vector in 25-dimensional amino acid space. If we restrict the set to coded amino acids only, the space reduces to the more familiar 20-dimensional amino acid space. However, if we were to leave the entries as equilibrium constants, the small ones (see Figures 2 and 3) would have no measurable effect on the direction of the vector, which would be determined by the highest value or a few highest values. This is avoided when logarithms of $K_a(X^{18})$ or association free energies, ΔG^o (X^{18}), are considered. Therefore, Figure 3 is the appropriate reduction of data of Table 1, as it both restricts to the coded set and provides free energies of association. The restriction to the coded set is easy to justify as proteinases act on substrates in which the residues are overwhelmingly members of the coded set with but a few post-translationally modified outliers.

With the data of Figure 3 in hand, there are numerous procedures for calculating the relationship between the vectors given by the columns. We are still evaluating these procedures. The Pearson product-moment correlation (Campbell, 1989), is very familiar and we employ it here:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \qquad (2)$$

where $r = +1$ perfect positive correlation, $r = 0$ no correlation, $r = -1$ perfect negative correlation.

In the specific application in this section the index, i, refers to the kind of coded residue present at position 18 in X^{18}OMTKY3, n is 20, the number of kinds of coded amino acids. x values are $-\Delta G_a^o(X^{18})$ values for the interaction with one of the enzymes and the y values are for the interaction with the other.

Table 3 is a list of all the possible pairwise comparisons for six enzymes. As expected, the closely related SGPA and SGPB have the most positive r value (0.96). The next closest relationship is between PPE and HLE (0.94), in spite of differences that appear to arise from considerable rigidity of PPE and relative flexibility of HLE. This high correlation is a testimonial to the insight of the late Aaron Janoff, who named HLE elastase (Janoff & Basch, 1971). All six enzymes we have studied are strongly inhibited by wild-type OMTKY3, whose P_1 residue is Leu^{18}. All of them have hydrophobic S_1 cavities. Therefore, we were startled to find a negative correlation within the set. It is (−0.10) between CHYM and PPE. Many proteinase chemists are also startled when they are told this. Yet the explanation seems straight forward. Phe^{18}, Tyr^{18} and

† The stronger binding of Lys^{18} to SGPB than to SGPA was recognized in our group sometime ago (Park, 1985; Lin 1986) based on results with Japanese quail ovomucoid third domain, which has…CysProLys↓Asp…sequence at the reactive site. An explanation for the difference was advanced on the basis of a presumed difference in the amino acid sequence between SGPA and SGPB. The reported amino acid sequences showed that, in SGPA residue, 191 was Gln (Johnson & Smillie, 1974) while in SGPB (Jurasek *et al.*, 1974), it was Glu (Chymotrypsinogen numbering and Fujinaga *et al.* (1985) alignments are employed.). However, subsequent to our explanations, Henderson *et al.* (1987) published nucleotide sequences of SGPA and SGPB. Residue 191 is Glu in both sequences. In order to confirm it, we have isolated the relevant peptide from both enzymes and obtained its amino acid sequence. It was found to be Glu^{191} in both. After this was communicated to Professor Smillie, he kindly rechecked the original assignment of Gln^{191} in SGPA and found that, along with many amide assignments of that period in protein sequencing, it was largely a guess. Therefore, it seems quite clear that SGPA and SGPB both have Glu^{191}. On the other hand, this leaves us without molecular explanation for the SGPB preference over SGPA for cationic side-chains.

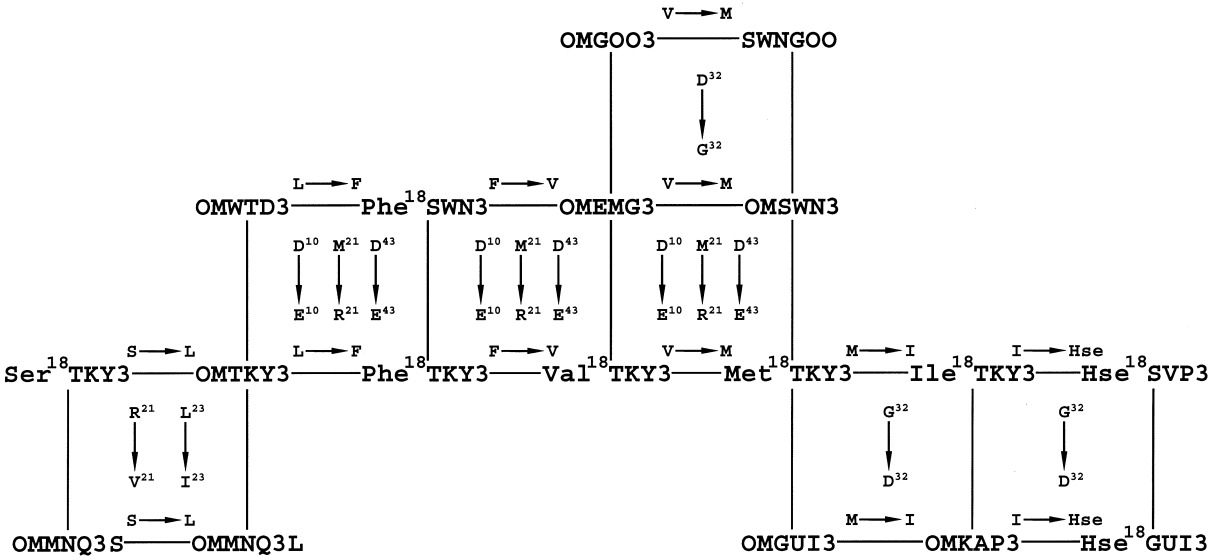

Figure 6. Additivity cycles involving single P_1 substitutions. Each entry differs from its horizontal neighbor by a single substitution at P_1. The entries in the long horizontal line are all from Table 1. WTD stands for West Indian tree duck, EMG for emperor goose, SWN for black-necked swan, GOO for greylag goose, MNQ for Montezuma quail, GUI for guinea fowl, KAP for Kalij pheasant, SVP for silver pheasant. The Hse^{18} OMSVP3 and Hse^{18} OMGU13 were prepared by opening the Met^{18} – Glu reactive site peptide bond by cyanogen bromide cleavage followed by enzymatic resynthesis. SWNGOO is a chimera or residues 6 to 18 of OMSWN3 and 19 to 56 of OMGOO3 (see Wieczorek & Laskowski, 1983 and Bigler *et al*., 1993 for more detailed descriptions).

Trp^{18} are the best P_1 residues for CHYM. They are nearly the worst for PPE. As it lacks enzymes with a preference for positve and negative P_1 residues, our set appears incomplete. We have already remedied some of these deficiencies and a more extended publication on the use of inhibitor binding constants to define the specificity of serine proteinases is in preparation.

Generality of the data

Table 1, whose data were exhaustively discussed above, applies only to X^{18} OMTKY3 and to that only at 21 (±2)°C and at pH 8.30. However, it is reasonable to expect that the relative data in each column of Table 1 may apply to a closely related set of avian ovomucoid third domains or even to all members of the Kazal family of serine proteinase inhibitors. If the data for these sets agree well, we talk about intrascaffolding additivity as all the members of these sets share the common Kazal family scaffolding. We could be more venturesome yet. As the canonical standard mechanism inhibitors have a common conformation of the contiguous combining region, interscaffolding additivity might be expected. If this is true, relative values in the six columns of Table 1 can be used to predict or confirm the relative K_a values of P_1 variants of BPTI, SLPI, SSI and eglin c. When this is found to hold (approximately), we can turn even more venturesome. As ''inhibitors are substrates'' (Laskowski & Kato, 1980) we can also seek a correlation between $\log(k_{cat}/K_m)$ values for a series of substrates that differ from one another only at P_1 and $\log K_a$ values of X^{18} OMTKY3.

Strict additivity criterion for avian ovomucoid third domains

Over decades of work on avian ovomucoid third domains, the Purdue group acquired 92 unique (Apostol *et al*., 1993) and 32 semisynthetic (Wieczorek & Laskowski, 1983; Bigler, 1991; Bigler *et al*., 1993), chemically modified, and totally synthetic (Warne, 1990; Ogunjobi & Laskowski, un-

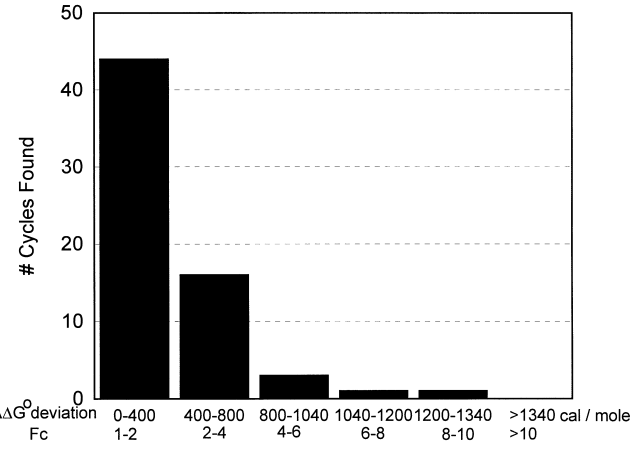

Figure 7. Results of additivity tests on proteins in Figure 6. Of the $6 \times 12 = 72$ potential cycles, data are available for 65. As shown in this graph, 44 were found with $1 < F_c < 2$ or within experimental error, 24 with $2 < F_c < 4$, five with $4 < F_c < 6$ and only two with $F_c > 6$. The F_c values, which were smaller than unity, were replaced by their inverses for ease of comparison.

Table 4. Correlation with literature data

Free energies of transfer					
Enzyme	Transfer	No. of residues compared	*r*	*s*	Reference
CHYM	Water to *n*-octanol	6	0.99	0.51	a
CHYM	Water to ethanol	17	0.85	0.45	b,c
Standard mechanism Inhibitors					
Enzyme	Inhibitor[(d)r]	No. of residues compared	*r*	*s*	Reference
CHYM	X^{15}BPTI	4	0.96	0.39	e,f,g
CHYM	X^{72} SLPI	6	0.99	0.80	h
PPE	X^{15} BPTI	7	0.95	1.34	e,f,i
CARL	X^{73} SSI-BPN′[(j)]	19	0.93	0.72	k
HLE	X^{15} BPTI	10	0.96	0.74	e,f,l
HLE	X^{72} SLPI	6	0.98	0.70	h
Substrates, k_{cat}/K_m values					
Enzyme	Substrate[(d)]	No. of residues compared	*r*	*s*	Reference
CHYM	"statistical"[(n)]	16	0.97	0.82	n
CHYM	Ac-X-OCH_3	8	1.00	0.94	o
CHYM	Boc-AAX-SBzl	9	0.94	0.64	p
PPE	Boc-AAX-SBzl	9	0.85	0.26	p
CARL	Abz-FGPX-Y′D[(q)] Savinase	11	0.87	1.15	r
CARL	Abz-FGPX-Y′D BPN′	11	0.88	1.05	r
CARL	Suc-AAPX-pNA	14	0.58	1.12	s
SGPB	Ac-PAPX-NH_2	7	0.86	1.25	t
HLE	Boc-AAX-SBzl	9	0.98	0.54	p

a, Hansch & Coates (1970) G, A, Abu, Ape, Ahx, Ahp. **b**, Tanford (1962) all coded except H,I,P. **c**, The D and E values corrected according to Levitt (1976). **d**, See abbreviations. **e**, Beckmann *et al.* (1988). **f**, Beckman *et al.* (1989). **g**, K, Ape, Ahx, M. **h**, Eisenberg *et al.* (1990) L, G, F, V, R, K. **i**, L, A, Abu, Ape, V, Ahx, M. **j**, Interaction of OMTKY3 variants with CARL is compared to the interaction of SSI variants with BPN′. **k**, Kojima *et al.* (1991) all 20 coded except C. **l**, K, L, A, Abu, Ape, V, Ahx, M, I, G. **m**, Schellenberger *et al.* (1991) analyzed all published k_{cat}/K_m values for chymotrypsin, assuming additivity. **n**, Schellenberger *et al.* (1991) Y, W, F, M, Ahp, Ahx, L, Ape, N, Abu, S, K, A, R, V, G. **o**, Dovorska *et al.* (1972) G, A, Abu, Ape, Ahx, F, V, I. **p**, Harper *et al.* (1984) A, S, Ape, Ahx, L, M, N, E, F. **q**, The log K_a for the CARL-OMTKY3 variants is compared to log(k_{cat}/K_m) to the listed substrate variants hydrolyzed by savinase. In the next line the results are compared to the hydrolysis by BPN′. As already pointed out, BPN′ and CARL interactions with OMTKY3 variants are highly correlated. Similar high correlation is expected for savinase but it was not yet experimentally tested. **r**, Grøn *et al.* (1992) G, A, V, I, F, W, N, P, D, R. s, Estell, D. (personal communication) G, A, S, V, T, L, M, K, Q, E, H, F, R, Y. t, Bauer (1978) G, A, V, L, F, Y, W.

published results) variants. For most of these variants K_a values for interaction with the six enzymes studied in this paper were determined (Empie & Laskowski, 1982; Laskowski *et al.*, 1983, 1989; Park, 1985; Wynn, 1990; Lu, 1994; Park *et al.* unpublished results). More strikingly, these data have been acquired by the same measuring technique as that employed for determination of data in Table 1 and possibly even more importantly they were acquired at pH 8.30 and at 21 (±2)°C. As pointed out above, pH affects the relative K_a values for Asp, Glu and Lys residues. We suspect that temperature also affects the relative values for many residues. However, only a subset of those data, shown in Figure 6, can be usefully compared. This subset consist of those inhibitor variants which differ from another member of the subset by a single substitution at P_1†. We wish to apply the additivity criterion (Laskowski *et al.*, 1983; Wells, 1990; Wynn, 1990; Horovitz *et al.*, 1991) to these data. Figure 6 is a listing of available additivity cycles. On the upper right of the left hand square, we have OMTKY3, our wild-type. Below it lies Montezuma quail ovomucoid third domain leucine form‡, OMMNQ3L. OMMNQ3L differs from OMTKY3 by changes of $R^{21} \rightharpoonup V^{21}$ and $L^{23} \rightharpoonup I^{23}$. Now note that moving to the left changes L^{18} (the reactive site P_1 residue) to S^{18} in both variants. In spite of the fact that K_a values for OMTKY3 and OMMNQ3L are different, $-\Delta\Delta G^o$ (L18S) values are the same within experimental error for all six enzymes. Another way to talk about $-\Delta\Delta G^o$

† This is of course, trivially true of all the variants listed in Table 1. We are now looking for additional variants with this property.

‡ Montezuma quail ovomucoid is polymorphic (Laskowski *et al.*, 1987). Some eggs contain only OMMNQ3L, others OMMNQ3S, yet others a 50:50 mixture (see Bogard *et al.*, 1980 for similar phenomenon in OMJPQ3). Third domain mixtures can be readily separated by reverse phase HPLC (Lin, 1986).

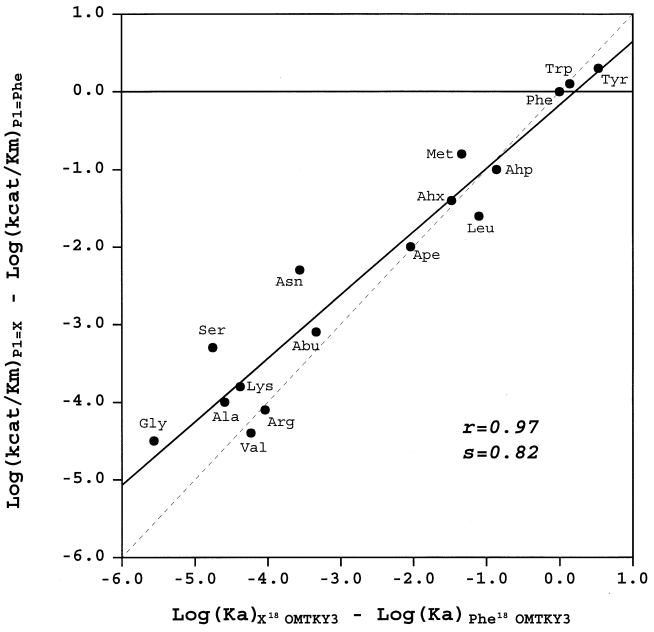

Figure 8. The dependence of $\log(k_{cat}/K_m)$ values of all chymotrypsin substrates in the world literature (Schellenberger *et al.*, 1991) upon log K_a values from Table 1. r is the Pearson product moment correlation coefficient, s *is* the power dependence (see the text). The thin diagonal is a slope 1.00 line drawn through the common Phe reference without any fitting.

(L18S) is its non-logarithmic version: the ratio of K_a values for the S^{18} and L^{18} variants. Therefore, identity of $K_a^{S(18)}/K_a^{L(18)}$ ratios is another measure of additivity. For the case discussed here the six ratios of ratios range from 1.1 to 1.8. We call such a ratio of ratios, F_c, the additivity factor. The determination of F_c requires an experimental determination of four equilibrium constants. Frequently, such constants differ by several orders of magnitude. An analysis of our data sets suggests that for $1.00 < F_c < 2.00$, the deviation from unity is likely to be caused by experimental errors. It follows that $|\Delta\Delta G^o|$ dev due to experimental errors is about 400 cal/mol. Figure 7 shows the F_c and $|\Delta\Delta G^o|$ dev values for the 65 cycles for which we have data. The reader can verify that there are 12 cycles depicted in Figure 6. This should yield six enzymes $\times$ 12 cycles $=$ 72 cycles. However, some data are missing and only 65 can be calculated. With few exceptions, the agreement is excellent.

We decided to examine the two most aberrant cases, where the F_c values are greater than 6. Both are for HLE, and both involve comparisons between P_1 Val and non-β-branched amino acid. HLE is the only one of the six enzymes we have studied that does not disfavor Val and Ile compared to Ape, Ahx and Leu. We have attributed this to exceptional flexibility of the HLE S_1 pocket. We think that the usual non-additivity we observe for HLE is due to the same phenomenon. Most S_1 pockets are rigid and less influenced by the nature of other contact residues, but HLE deforms easily.

Comparison with literature data

Unlike the data from our laboratory, the data of others were not obtained at the same pH and temperature as our own. Sometimes they were not even obtained for the same enzyme. We use subtilisin Carlsberg (CARL), others use BPN′ and savinase. These are closely related but not the same. The measuring methods also differ. Therefore, the differences, even in relative values, are likely to be and are large, and less strict comparisons than additivity must be employed.

The Pearson product-moment correlation (Campbell, 1989) which we already employed for the examination of enzyme specificity (equation (2)) is such a measure. Therefore, for each of the literature data sets we calculate such a correlation. The data are listed in Table 4. The correlation coefficients for many sets are quite near to unity.

All the plots that we construct are log/log plots. For example they are $-\Delta G_a^o$ for P_1 scaffold *versus* $-\Delta G_a^o$ of X^{18} OMTKY3 variants or:

$$\log \frac{k_{cat}}{K_m}$$

for a series of related substrates with only P_1 residue varied *versus* log K_a for X^{18} OMTKY3. The best straight line through such plots is characterized by a slope, s (see Figure 8). The value of s is simply the power dependence $y = ax^s$. In all of our plots, X^{18} OMTKY3 data are the x-axis or the abscissa, while the data of others are the ordinate or the y-axis. The $r = 1.00$, $s = 1.00$ indicate direct proportionality and therefore, additivity. However, s deviations from unity are far more common than r deviations. All of the plots on which Table 4 is based are in Lu (1994). Here only one of them (Figure 8) is given as an example.

Consider first the comparisons between free energies of transfer of amino acid side-chains from water to organic solvents to the binding to S_1 pockets. In fact, we restricted the comparisons to CHYM as only CHYM has a commodious enough pocket so that global steric effects do not matter. First, we look at the transfer of simple homologous, aliphatic side-chains to octanol and to chymotrypsin. The r value of 0.99 is superb but $s = 0.51$ not 1.00. This is a reflection of the well known phenomenon that ''enzymes are more hydrophobic than organic solvents'' (Fersht, 1977, 1985). (We feel that now it is better dealt with by the examination of the correlation $-\Delta\Delta G^o$ (Gly18X) to the buried hydrophobic surface area computed from X-ray crystallographic determinations and from modeling (Huang *et al.*, 1995). As we extend the size of the set to most of the residues, the correlation gets somewhat worse (it would have been worse yet if His, Ile and Pro were included) but the s value remains near one

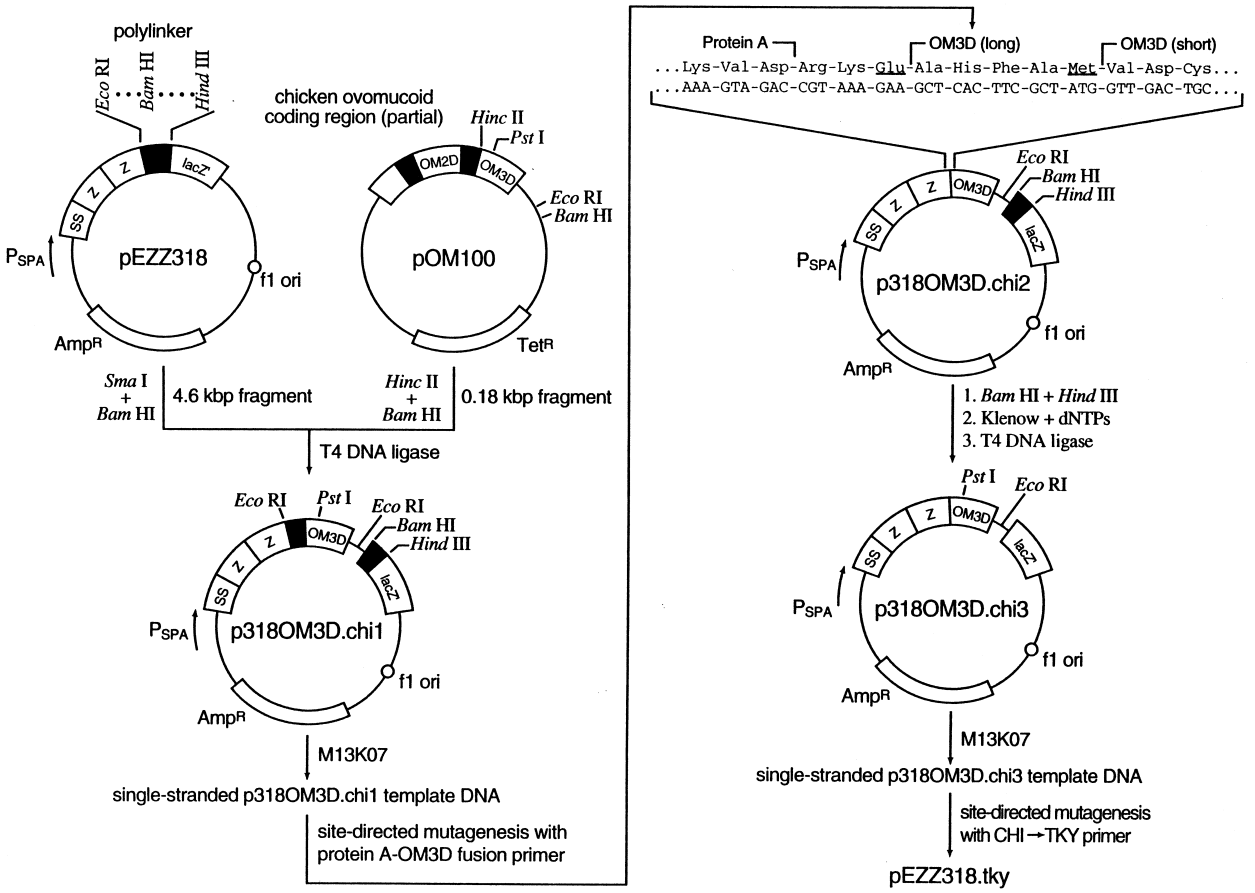

Figure 9. Schematic diagram of OM3D expression vector construction. The expression vector, pEZZ318.tky, was assembled starting from a chicken ovomucoid cDNA clone, pOM100, as described in Materials and Methods. The insert in pOM100 contains sequences representing the 3′-untranslated region and the coding region for amino acids 18 to 186 in the chicken ovomucoid mRNA (Stein *et al.*, 1978; Lai *et al.*, 1979); this insert thus contains the complete coding information for the OM2D and OM3D domains, but the sequence for the OM1D domain is incomplete. Black rectangles represent coding sequences for polypeptide linker regions connecting discrete protein domains. The Glu and Met residues in the short sequence (shown) connecting the ZZ and OM3D domains, which are the sites of cleavage in the fusion proteins by V8 porteinase or cyanogen bromide for producing the long and short forms of OM3D (see the text), respectively are underlined. Other symbols are as follows: P_{SPA}, staphylococcal protein A promoter; SS, signal sequence for secretion; Z, protein A IgG-binding domain; lacZ′, *E. coli* β-galactosidase α fragment; fl ori, fl phage origin of DNA replication; Amp^R, ampicillin resistance marker; Ter^R, tetracycline resistance marker.

half. Obviously, similar correlations could have been done for many different solvents and for more enzymes, not just CHYM. We did not persist as these two lines are in Table 4 only to illustrate our analysis technique.

The next part of Table 4 deals with interscaffolding comparisons between inhibitors which are not at all homologous (they are members of different families) but they share being standard mechanism canonical inhibitors. The *r* values in Table 4 are very near to unity, but *s* values deviate greatly. It is not clear whether this is due to differences in experimental conditions and techniques or the reflection of the true situation. In order to answer this, we recently acquired a set of seven eglin c variants at P_1 and measured them in our laboratory under the conditions of Table 1. Both *r* and *s* values are near unity.

The third section of Table 4 deals with $\log(k_{cat}/K_m)$ plots *versus* log K_a plots (Figure 8). Here Schellenberger *et al.* (1991) combined all of the known k_{cat}/K_m values from the world's literature and calculated their dependence on P_1 (and other residue positions). We found the agreement remarkably good. An even better agreement is found for a restricted set of acetyl methyl esters of simple hydrophobic amino acids reported in a classic paper by Dorovska *et al.* (1972). For other enzymes the *r* and *s* agreements are rather good but not as good as for CHYM. The excellent agreement overall between $\log(k_{cat}/K_m)$ and our log K_a presumably implies that the P_1 side-chains are imbedded in the enzyme's S_1 pockets in a very similar manner in the kinetic transition states and in the enzyme-inhibitor complexes. As three-dimensional structures of the latter are already available or are being ob-

tained, this is good news for kineticists, who can look at part of the transition states.

Materials and Methods

Construction of the OM3D expression vector

The steps in the construction of the expression vector for the turkey ovomucoid third domain as a downstream fusion to secreted IgG-binding domains of staphylococcal protein A are shown schematically in Figure 9. Briefly, a 0.18 kbp *Hinc*II/*Bam*H I fragment from a plasmid containing a molecular clone of the chicken ovomucoid cDNA, pOM100 (Stein *et al*., 1978; Lai *et al*., 1979), was purifed by gel electrophoresis and inserted into *Sma*I/*Bam*HI-digested pEZZ318 (Nilsson & Abrahmsén, 1990). This *Hinc*II/*Bam*HI fragment contained the coding region for the third domain of chicken ovomucoid (Figure 9). Single-stranded template was obtained from the above plasmid using the M13K07 helper phage (Vieira & Messing, 1987). The OM3D coding region was then fused in-frame to the coding sequences for the IgG-binding (''Z'') domains of staphylococcal protein A via deletion mutagenesis (Taylor *et al*., 1985; Nakamaye & Eckstein, 1986) with the oligonucleotide, 5′ GCA GTC AAC CAT AGC GAA GTG AGC TTC TTT ACG 3′ (protein A-OM3D fusion primer). To remove the rest of the polylinker sequences the plasmid was cleaved with *Hin*dIII and *Bam*HI, filled-in with Klenow DNA polymerase and dNTPs and religated. The chicken OM3D sequence was mutated to the turkey OM3D sequence and a unique *Bam*HI site introduced by site-directed mutagenesis of single-stranded template DNA using the oligonucleotide primer, 5′ GTT GTC *GGA* TCC ACA GAG AGG TCT GTA TTC CAG CGT GCA TGC AGG CTT AGG 3′ (CHI → TKY primer). Nucleotides that were altered in the mutagenesis from the chicken to the turkey sequence are underlined; the silent mutation in the codon for Gly^{25} that introduced the *Bam*HI site (italics) is also underlined. The amino acid and codon changes were as follows: D15A(GAC → GCA), A18L(GCA → CTG), and D20Y(GAC → TAC). The resultant expression plasmid was termed pEZZ318.tky.

Introduction of mutations at the P_1 position

Standard oligonucleotide-mediated site-directed mutagenesis, using the phosphorothioate incorporation method (Taylor *et al*., 1985; Nakamaye & Eckstein, 1986), was employed to introduce amino acid substitutions at the P_1 position in the recombinant turkey OM3D gene in pEZZ318.tky. Mutagenic oligonucleotides were also designed to revert the silent *Bam*HI site mutation in codon 25 (see above); since this *Bam*HI site was unique in pEZZ318.tky, mutant plasmids could be easily distinguished from the parental plasmid based upon their resistance to *Bam*HI digestion. Mutagenized plasmids were transformed into *Escherichia coli* strain XL-1 Blue (Stratagen) for selection, screening, and characterization. The entire OM3D coding region for each candidate mutant was resequenced by the dideoxy chain termination method (Sanger *et al*., 1977) to verify the presence of the intended mutation and the absence of adventitious second-site mutations. Each mutant plasmid was then retransformed into *E. coli* strain RV308 (Maurer *et al*., 1980) for protein expression.

Protein expression and purification

The fusion protein was designed to be secreted into the periplasmic space of *E. coli* to insure facile and correct formation of disulfide bridges (see Figure 1). Having the two IgG-binding Z domains also allowed us to use affinity chromatography on IgG column as a method of purification. In brief, after inoculation on an LB plate, a single colony of RV308 cells was grown overnight at 30°C in 75 ml of 2 × TY broth (16 g/l Tryptone, 10 g/l Yeast Extract, 5 g/l NaCl, 100 mg/l amplicillin and 0.1% (w/v) glucose). The medium then was transferred into a 2.8 liter culture flask containing 1.5 liters of 2 × TY broth and was continously shaken for 24 hours at 30°C. An osmotic shock procedure (Randall & Hardy, 1986) was used to obtain periplasmic proteins. After osmotic shock and centrifugation, the supernatant containing the periplasmic fraction was loaded onto an IgG-Sepharose 6 Fast Flow (Pharmacia) affinity chromatography column. The affinity purification was performed according to the standard procedures provided by the manufacturer (Pharmacia). The typical yield of fusion proteins is about 30 mg/l.

The fusion protein was subjected to CNBr cleavage at the Met^5–Val^6 bond in 5% TFA. The more common 70% formic acid was avoided as traces of formylation were seen in many purified products obtained in that medium. The reaction mixture was then subjected first to size exclusion chromatography on Bio-gel P-10 to separate third domain from the two Z domains of SPA followed by its purification by ion exchange chromatography on S-Sepharose and Q-Sepharose columns at pH 3.6 and 9.1, respectively.

In the cases of the Cys^{18} and Met^{18} variants, *Staphylococcus aureus* protease V8, instead of CNBr, was used to split off the connecting peptide at the Glu^{-1}–Ala^1 bond, which yielded third domains five residues longer at their N terminus. For the Cys^{18} variant, since there existed a significant amount of dimerized fusion protein covalently linked by Cys^{18}–Cys^{18} in the expression product, an alternative cleavage by CNBr was also performed to generate a third domain dimer. The cystine dimer was reduced by DTT into Cys^{18}OMTKY3 monomer without the N-terminal extra five residues. The yields of isolated highly purified X^{18}OMTKY3 variants were remarkably consistent (3.5 to 4.0 mg/l; nine liters were shaken at a time).

Characterization

Probably the most decisive characterization of the variants was that complexes of X^{18}OMTKY3 with SGPB were crystallized and their three-dimensional structures were determined (Huang *et al*., 1995; Huang, Bateman & James, personal communication). The conformation of the inhibitor in these complexes is strikingly similar to that found for the natural turkey ovomucoid third domain (Fujinaga *et al*., 1982; Read *et al*., 1983). These findings rule out a great number of possible folding concerns, such as improper closure of disulfide bridges. A possible objection is that only a part of the preparation, the part that crystallized, is being tested.

All of the samples were characterized by amino acid analysis, electrospray mass spectrometry, analytical anion exchange chromatography (Ardelt & Laskowski, 1982, 1991), thermal denaturation studies except for Cys^{18} and Met^{18}, and in many cases, N-terminal sequencing beyond the P_1 position. In each, the amino acid composition, the molecular weight, the chromatographic

behavior, the melting transition and t_m value as well as the N-terminal sequence (beyond the P_1 position) all agreed well with what was expected (Lu, 1994). It should be pointed out that, the Cys^{18} and Met^{18} variants generated by *Staphylococcus aureus*V8 hydrolysis contained trace amounts of non-covalently bound peptide fragments from the SPA domains, as indicated by N-terminal sequencing and poor amino acid analysis results. Therefore, the identification of these two variants largely relied on mass spectral analysis and protein sequencing, which agreed well with expectations.

The K_a values of OMTKY3 and OMSVP3 natural materials were respectively compared with the Leu^{18}OMTKY3 and Met^{18}OMTKY3 recombinant variants; in both cases, the K_a values are well within the ±20% experimental error (Lu, 1994). Note that in these comparisons for OMTKY3, the natural long form (Figure 1) is compared with the recombinant short form. For the OMSVP3, the natural long form is compared to the recombinant long form with a different N-terminal pentapeptide. The identity of the K_a values extends the earlier conclusions of Wieczorek *et al*. (1987), who showed that absence of the natural pentapeptide has no effect on K_a. We now show that the recombinant pentapeptide (Figure 1) also has no effect on K_a.

K_a determination

Determinations of equilibrium constants were carried out by an extensively modified procedure of Green & Work (1953), as described by Empie & Laskowski (1982) with later modifications by Park (1985) and Wynn (1990). The sources of the enzymes are given by Bigler *et al*. (1993). For most enzymes we studied, the procedure allows for the determination of K_a values over 10 orders of magnitude (K_a from about 10^3 to $10^{13}\ M^{-1}$) with an accuracy of ±20%. The data were collected on a Hewlett-Packard HP8450A diode array spectrophotometer at 21 (±2)°C at pH 8.30 in 0.1 M Tris-HCI containing 0.005% Triton X-100 and 0.02 M $CaCl_2$. Most K_a values were obtained using appropriate tetrapeptide *p*-nitroanilide chromogenic substrates for the determination of the residual enzyme concentration in the incubation mixture. However, for the Phe^{18}, Tyr^{18} and Trp^{18} with CHYM, a more sensitive 7-amido-4-methylcoumarin-based fluorogenic substrate was used and the signal change was monitored on a Perkin Elmer LS-50 fluorometer. In addition, K_a values for Cys^{18} were measured in the presence of 1 mM β-mercaptoethanol to prevent possible dimerization.

Acknowledgements

B. O'Malley's gift of pOM100 made the molecular biology much easier. We thank B. Nilsson for the gift of pEZZ318 and for the suggestion of RV308 as an expression host. Discussions with K. Huang, K. S. Bateman and M. N. G. James about their X-ray structures had a big impact on the interpretation. We are grateful to L. Smillie, J. Travis and the late M. Laskowski Sr for gifts of enzymes that are used as standards for validating the commercial ones. L. Smillie additionally reviewed his old data on SGPA sequence. D. A. Estell supplied unpublished k_{cat}/K_m values for CARL and BPN'. The former and present members of the Purdue ovomucoid third domain teams defined the problem, developed characterization and especially K_a measurement techniques and helped with the manuscript in many ways. The work was supported at Purdue by National Institutes of Health grant GM10831 and at Rutgers by National Institutes of Health grants AG11525 and AG10462.

References

Anfinsen, C. B. & Corley, L. G. (1969). An active variant of staphylococcal nuclease containing norleucine in place of methionine. *J. Biol. Chem.* **244**, 5149–5152.

Apostol, I., Komiyama, T., Yoshida, N. & Laskowski, M., Jr (1991). Glu specific *Streptomyces griseus* proteinase-Glu SGP- is homologous to SGPA, SGPB and α lytic proteinasePoster at 5th Symposium of the Protein Society.

Apostol, I., Giletto, A., Komiyama, T., Zhang, W. & Laskowski, M., Jr (1993). Amino acid sequences of ovomucoid third domains from 27 additional species of birds. *J. Protein Chem.* **12**, 419–433.

Ardelt, W. & Laskowski, M., Jr (1982). Analytical ion exchange chromatography of proteins. *Anal. Biochem.* **120**, 198–203.

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.

Bateman, K. S., Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr & James, M. N. G. (1996). X-ray crystal structures of SGPB in complex with three aromatic P_1 variants of OMTKY3. Poster presented at Keystone Symposia on Proteolytic Enzymes and Inhibitors in Biology and Medicine.

Bauer, C.-A. (1978). Active centers of *Streptomyces griseus* Protease 1, *Streptomyces griseus* Protease 3, and α-chymotrypsin: Enzyme-substrate interactions. *Biochemistry,* **17**, 375–380.

Beckmann, J., Mehlich, A., Schröder, W., Wenzel, H. R. & Tschesche, H. (1988). Preparation of chemically ''mutated'' aprotinin 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.

Bigler, T. (1991). Semisynthesis of ovomucoid third domain variants: structure-function studies. PhD thesis, Purdue University, West Lafayette, IN, USA.

Bigler, T., Lu, W., Park, S. J., Tashiro, M., Wieczorek, M., Wynn, R. & Laskowski, M., Jr (1993). Binding of amino acid side chains to preformed cavities: interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P_1 residues. *Protein Sci,* **2**, 786–799.

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

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., 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.

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.

Bogard, W. C., Jr, Kato, I. & Laskowski, M., Jr (1980). A Ser^{162}/Gly^{162} polymorphism in Japanese quail ovomucoid. *J. Biol. Chem.* **255**, 6569–6574.

Born, M. (1920). Volumen und hydralabonswärme der ionen. *Z. Phys.* **1**, 45–49.

Campbell, R. C. (1989). *Statistics for Biologists* Vol. 8, pp. 316–320, Cambridge University Press.

Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone receptor interface. *Science,* **267**, 383–386.

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.

Doig, A. J. & Sternberg, M. J. E. (1995). Side-chain conformational entropy in protein folding. *Protein Sci.* **4**, 2247–2251.

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 Letters,* **23**, 122–124.

Eisenberg, S. P., Hale, K. K., Heimdal, P. & Thompson, R. C. (1990). Location of the protease-inhibitory region of secretory leukocyte protease inhibitor. *J. Biol. Chem.* **265**, 7976–7981.

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.

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

Fersht, A. R. (1985). *Enzyme Structure and Mechanism* 2nd edit., W. H. Freeman and Company, Reading & Sons, San Francisco.

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**, 4868–4872.

Fujinaga, M., Delbaere, L. T. J., Brayer, G. D. & 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.* **183**, 479–502.

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 resolutions. *J. Mol. Biol.* **195**, 397–418.

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

Grøn, H., Meldal, M. & Breddam, K. (1992). Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. *Biochemistry,* **31**, 6011–6018.

Hansch, C. & Coats, E. (1970). α-Chymotrypsin: a case study of substituent constants and regression analysis in enzymic structure-activity relationships. *J. Pharm. Sci.* **59**, 731–743.

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 pancreative elastase, rat mast cell proteases I and II, bovine chymotrypsin A_α, and *Staphylococcus aureus* protease V-8 using tripeptide thiobenzyl ester substrates. *Biochemistry,* **23**, 2995–3002.

Henderson, G., Krygsman, P., Liu, C. J., Davey, C. C. & Malek, L. T. (1987). Characterization and structure of genes for proteases A and B from *Streptomyces griseus*. *J. Bacteriol,* **169**, 3778–3784.

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.

Huang, K. (1995). Structural studies of the interactions between serine proteinases and protein inhibitors. Ph.D. thesis, University of Alberta, Edmonton, Alberta, Canada.

Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr & James, M. N. G. (1995). Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu^{18}, Ala^{18}, and Gly^{18} variants of turkey ovomucoid inhibitor third domain with *Streptomyces griseus* proteinase B. *Protein Sci.* **4**, 1985–1997.

Janoff, A. & Basch, R. S. (1971). Further studies on elastase-like esterases in human leukocyte granules. *Proc. Soc. Exp. Biol. Med.* **136**, 1045.

Johnson, P. & Smillie, L. B. (1974). The amino acid sequence and predicted structure of *Streptomyces griseus* protease A. *FEBS Letters,* **47**, 1–6.

Jurasek, L., Carpenter, M. R., Smillie, L. B., Gertler, A., Levy, S. & Ericsson, H. L. (1974). Amino acid sequence of *Streptomyces griseus* protease B, a major component of pronase. *Biochem. Biophys Res. Commun.* **61**, 1095–1100.

Kato, I., Schrode, J., Wilson, K. A. & Laskowski, M., Jr (1976). Evolution of Proteinase Inhibitors. *XXIIIrd Annual Colloquium, Protides of the Biological Fluids,* pp. 235–243, Brugge, Belgium.

Kato, I., Kohr, W. J. & Laskowski, M., Jr (1978). Evolution of Avian Ovomucoids. *Regulatory Proteolytic Enzymes and Their Inhibitors* (Magnusson, S. *et al.*, ed.), vol. 47, pp. 197–201, FEBS 11th Meeting, Copenhagen.

Kato, I., Schrode, J., Kohr, J. & Laskowski, M., Jr (1987). Chicken ovomucoid: determination of its amino acid sequence, determination of the trypsin reactive site, and preparation of all three of its domains. *Biochemistry,* **26**, 193–201.

Kellis, J. T. Jr, Nyberg, K., Sali, D. & Fersht, A. R. (1988). Contribution of hydrohobic interactions to protein stability. *Nature,* **333**, 784–786.

Kellis, J. T. Jr, Nyberg, K. & Fersht, A. R. (1989). Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry,* **28**, 4914–4922.

Kojima, S., Nishiyama, Y., Kumajai, I. & Miura, K. (1991). Inhibition of subtilisin BPN′ by reactive 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 *S*treptomyces *g*riseus *P*roteinase (GluSGP). *J. Biol. Chem.* **266**, 10727–10730.

Lai, E. C., Stein, J. P., Catterall, J. F., Woo, S. L. C., Mace, M. L., Means, A. R. & O'Malley, B. W. (1979). Molecular structure and flanking nucleotide sequences of the natural chicken ovomucoid gene. *Cell,* **18**, 829–842.

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

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

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. *Proteinase Inhibitors: Medical and Biological Aspects* (Katunama, N., et al, ed.), pp. 55–68, Japan Scientific Society Press, Tokyo/Springer Verlag, Berlin.

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, Sequence 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. *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, 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 ovommucoid third domain from 25 additional species of birds. *J. Protein Chem.* **9**, 715–725.

Lee, B. (1993). Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of smaller size. *Protein Sci,* **2**, 733–738.

Levitt, M. (1976). a simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104**, 59–107.

Lin, T. Y. (1986). Detailed characterization of ovomucoid third domains and of their interaction with enzymes. Ph. D. thesis, Purdue University, West Lafayette, IN, U.S.A.

Lu, W. (1994). Energetics of the interactions of ovomucoid third domain variants with different serine proteinases. Ph.D. thesis, Purdue University, West Lafayette, IN, U.S.A.

Lu, W., Komiyama, T., Warne, N., Laskowski, M., Jr, Chaing, Y., Rothberg, I., Ryan, K. & Anderson, S. (1992). Expression of turkey ovomucoid third domain and of several variants with altered P_1 residues *Abstracts of the 6th Symposium of the Protein Society* Abstr. S191, pp. 79, The Protein Society, San Diego.

Lu, W., Zhang, W., Molloy, S. S., Thomas, G., Ryan, K., Chiang, Y., Anderson, S. & Laskowski, M., Jr. (1993a). Arg^{15}-Lys^{17}-Arg^{18} turkey ovomucoid third domain inhibits human furin. *J. Biol. Chem,* **268**, 14583–14585.

Lu, W., Zhang, W., Rothberg, I., Chaing, Y., Anderson, S. & Laskowski, M., Jr (1993b). Binding of amino acid side chains at P_1 of protein proteinase inhibitors to the preformed S_1 cavities in serine proteinases. *Protein Sci.* **2**(suppl.), 122.

March, C. J. (1980). Multiple domain Kazal inhibitors. PhD thesis, Purdue University, West Lafayette, IN, USA.

Matsumura, M., Becktel, W. J. & Matthews, B. W. (1988). Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile3. *Nature,* **334**, 406–410.

Matthews, B. W. (1993). Structural and genetic analysis of protein folding and stability. *Curr. Opin. Struct. Biol.* **3**, 589–593.

Maurer, R., Meyer, B. J. & Ptashne, M. (1980). Gene regulation at the right operator (OR) of bacteriophage γ, I. OR3 and autogenous negative control by repressor. *J. Mol. Biol.* **139**, 147–161.

Molina, M. A., Marino, C., Oliva, B., Avilés, F. X. & Querol, E. (1994). C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase A. *J. Biol. Chem,* **269**, 21467–21472.

Musil, D., Bode, W., Huber, R., Laskowski, M., Jr, Lin, T.-Y. & Ardelt, W. (1991). Refined X-ray crystal structures of the reactive site modified ovomucoid inhibitor third domains from silver pheasant (OMSVP3*) and from Japanese quail (OMJPQ3*). *J. Mol. Biol.* **220**, 739–755.

Nakamaye, K. L. & Eckstein, F. (1986). Inhibition of restriction endomuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis. *Nucl. Acids Res.* **14**, 9679–9698.

Nienaber, V., Breddam, K. & Birktoft, J. (1993). A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry,* **32**, 11469–11475.

Nilsson, B. & Abrahmsén, L. (1990). Fusions to staphylococcal protein A. *Methods Enzymol.* **185**, 144–161.

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

Park, S. J. (1985). Effect of amino acid replacements in ovomucoid third domains upon their association with serine proteinases. PhD thesis, Purdue University, West Lafayette, IN, USA.

Qasim, M. A., Ranjbar, M. R., Wynn, R., Anderson, S. & Laskowski, M., Jr (1995). Ionizable P_1 residues in serine proteinase inhibitors undergo large p*K* shifts on complex formation. *J. Biol. Chem.* **270**, 27419–27422.

Radzicka, A. & Wolfenden, R. (1988). Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry,* **27**, 1664–1670.

Randall, L. L. & Hardy, S. J. S. (1986). Correlation of competence for export with lack of tertiary structure of the mature species: a study in vivo of maltose-binding protein in *E. coli*. *Cell,* **46**, 921–928.

Read, R. J. & James, M. N. G. (1986). Introduction to the protein inhibitors: X-ray crystallography. *Proteinase Inhibitors* (Barrett, A. J. & Salvesen, G., eds), pp. 301–336, Elsevier Science Publishers B.V., Amsterdam.

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.

Rhodes, M. B., Bennet, N. & Feeney, R. E. (1960). The trypsin and chymotrypsin inhibitors from avian egg whites. *J. Biol. Chem.* **235**, 1686–1693.

Sandberg, W. S. & Terwilliger, T. C. (1991). Energetics of repacking a protein interior. *Proc. Natl Acad. Sci. USA,* **88**, 1706–1710.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA,* **74**, 5463–5467.

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

Schellenberger, V., Braune, K., Hofmann, H. & Jakubke, H. (1991). The specificity of chymotrypsin. *Eur. J. Biochem.* **199**, 623–636.

Serrano, L., Kellis, J. T., Jr, Cann, P., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. II. Substructure of Barnase and the contribution of different interactions of protein stability. *J. Mol. Biol.* **224**, 783–804.

Shortle, D., Stites, W. E. & Meeker, A. K. (1990). Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry,* **29**, 8033–8041.

Stein, J. P., Catterall, J. F., Woo, S. L. C., Means, A. R. & O'Malley, B. W. (1978). Molecular cloning of ovomucoid gene sequences from partially purified ovomucoid messenger RNA. *Biochemistry,* **17**, 5763–5772.

Tamura, A., Mitsushita, M., Naito, A., Kojima, S., Miura, K.-I. & Akasaka, K. (1996). Dynamics of the three methionyl side chains of *Streptomyces* subtilisin inhibitor. Deuterium NMR studies in solution and in the solid state. *Protein Sci.* **5**, 127–139.

Tanford, C. (1962). Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* **84**, 4240–4247.

Taylor, J. W., Schmidt, W., Cosstick, R., Okruszek, A. & Eckstein, F. (1985). The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA. *Nucl. Acids Res.* **13**, 8749–8764.

Thomson, J., Ratnaparkhi, G. S., Varadarajan, R., Sturtevant, J. M. & Richards, F. M. (1994). Thermodynamic and structural consequences of changing a sulfur atom to a methylene group in the M13Nle mutation in Ribonuclease-S. *Biochemistry,* **33**, 8587–8593.

Vieira, J. & Messing, J. (1987). Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**, 3–11.

Warne, N. W. (1990). Total chemical synthesis of turkey ovomucoid third domain and an HLE specific inhibitor. Ph.D. thesis, Purdue University, West Lafayette, IN, U.S.A.

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, IN, U.S.A.

Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K. & Makisumi, S. (1988). Purification and characterization of an acidic amino acid specific endopeptidase of *Streptomyces griseus* obtained from a commercial preparation (Pronase). *J. Biochem,* **104**, 451–456.

Edited by R. Huber

(Received 9 July 1996; received in revised form 7 November 1996; accepted 9 November 1996)