Designing Subtilisin BPN′ To Cleave Substrates Containing Dibasic Residues†

Marcus D. Ballinger,‡ Jeffrey Tom,§ and James A. Wells*,‡

Departments of Protein Engineering and Bioorganic Chemistry, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080

Received May 8, 1995; Revised Manuscript Received July 17, 1995⊗

ABSTRACT: The bacterial serine protease, subtilisin BPN′, has been mutated so that it will efficiently and selectively cleave substrates containing two consecutive basic (dibasic) residues. Mutants were designed on the basis of both the structure of subtilisin BPN′ and considerations of sequence differences between it and eukaryotic homologs, Kex2, PC2, and furin, which are known to cleave dibasic substrates. These eukaryotic proteases have high sequence homology to one another but differ substantially from subtilisin BPN′ in loops that interact with the substrate. When these loops were grafted into subtilisin BPN′, the mutated enzyme could not be expressed, presumably due to destabilization of the folded enzyme. We noted that several neutral residues in subtilisin BPN′ (Gly 166, Ser 33, and Asn 62) that are positioned to interact with a dibasic substrate are acidic residues at analogous positions in Kex2. Mutating these residues individually to either Glu or Asp in subtilisin BPN′ resulted in systematic shifts in substrate specificity (k_{cat}/K_m) toward basic residues and away from the natural preference for hydrophobic substrates. A combination mutant, where Asn 62 was changed to Asp and Gly 166 was changed to Asp (N62D/G166D), had a larger than additive shift in specificity toward dibasic substrates. This unexpectedly large change was confirmed by detailed analysis with a variety of synthetic substrates. Additional substrate determinants were revealed by sorting a library of phage particles (substrate phage) containing five contiguous randomized residues. This method identified a particularly good substrate (Asn-Leu-Met-Arg-Lys) that was selectively cleaved in the context of a fusion protein by the N62D/G166D subtilisin. This variant subtilisin may be useful for cleaving fusion proteins with dibasic substrate linkers and processing hormones or other proteins (*in vitro* or *in vivo*) that contain dibasic cleavage sites.

Site-specific proteolysis is one of the most common forms of posttranslational modifications of proteins [for a review see Neurath, 1989]. In addition, proteolysis of fusion proteins *in vitro* is an important research and commercial tool [for reviews see Uhlen and Moks (1990), Carter (1990), Nilsson *et al.* (1992), and Beck *et al.* (1994)]. Fusion domains can facilitate purification and augment expression and/or secretion. However, to liberate the protein product from the fusion domain requires selective and efficient cleavage of the fusion protein. Enzymatic methods are generally preferred as they tend to be more specific and can be performed under mild conditions that avoid denaturation or unwanted chemical side reactions.

A number of natural and designed enzymes have been employed for site-specific proteolysis. Although some are generally more useful than others (Forsberg *et al.*, 1992), no enzyme is applicable to every situation given the sequence requirements of the fusion protein junction and the possible existence of protease-sensitive sequences within the desired protein product. Thus, an expanded array of sequence-specific proteases, analogous to restriction endonucleases, would broaden the utility of site-specific proteolysis for processing fusion proteins or generating protein/peptide fragements either *in vitro* or *in vivo*.

A common site-specific proteolysis event is the maturation of prohormones by the Kex2 family of enzymes that is present in eukaryotic cells [for reviews see Steiner *et al.* (1992) and Smeekens (1993)]. This family of proteases, which includes the yeast Kex2 and the mammalian PC2 and furin enzymes, is homologous to the bacterial serine protease subtilisin. Bacterial subtilisins have broad substrate specificities [for review see Kraut (1977)] that reflect their roles as scavenger proteases. In contrast, the eukaryotic enzymes are very specific for cleaving at peptide sequences containing at least two, and in some cases three, basic residues; thus these enzymes are well suited for site-specific proteolysis.

Kex2 requires Arg at the P1 position of substrates and prefers Arg or Lys at P2. Furin and several other prohormone processing enzymes from higher eukaryotes have an additional requirement for Arg at P4 (Bresnahan *et al.*, 1990; Wise *et al.*, 1990; Hosaka *et al.*, 1991; Matthews *et al.*, 1994). Kex2 hydrolyzes dibasic substrates such as acetyl-Pro-Met-Tyr-Lys-Arg-MCA[1] with very high catalytic efficiency (k_{cat}/K_m = 1.1 × 10^7 M^{-1} s^{-1}; Brenner & Fuller, 1992). The specificities of Kex2-related subtilisins are

† This work was supported in part by an NRSA fellowship (M.D.B.) from NIH Postdoctoral Training Grant GM16549-01.

* To whom correspondence should be addressed.
‡ Department of Protein Engineering.
§ Department of Bioorganic Chemistry.

⊗ Abstract published in *Advance ACS Abstracts,* September 15, 1995.

[1] Abbreviations: subtilisin BPN′, subtilisin from *Bacillus amyloliquefaciens*; pNA, *p*-nitroanilide; MCA, methylcoumarin amide; hGH, human growth hormone; hGHbp, extracellular domain of the hGH receptor; PBS, phosphate−buffered saline; AP, alkaline phosphatase; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis. Mutants are designated by the single-letter amino acid code for the wild-type residue followed by their position and replacement residue. Multiple mutants are indicated by component single mutants separated by slashes. Substrates are described in triplet or single letter code as P*n*...P2-P1-P1′-P2′...P*n*′, where the scissile bond is between the P1 and P1′ residues (Schechter & Berger, 1967.

0006-2960/95/0434-13312$09.00/0

Table 1: Oligonucleotides Used for Site-Directed Mutagenesis, Specificity Pockets Affected by the Mutations, and Relative Expression of Activity for Subtilisin BPN′ Variants[a]

mutant	oligonucleotide	specificity pocket	activity expressed
S33D	5′-GCGGTTATCGAC**GA**CGGTATCGATTCT-3′	S2	+
S33K	5′-GCGGTTATCGACA**AA**GGTATCGATTCT-3′	S2	+
S33E	5′-GCGGTTATCGAC**GAA**GGTATCGATTCT-3′	S2	+
N62D	5′- CCAAGACAAC**G**ACTCTCACGGAA-3′	S2	+
N62S	5′-CCAAGACAACA**G**CTCTCACGGAA-3′	S2	+
N62K	5′-CCAAGACAACAAA**T**CTCACGGAA-3′	S2	+
S33D/N62D	see individual mutations	S2	+
G166D	5′-CACTTCCGGCAGCTC**GTCG**ACAGTGG**AC**TACCCTGGCAAATA-3′ (inserts *Sal*I site)	S1	+
G166E	5′-CACTTCCGGCAGCTC**GTCG**ACAGTGG**A**GTACCCTGGCAAATA-3′ (inserts *Sal*I site)	S1	+
G128P/P129A	5′- TTAACATGAGCCTCGGC**CCAGCTAGC**GGTTCTGCTGCTTTA-3′ (inserts *Nhe*I site)	S1	−
G128P/P129A/S130D/G131D	5′- TTAACATGAGCCTCGGC**CCCGCGGATGA**TTCTGCTGCTTTAAA-3′ (inserts *Sac*II site)	S1	−
T164N/V165D	5′-CGGCAGCTCAAGC**AACGAT**GGCTA**T**CCTGGC AAATACCCTTCTGTCA-3′ (inserts *Bsa*BI site)	S1	−
T164Y/V165D	5′-CGGCAGCTCAAGC**AACGAT**GGCTA**T**CCTGGC AAATACCCTTCTGTCA-3′ (inserts *Bsa*BI site)	S1	−
T164N-Y(insert)-V165D	5′-ACTTCCGGCAGCTC**TTCGAACTACGAC**GGGT ACCCTGGCAAATA-3′ (inserts *Bst*BI site)	S1	−
N62D/G166D	see individual mutations	S1/S2	+
N62D/G166E	see individual mutations	S1/S2	+

[a] Bold type indicates base changes from the pSS5 (wild-type) template. For activity expressed, (+) indicates a high level of skim milk digestion surrounding transformed colonies, and (−) indicates no activity detectable.

expected to arise from clustering of acidic residues in the side-chain binding sites (Siezen *et al.*, 1991). The impairment of protein processing accompanying mutations of selected Asp and Glu residues in furin supports this hypothesis (Creemers *et al.*, 1993).

The eukaryotic proteases are expressed in small amounts (Bravo *et al.*, 1994, Matthews *et al.*, 1994), which makes them impractical to apply presently to processing of fusion proteins *in vitro*. Given the fact that subtilisin BPN′ can be expressed in large amounts (Wells *et al.*, 1983), we wondered if it would be possible to alter its specificity to resemble that of Kex2, thus combining the advantages of both enzymes. Here we describe the production of a subtilisin variant with dibasic specificity that was generated using a combination of structure-based protein design and attention to the sequences of the prohormone processing enzymes. This engineered enzyme should be useful for processing fusion proteins or generating protein fragments by cleavage at designed dibasic sites.

MATERIALS AND METHODS

Materials. Oligonucleotides and *p*-nitroanilide substrates were synthesized and purified by the Bioorganic Chemistry Department at Genentech. Chromatography columns were from Pharmacia Biotechnology.

Construction and Purification of Subtilisin Mutants. Site-directed mutations were introduced into the subtilisin BPN′ gene cloned into the phagemid pSS5 (Carter & Wells, 1987) using the method of Kunkel *et al.* (1991). In the case of multiple S1 site variants [G128P/P129A, G128P/P129A/S130D/G131D, T164N/V165D, T164Y/V165D, and T164N-Y(insert)-V165D], mutations were introduced into a template containing the mutation S24C (Carter and Wells, 1987). For the S33D/N62D, N62D/G166D, and N62D/G166E double mutants, mutagenesis at positions 33 and 166 was performed on the N62D template. The oligonucleotides used for mutagenesis are listed in Table 1. In some cases, additional silent mutations were simultaneously introduced to provide a restriction site that could be used as additional verification of mutagenesis. All mutations were confirmed by dideoxy sequencing (Sanger *et al.*, 1977), and double-stranded mutant DNA was transformed into a protease-deficient strain (BG2036) of *Bacillus subtilis* (Yang *et al.*, 1984). The presence or absence of clear halos surrounding transformed colonies, indicative of skim milk digestion, was noted to give an indication of secreted protease activity.

Subtilisin variants were cultured and purified by ethanol precipitation as described (Carter & Wells, 1987), followed by chromatography over a Mono-S cation-exchange column using an FPLC system (Pharmacia). For the chromatography, variant subtilisins were loaded and washed in 5 mM $CaCl_2$ and 25 mM MES (pH 5.5) and eluted using a linear gradient from 0 to 0.15 M NaCl in the same buffer. Peak fractions of the subtilisin variants that could be expressed were judged to be >95% pure as assessed by SDS−PAGE. For the multiple S1 site variants, attempts were also made to coculture with A48E "helper" subtilisin. These mutant subtilisins were subjected to activated thiol affinity chromatography as described (Carter & Wells, 1987) in addition to the purification steps above.

Kinetic Characterizations. Subtilisins were assayed by measuring the initial rates of hydrolysis of pNA tetrapeptide substrates in 0.4 mL of 20 mM Tris-HCl (pH 8.2) and 4% (v/v) dimethyl sulfoxide at 25 ± 0.2 °C as described previously (Estell *et al.*, 1986). Kinetic measurements were made using a Kontron Uvikon 860 spectrophotometer. Enzyme concentrations were determined spectrophotometrically using $\epsilon_{280}^{0.1\%} = 1.17$ (Matsubara *et al.*, 1965) and were typically 2−20 nM for enzyme assays. Initial rates were determined for 9−12 different substrate concentrations over the range of 0.001−2.0 mM that bracketed K_m. Substrate concentrations were determined by allowing pNA hydrolysis

to proceed to completion and measuring the absorbance at 410 nm ($\epsilon_{410} = 8480\ cm^{-1}\ M^{-1}$). Plots of initial rates versus substrate concentration were fitted to the Michaelis−Menten equation using the program KaleidaGraph (Synergy Software, Reading, PA).

Substrate Phage. Substrate phage selections were performed as described by Matthews & Wells (1993), with minor modifications. Phage sorting was carried out using a library in which the linker sequence ($GPGGX_5GGPG$, where X is any of the naturally occurring L-amino acids) was inserted between the gene III coat protein and a tight-binding variant of hGH. The library contained 2×10^6 independent transformants. Phage particles were prepared by infecting 1 mL of log phase 27C7 F′tet *lacIq Escherchia coli* cells (a derivative of W3110−Δ*fhuA* Δ*ptr* Δ*phoA* Δ(*argF-lac*)169 Δ*ompT* Δ*htrA::kan* [*F′proAB lacIqZΔM15 Tn10*]; S. Bass, unpublished results) with $\sim 10^8$ library phage for 1 h at 37 °C, followed by 18−24 h of growth in 25 mL of 2YT medium containing 10^{10} M13K07 helper phage and 50 μg/mL carbenicillin at 37 °C. Wells of a 96-well microtiter plate (Nunc Maxisorb) were coated with 2 μg/mL hGHbp in 50 mM $NaHCO_3$ (pH 9.6) overnight at 4 °C and blocked with PBS (150 mM NaCl and 10 mM sodium phosphate at pH 7.4) containing 2.5% (w/v) skim milk for 1 h at room temperature. Between 10^{11} and 10^{12} phage in 0.1 mL of 100 mM NaCl, 1 mM EDTA, and 10 mM Tris-HCl (pH 7.6) were incubated in the wells at room temperature for 2 h with gentle agitation. The plate was washed with 20 rinses of PBS plus 0.05% Tween 20 and twice with 20 mM Tris-HCl (pH 8.2). The N62D/G166D subtilisin was added in 0.1 mL of 20 mM Tris-HCl (pH 8.2) and protease-sensitive phage were eluted after a variable reaction time. The concentration of protease and incubation times for elution of sensitive phage were decreased gradually over the course of the sorting procedure to increase selectivity, as follows: protease concentrations of 0.2 nM (rounds 1−3) and 0.1 nM (rounds 4−9) and reaction times of 5 min (rounds 1−6), 2.5 min (round 7), 40 s (round 8), and 20 s (round 9). Control wells in which no protease was added were also included in each round. For the resistant phage pool, the incubation time with protease remained constant at 5 min. The wells were then washed 10 times with PBS plus 0.05% Tween 20 and resistant phage eluted by treatment with 0.1 mL of 0.2 M glycine (pH 2.0) in PBS plus 0.05% Tween 20 for 1 min at room temperature. Protease-sensitive and -resistant phage pools were titered and used to infect log phase *E. coli* 27C7 F′tet *lacIq* cells for 1 h at 37 °C, followed by centrifugation at 4000 rpm, removal of supernatant, and resuspension in 1 mL of 2YT medium. The infected cells were then grown for 18−24 h in the presence of helper phage as described above, and the process was repeated 9 times.

Selected substrates were introduced into AP fusion proteins and expressed as previously described (Matthews & Wells, 1993; Matthews *et al.*, 1994), except that the PEG 3350 precipitation step in the fusion protein preparation was omitted due to poor yields. To assay for relative rates of cleavage, 10 μg of each hGH−AP fusion protein was added to 100 μL of hGHbp−affi-Gel resin slurry, and the mixture was stored at 4 °C overnight. The resin was washed three times with TE and twice in 20 mM Tris-HCl (pH 8.2), followed by treatment with 0.5 nM (final concentration) N62D/G166D subtilisin in 150 μL of 20 mM Tris-HCl (pH 8.2). Aliquots (25 μL) were withdrawn at various times and filtered through an Ultra-free MC 0.45 μm filter unit (Whatman). Twenty microliters of each aliquot was incubated with 180 μL of 6.6 mM *p*-nitrophenyl phosphate in 0.6 M Tris-HCl (pH 8.0), and the rate of *p*-nitrophenol release was monitored on an SLT microplate reader using Deltasoft II software. Cleavage rates were compared to a standard curve to quantitate AP release.

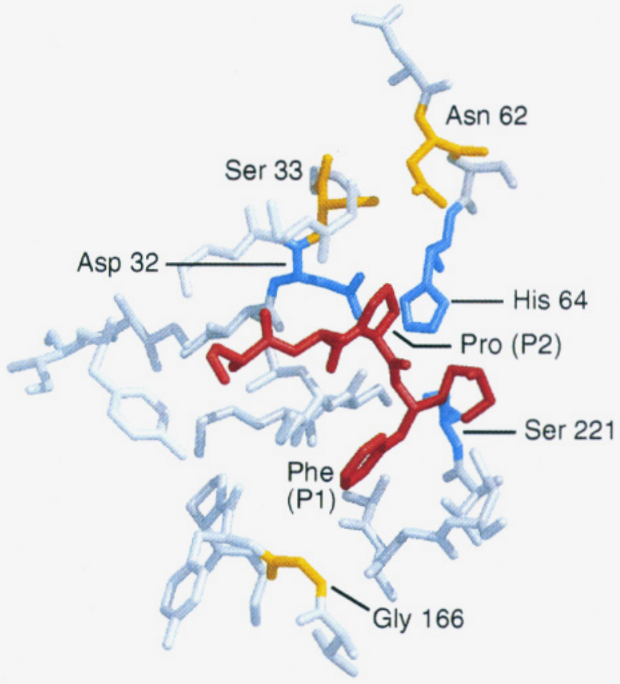

FIGURE 1: Structure of a succinyl-Ala-Ala-Pro-BoroPhe inhibitor (red) bound to the active site of subtilisin BPN′ (K. Butcher and T. Kossiakoff, unpublished results), showing residues in the S2 and S1 pockets subjected to mutagenesis (Ser 33, Asn 62, and Gly 166, in yellow) and the catalytic triad residues (Asp 32, His 64, and Ser 221, in blue).

Recovery and N-terminal sequencing of cleaved AP products were carried out as previously described (Matthews & Wells, 1993; Matthews *et al.* 1994).

RESULTS

Mutant Design and Preparation. A number of subtilisin structures have been solved with a variety of inhibitors and transition-state analogs bound (Wright *et al.*, 1969; McPhalen and James, 1988; Bode *et al.*, 1986; Bott *et al.*, 1988; Butcher & Kossiakoff, unpublished results). The structure of subtilisin BPN′ bound to a succinyl-Ala-Ala-Pro-Phe-boronate inhibitor (Figure 1) was used to locate residues that are in close proximity to side chains at the P1 and P2 positions of the substrate. Previous work had shown that replacement of residues at positions 156 and 166 in the S1 site with various charged residues led to improved specificity for complementary charged substrates (Wells *et al.*, 1987). Although longer range electrostatic effects on substrate specificity have been noted (Russell & Fersht, 1987), these were generally much smaller than local ones. Therefore, it seemed reasonable that local differences in charge between subtilisin BPN′ and the eukaryotic enzymes may account for the differences in specificity.

A detailed sequence alignment of 39 different subtilisins (Siezen *et al.*, 1991) allowed us to identify differences between subtilisin BPN′ and the eukaryotic processing enzymes, Kex2, furin, and PC2. Within the S1 pocket there

	S1 Site			S2 Site	
	125-131	151-157	163-167	30-35	60-64
Subtilisin	SLGGPSG	AAAGN**E**G	ST-VGY	VI**D**SGI	**D**NNSH
Kex2	SWGPA**DD**	FASGNGG	CNY**D**GY	IV**DD**GL	S**DD**YH
Furin	SWGP**EDD**	WASGNGG	CNC**D**GY	IL**DD**GI	N**D**NRH
PC2	SWGPA**DD**	WASG**D**GG	CNC**D**GY	IM**DD**GI	WFNSH

FIGURE 2: Sequence alignments for selected regions of subtilisin BPN′, Kex2, furin, and PC2 [from Siezen *et al.* (1991)], showing residues lining the S1 and S2 sites. Acidic residues are in boldface type.

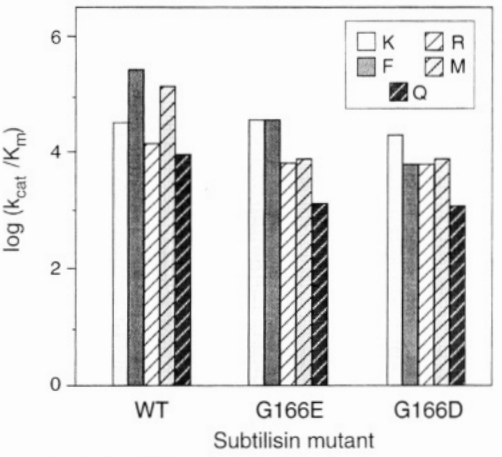

FIGURE 3: Catalytic efficiencies (k_{cat}/K_m) for S1 site mutants versus substrates having Arg, Lys, Phe, Met, or Gln at the P1 position. Data from Table 2.

are a number of charged residues that appear in the prohormone processing enzymes and not in subtilisin BPN′ (Figure 2). For example, the eukaryotic enzymes have two conserved Asp residues at positions130 and 131 as well as an Asp at position 165 that is preceded by insertion of a Tyr or Cys. However, in the region from residues 151−157, subtilisin BPN′ contains a Glu and the eukaryotes a conserved Gly. In the S2 site there were two notable differences in sequence. Subtilisin contains a Ser at position 33 whereas the prohormone processing enzymes contain Asp. There is not as clear a consensus in the region of residues 60−64, but one notable difference is at position 62. This side chain points directly at the P2 side chain (Figure 1) and is Asn in subtilisin BPN′, furin, and PC2 but Asp in Kex2. Thus, some but not all substitutions were suggestive of the specificity differences among these enzymes.

A variety of mutants were produced to probe and engineer the specificity of subtilisin BPN′ using oligonucleotides described in Table 1. After the mutant plasmids were produced, they were transformed into a protease-deficient strain of *B. subtilis* (BG2036) that lacks an endogenous gene for secretion of subtilisin. These were then tested for protease activity on skim milk plates.

The first set of mutants tested were ones where segments of the S1 binding site were replaced with sequences from Kex2. None of these segment replacements produced detectable activity on skim milk plates even though variants of subtilisin whose catalytic efficiencies are reduced by as much as 1000-fold do produce detectable halos (Wells *et al.*, 1986). Attempts to isolate these mutant subtilisins from shake-flask cultures were also unsuccessful. We went on to produce a series of single residue substitutions at the S1 and S2 sites that we anticipated to have less impact on the stability. These mutants at positions 166 in the S1 site and 33 and 62 in the S2 site were chosen on the basis of the modeling and sequence considerations described above. All single mutants as well as combination mutants that produced activity on skim milk plates were purified to near homogeneity.

Kinetic Analysis of Variant Subtilisins. To probe the effects of the G166E and G166D substitutions on specificity at the P1 position, we used substrates having the form succinyl-Ala-Ala-Pro-X-pNA, where X was either Lys, Arg, or Phe. The k_{cat}, K_m, and k_{cat}/K_m values were determined from initial rate measurements. Results are listed in Table 2 and graphically represented in Figure 3, along with data for Met and Gln P1 substrates from Wells *et al.* (1987). In terms of catalytic efficiency (k_{cat}/K_m), the wild-type enzyme preferred Phe > Met > Lys > Arg > Gln, whereas the G166E mutant preferred Lys ~ Phe > Arg ~ Met > Gln, and G166D preferred Lys > Phe ~ Arg ~ Met > Gln. Thus, both acidic substitutions at position 166 caused a shift towards a preference for basic residues as previously reported (Wells *et al.*, 1987). These shifts mainly reflect increases in K_m for uncharged side chains and decreases in K_m for basic P1 residues.

The effects of single and double substitutions in the S2 site were analyzed with substrates having the form succinyl-Ala-Ala-X-Phe-pNA, where X was either Lys, Arg, Ala, Pro, or Asp (Table 3; Figure 4). The wild-type enzyme preferred Ala > Pro > Lys > Arg > Asp at the P2 position, in agreement with previous data (Grøn *et al.*, 1992). In contrast, the S33D preferred Ala > Lys ~ Arg ~ Pro ≫ Asp, and the N62D preferred Lys > Ala > Arg > Pro ≫ Asp. Although the effects were most dramatic for the N62D mutant, the S33D variant also showed significant improvements toward basic P2 residues and corresponding reductions in hydrolysis of the Ala and Asp P2 substrates. We then

Table 2: Kinetic Analysis of S1 Site Mutants versus Substrates Having Variable P1 Residues[a]

	P1 residue														
	Lys[b]			Arg[b]			Phe[b]			Met[c]			Gln[c]		
mutant	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m
WT	32	950	3.3×10^4	8.9	630	1.4×10^4	29	110	2.6×10^5	13	91	1.4×10^5	3.3	370	8.9×10^3
G166D	12	630	1.9×10^4	2.7	500	5.5×10^3	6.1	1000	5.8×10^3	7.6	1200	6.5×10^3	2.8	2700	1.0×10^3
G166E	8.7	240	3.6×10^4	3.1	470	6.5×10^3	12	360	3.3×10^4	3.8	520	7.2×10^3	1.4	1200	1.1×10^3

[a] The kinetic constants were determined from plots of initial rates versus substrate concentration for the tetrapeptide series succinyl-Ala-Ala-Pro-X-pNa, where X was Lys, Arg, Phe, Met, or Gln. Units are as follows: k_{cat}, s^{-1}; K_m, μM; and k_{cat}/K_m, M^{-1} s^{-1}. Standard errors were less than 15%. [b] This work. [c] Data from Wells *et al.* (1987), in which assays were performed in 100 mM Tris-HCl (pH 8.6).

Table 3: Kinetic Analysis of S2 Site Mutants versus Substrates Having Variable P2 Residues[a]

	P2 residue														
	Lys			Arg			Ala			Pro			Asp		
mutant	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m
WT	13	79	1.6×10^5	5.0	62	8.1×10^4	8.3	3.5	2.4×10^6	29	110	2.6×10^5	3.5	94	3.7×10^4
S33D	26	98	2.7×10^5	9.2	36	2.6×10^5	45	60	7.4×10^5	52	260	2.0×10^5	2.7	590	4.6×10^3
N62D	26	26	1.0×10^6	3.3	44	7.5×10^4	17	130	1.4×10^5	22	440	4.9×10^4	1.5	1400	1.1×10^3
S33D/N62D	20	170	1.2×10^5	4.7	49	9.6×10^4	39	180	2.1×10^5	42	700	5.9×10^4	2.6	1000	2.6×10^3

[a] The kinetic constants were determined from plots of initial rates versus substrate concentration for the tetrapeptide series succinyl-Ala-Ala-X-Phe-pNa, where X was Lys, Arg, Ala, Pro, or Asp. Units are as follows: k_{cat}, s^{-1}; K_m, μM; and k_{cat}/K_m, M^{-1} s^{-1}. Standard errors were less than 15%.

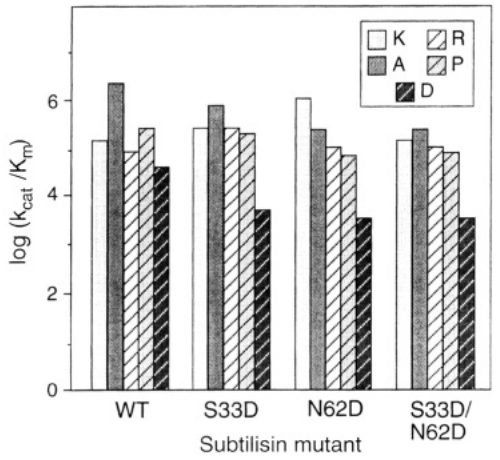

FIGURE 4: Catalytic efficiencies (k_{cat}/K_m) for S2 site mutants versus substrates having Arg, Lys, Ala, Pro, or Asp at the P2 position. Data from Table 3.

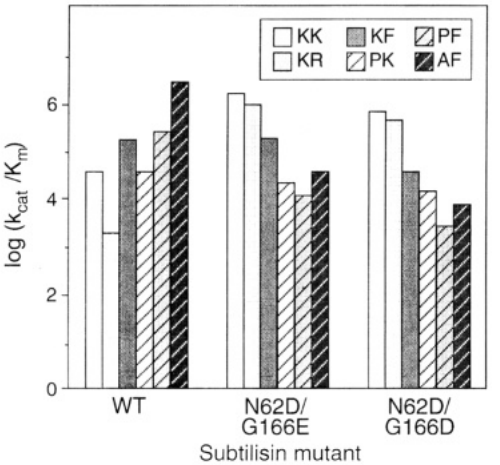

FIGURE 5: Catalytic efficiencies (k_{cat}/K_m) for S1 and S2 site double mutants versus substrates having Lys-Lys, Lys-Arg, Lys-Phe, Pro-Lys, Pro-Phe, or Ala-Phe at the P2-P1 positions. Data from Table 4.

analyzed the double mutant but found that it exhibited catalytic efficiencies nearly identical to the poorer of the two single mutants (S33D) for each of the substrates tested. Given the close proximity of positions 33 and 62, it is possible that charge repulsion could cause these less than additive effects [for a review see Wells (1990)].

Despite the less than additive effects seen for the two charged substitutions in the S2 site, we decided to combine the best S2 site variant (N62D) with the acidic substitutions in the S1 site. The two double mutants, N62D/G166E and N62D/G166D, were analyzed with substrates having the form succinyl-AAX_2X_1-pNA, where X_2X_1 was either KK, KR, KF, PK, PF, or AF (Table 4; Figure 5). The wild-type preference was AF > PF ~ KF > KK ~ PK. The KR substrate was not hydrolyzed to completion by the wild-type enzyme, possibly due to cleavage between K- and R-pNA. Both double mutants, however, had the preference KK > KR > KF > PK ~ AF > PF. Thus for the double mutants there was a dramatic improvement in the ability to cleave dibasic substrates and a decrease in activity toward hydrophobic substrates.

The greater than additive effect (or synergy) of these mutants can be seen from ratios of the catalytic efficiencies for the single and multiple mutants. For example, the G166E variant cannot distinguish Lys from Phe at the P1 position. Yet the N62D/G166E variant cleaves the Lys P2-Lys P1 substrate about 8 times faster than the Lys P2-Phe P1 substrate. Similarly, the G166D cleaves the Lys P1 substrate about 3 times faster than the Phe P1 substrate, but the N62D/G166D double mutant cleaves a Lys P2-Lys P1 substrate 18 times faster than a Lys P2-Phe P1 substrate. Thus, in contrast to the reduction in specificity seen for the double mutant in the S2 site, the position 62/166 double mutants in the S1 and S2 sites enhance specificity for basic residues as compared to component single mutants. The synergistic

Table 4: Kinetic Analysis of S1 plus S2 Site Double Mutants versus Substrates Having Variable P1 and P2 Residues[a]

	P2 residue																	
	Lys-Lys			Lys-Arg			Lys-Phe			Pro-Lys			Pro-Phe			Ala-Phe		
mutant	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m
WT	17	470	3.6×10^4	2.8	1700	1.7×10^3 [b]	13	79	1.6×10^5	32	950	3.3×10^4	29	110	2.6×10^5	8.3	3.5	2.4×10^6
N62D/G166D	11	21	5.3×10^5	15	41	3.7×10^5	3.8	126	3.0×10^4	11	1080	1.0×10^4	3.4	1800	1.9×10^3	5.1	1000	5.0×10^3
N62D/G166E	8.1	6.2	1.3×10^6	22	29	7.4×10^5	11	67	1.7×10^5	9.1	570	1.6×10^4	7.4	830	9.0×10^3	8.7	290	3.0×10^4

[a] The kinetic constants were determined from plots of initial rates versus substrate concentration for the tetrapeptide series succinyl-Ala-Ala-X_2-X_1-pNa, where X_2-X_1 was Lys-Lys, Lys-Arg, Pro-Lys, Lys-Phe, Ala-Phe, or Pro-Phe. Units are as follows: k_{cat}, s^{-1}; K_m, μM; and k_{cat}/K_m, M^{-1} s^{-1}. Standard errors were less than 15%. [b] Artificially low k_{cat}/K_m values are presumably due to competing cleavage between Lys and Arg.

effect on specificity arises through both an increase in catalytic efficiency for dibasic substrates and decreased catalytic efficiency for nondibasic substrates.

Attempts to reverse the polarity of designed ion pairs with respect to the enzyme and substrate were less successful. Previously, a G166K mutation had been shown to substantially increase the catalytic efficiency versus P1-Glu substrates relative to P1-Gln, Met, and Lys, although an overall preference for Met was maintained (Wells *et al.*, 1987). For the S2 site, an N62K mutant was analyzed. Similarly, this mutation lowered the catalytic efficiency versus an Ala P2 substrate relative to Asp, but the enzyme still preferred Ala at P2 (k_{cat}/K_m for succinyl-Ala-Ala-Ala-Phe-pNA = 3.3×10^4; k_{cat}/K_m for succinyl-Ala-Ala-Asp-Phe-pNA = 4.8×10^3). The data for this mutant and wild-type subtilisin indicate that Asp is an inherently poor P2 residue and thus a difficult target for a specificity design.

Substrate Phage Selection and Cleavage of a Fusion Protein. Subtilisin has the capability to bind substrates from the P4 to P3′ positions (McPhalen & James, 1988; Bode *et al.*, 1986). Given the extent and apparently cooperative nature of substrate binding, we wished to explore more broadly the substrate preferences for the enzyme. To do this we utilized the method of substrate phage selection (Matthews & Wells, 1993; Matthews *et al.*, 1994). In the technique, a five-residue substrate linker that is flanked by diglycine residues is inserted between an affinity domain (in this case a high-affinity variant of hGH) and the carboxy-terminal domain of gene III, a minor coat protein displayed on the surface of the filamentous phage, M13. The five-residue substrate linker is fully randomized to generate a library of 20^5 different protein sequence variants. These are displayed on the phage particles which are allowed to bind to hGHbp on a solid support. The protease of interest is added, and if it cleaves the phage particle at the substrate linker, that particle is released and recovered by elution. The particles released by protease treatment can be propagated and subjected to additional rounds of selection to further enrich for good protease substrates. Sequences that are retained can also be propagated to enrich for poor protease substrates. By sequencing the isolated phage genes at the end of either selection, one can identify the best and worst substrates for further analysis.

We chose to focus on the N62D/G166D variant as it was slightly better than N62D/G166E at discriminating in favor of synthetic dibasic substrates. We subjected the substrate phage library to nine rounds of selection with the subtilisin variant and isolated clones that were either increasingly sensitive or resistant to cleavage. Of 21 clones sequenced from the sensitive pool, 18 contained dibasic residues and 11 of them had the substrate linker sequence Asn-Leu-Met-Arg-Lys (Table 5). Three of the sensitive sequences were monobasic, Asn-Leu-Thr-Ala-Arg, reflecting that the wild-type preference for Ala at the P2 position was not entirely suppressed. Only three of ten sequences isolated from the protease-resistant pool contained basic residues, and those that were had a Pro following the mono- or dibasic residue. It is known that subtilisin does not cleave substrates containing Pro at the P1′ position (Carter *et al.*, 1989).

We wished to analyze how efficiently the most frequently selected sequences were cleaved in the context of a fusion protein. For this we applied an alkaline phosphatase–fusion protein assay (Matthews & Wells, 1993; Matthews *et al.*, 1994). The hGH substrate linker domains were excised from the phage vector by PCR and fused in front of the gene for *E. coli* AP. The fusion proteins were expressed, bound to an hGHbp affinity resin, and treated with the subtilisin variant. The rate of cleavage of the fusion protein from the plate was monitored by collecting soluble fractions as a function of time and assaying for AP activity (Figure 6). The most frequently isolated substrate sequence, Asn-Leu-Met-Arg-Lys, was cleaved about 10 times faster than the next frequently isolated clones (Thr-Ala-Ser-Arg-Arg and Asn-Leu-Thr-Ala-Arg). We also tested the dibasic sequence isolated from the resistant pool, namely, Arg-Lys-Pro-Thr-His. We did not detect cleavage above background for this substrate during the assay.

Table 5: Substrate Phage Sequences Sensitive or Resistant to Cleavage by N62D/G166D Subtilisin[a]

Protease-Sensitive Pool		
no basic sites (0)	monobasic sites (3)	dibasic sites (18)
	N L T A **R**(3)	N L M **R** **K**(11)
		T A S **R** **R**(4)
		L T **R** **R** S
		A L S **R** **K**
		L M L **R** **K**
Protease-Resistant Pool		
no basic sites (7)	monobasic sites (2)	dibasic sites (1)
A S T H F	Q **K** P N F	**R** **K** P T H
I Q Q Q Y	**R** P G A M	
Q G E L P		
A P D P T		
Q L L E H		
V N N N H		
A Q S N L		

[a] Sequences were derived from sorting a library containing a GPGG-XXXXX-GGPG linker, where X is any of the 20 natural L-amino acids, for nine rounds of selection. Numbers in parentheses indicate the number of times a particular sequence or type of sequence was isolated, and basic residues are shown in bold type.

The liberated AP products for the three substrate sequences were also subjected to N-terminal sequencing to determine the site of cleavage (Figure 6). As expected, the protease-sensitive fusions were cut following the dibasic or monobasic sites. The cleavage sites also illuminate an additional feature that is found in all the substrate sequences isolated, namely, a hydrophobic residue (Leu, Met, or Ala) in the P4 position. This is consistent with the S4 subsite preference of wild-type subtilisin BPN′ (Grøn *et al.*, 1992).

DISCUSSION

Previous attempts to remove or reverse charge specificity in enzyme active sites have generally been met with difficulty. A classic example is the attempt to recruit chymotrypsin-like (hydrophobic P1) specificity into the basic P1-specific protease trypsin [for a review, see Perona and Craik (1995)]. Mutation of the ion pair-forming Asp 189 to Ser in the S1 pocket was insufficient to confer the change in specificity (Graf *et al.*, 1988). In order to obtain

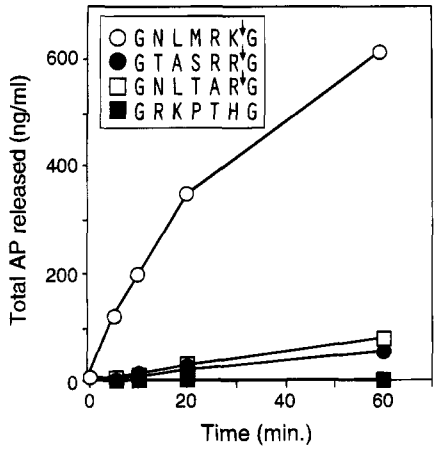

FIGURE 6: Results of hGH−AP fusion protein assay and N-terminal analysis of resulting AP cleavage products. hGH−AP fusion proteins were constructed, bound to hGHbp-coupled resin, and treated with 0.5 nM N62D/G166D subtilisin. Aliquots were withdrawn at various times, and AP release was monitored by activity assay in comparison to a standard curve as described (Matthews & Wells, 1993; Matthews *et al.*, 1994). Released AP products were also subjected to N-terminal sequencing to determine the site of cleavage (denoted by the arrow). The fusion protein containing the GRKPTHG linker did not release a detectable quantity of AP product.

chymotryptic-like specificity, it was necessary to transplant two more distant surface loops from chymotrypsin and install a Y172W mutation in addition to D189S (Hedstrom *et al.*, 1992, 1994). Moreover, charge specificity reversal was not gained by an Asp 189 to Lys mutation (Graf *et al.*, 1987). Similar results have been seen for analogous ion pair reversal experiments in aspartate aminotransferase (Cronin & Kirsch, 1988). These were attributed to a lack of stabilization of an introduced charged side chain or enzyme−substrate ion pair complex by the wild-type enzyme environment, in which strong specificity is attained not only by setting up a correctly positioned ion pair but also by generating an overall electrostatic environment that rejects noncharged or incorrectly charged substrates (Hwang & Warshel, 1988).

Attempts to introduce charge specificity into otherwise hydrophobic or nonspecific binding sites have been more successful (Wells *et al.*, 1987; Bocanegra *et al.*, 1993). Nonetheless, it can be difficult to set up an ionic interaction precisely enough for it to provide the desired specificity determinant. For example, Stennicke *et al.* (1994) made acidic (Asp or Glu) mutations at five residues in the P1′ binding site of carboxypeptidase Y in an attempt to change the P1′ preference from Phe to Lys or Arg. Only the L272D and L272E mutations were found to significantly alter the specificity in the desired direction, up to 1.5-fold preference in Lys or Arg over Phe. This indicates that the positioning of the charge was crucial for both forming the desired enzyme−substrate ion pair and discriminating against noncharged or oppositely charged substrates.

In the present work, we have reaffirmed that specificity enhancements toward basic substrates can be gained by introducing acidic residues into subtilisin BPN′ (Wells *et al.*, 1987). The effects are due to both improved catalytic efficiency for substrates having positively charged side chains and decreased catalytic efficiencies for other substrates. The discriminations are due to both k_{cat} and K_m effects, although the K_m effects predominate. The decreases in K_m for basic substrates and increases in K_m for nonbasic substrates suggest that the acidic residues can successfully form ion pairs with basic substrates and provide repulsion toward nonbasic substrates. We have found that the selectivity for basic side chains brought about by individual acidic mutations was modest. However, when two choice substitutions were simultaneously incorporated, the resulting variants had higher specificity for basic residues in each of the subsites. This apparent synergism is likely due to an interdependency of side-chain binding in the S1 and S2 subsites.

Nonadditive effects between subsites in subtilisin BPN′ have previously been reported (Grøn & Breddam, 1992), where it was shown that favorable interactions in one or more subsites (particularly S1 and S4) can mask the effects of poor interactions in others. Conversely, this would imply that detracting from important favorable interactions in one subsite should make substrate binding more dependent on the interactions in other sites. Thus by deterring hydrophobic P1 binding in subtilisin BPN′ by inserting the G166D mutation, the interactions with Asp 62 in the S2 pocket would become more influential to overall substrate binding. Alternatively, the nonadditivity could be explained by the peptide adopting a slightly different bound conformation when multiple ion pairs are formed. For example, a Asp 62−Lys P2 salt bridge in the S2 pocket of N62D/G166D subtilisin might be translated through the peptide and affect the binding of the P1 side chain. This could result in a slightly improved binding of basic residues and enhanced discrimination against others. In a wild-type S2, Pro P2 background, the presentation of the P1 side chain to Asp 166 might not be as optimal for discrimination. It should be emphasized that the nonadditivity seen in this case is slight, amounting to less than 1.1 kcal/mol in free energy; nonetheless, in terms of producing a useful site-specific protease the effects are extremely beneficial.

The interdependence of subsite interaction is also evident in the results of substrate phage sorting for N62D/G166D subtilisin. Although the kinetic data for tetrapeptide-pNA substrates imply that Lys is preferred at the P2 position for this variant, the dibasic substrates selected contained exclusively Arg at the P2 position. While this may have resulted from an expression bias favoring Arg (Matthews & Wells, 1993), it implies that the binding of full-length peptides results in an enhancement of specificity toward Arg at P2. In addition, the different amino acids selected at postions other than P1 and P2 relative to those fixed in the kinetic analyses of pNA substrates could influence the selectivity at the P1 and P2 sites.

The N62D/G166D subtilisin mutant resembles Kex2 and related enzymes with respect to the preference for dibasic P1-P2 substrates. The results provide strong evidence that the acidic substitutions examined are indeed important specificity determinants in Kex2. An important specificity difference between N62D/G166D subtilisin and the prohormone processing enzymes is that the latter has an absolute requirement for Arg in the P1 position (Brenner & Fuller, 1992), whereas N62D/G166D subtilisin accepts both basic amino acids with nearly equal catalytic efficiency. Furthermore, the specificity for arginine in the Kex2 enzymes arises through effects not only on K_m but also on k_{cat} (Brenner & Fuller, 1992). This suggests that binding of Arg in the S1 site helps to drive the enzyme−substrate complex to the tetrahedral intermediate. The differences throughout the loops comprising the P1 pocket relative to subtilisin are likely

to be involved in this intricate mechanism.

The combination of N62D and G166D mutations in subtilisin BPN′ produces an enzyme having enough specificity that it should be useful for cleavage of fusion proteins. Using substrate phage, we have identified a particularly good substrate sequence for this enzyme, Asn-Leu-Met-Arg-Lys. This sequence could be installed as a linker in a fusion of an affinity domain and the protein of interest to provide a method of expression and purification as demonstrated by the AP cleavage assay. This designed enzyme and matched substrate should make it possible to cleave other fusion proteins to generate protein products *in vitro* or *in vivo*.

ACKNOWLEDGMENT

We thank the oligonucleotide synthesis and protein sequencing groups for their support, Kerrie Andow for graphics, and Wayne Anstine for figures. We are also grateful to Thomas Chang and David Matthews for expert advice.

REFERENCES

Beck, J. T., Marsters, S. A., Harris, R. J., Carter, P., Ashkenazi, A., & Chamow, S. (1994) *Mol. Immunol. 31*, 1335−1344.

Bocanegra, J. A., Scrutton, N. S., & Perham, R. N. (1993) *Biochemistry 32*, 2737−2740.

Bode, W., Papamokos, E., Musil, D., Seemueller, U., & Fritz, M. (1986) *EMBO J. 5*, 813-818.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., & Power, S. (1988) *J. Biol. Chem. 263*, 7895−7906.

Bravo, D. B., Gleason, J. B., Sanchez, R. I., Roth, R. A., & Fuller, R. S. (1994) *J. Biol. Chem. 269*, 25830−25837.

Brenner, C., & Fuller, R. S. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 922−926.

Bresnahan, P. A., Leduc, R., Thomas, L., Thorner, J., Gibson, H. L., Brake, A. J., Barr, P. J., & Thomas, G. (1990) *J. Cell Biol. 111*, 2851.

Carter, P. (1990) in *Protein Purification: From Molecular Mechanisms to Large-Scale Processes* (Landisch, M. R., Wilson, R. C., Painton, C.-D., & Builder, S. E., Eds.) ACS Symposium Series 427, Chapter 13, pp 181−193, Americal Chemical Society, Washington, DC.

Carter, P., & Wells, J. A. (1987) *Science 237*, 394−399.

Carter, P., Nilsson, B., Burnier, J., Burdick, D., & Wells, J. A. (1989) *Proteins: Struct., Funct., Genet. 6*, 240−248.

Creemers, J. W. M., Siezen, R. J., Roebroek, A. J. M., Ayoubi, AT. A. Y., Huylebroeck, D., & Van de Ven, W. J. M. (1993) *J. Biol. Chem. 268*, 21826−21834.

Cronin, C. N., Kirsch, J. F. (1988) *Biochemistry 27*, 4572−4579.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., & Wells, J. A. (1986) *Science 233*, 659−663.

Forsberg, G., Baastrup, B., Rondahl, H., Holmgren, E., Pohl, G., Hartmanis, M., & Lake, M. (1992) *J. Protein Chem. 11*, 201−211.

Graf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J., & Rutter, W. J. (1987) *Biochemistry 26*, 2616−2623.

Graf, L., Jancso, A., Szilagyi, L., Hegyi, G., Pinter, K, Naray-Szabo, G., Hepp, J., Medzihradszky, K., & Rutter, W. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4961−4965.

Grøn, H., & Breddam, K. (1992) *Biochemistry 31*, 8967−8971.

Grøn, H., Meldal, M., & Breddam, K. (1992) *Biochemistry 31*, 6011−6018.

Hedstrom, L., Szilagyi, L., & Rutter, W. J. (1992) *Science 255*, 1249−1253.

Hedstrom, L., Perona, J. J., & Rutter, W. J. (1994) *Biochemistry 33*, 8757−8763.

Hosaka, M., Nagahama, M., Kim, W.-S., Watanabe, T., Hatsuzakawa, K., Ikemizu, J., Murakami, K., & Nakayama, K. (1991) *J. Biol. Chem. 266*, 12127−12130.

Hwang, J.-K., & Warshel, A. (1988) *Nature 334*, 270−272.

Kraut, J. (1977) *Annu. Rev. Biochem. 46*, 331−358.

Kunkel, T. A. , Bebenek, K., McClary, J. (1991) *Methods Enzymol. 204*, 125−139.

Matsubara, H., Kasper, C B., Brown, D. M., & Smith, E. L. (1965) *J. Biol. Chem. 240*, 1125−1130.

Matthews, D. J., & Wells, J. A. (1993) *Science 260*, 1113−1117.

Matthews, D. J., Goodman, L. J., Gorman, C. M., & Wells, J. A. (1994) *Protein Sci. 3*, 1197−1205.

McPhalen, C. A., & James, N. G. (1988) *Biochemistry 27*, 6582−6598.

Neurath, H. (1989) *Trends Biochem. Sci. 14*, 268.

Nilsson, B., Forsberg, G., Moks, T., Hartmanis, M., & Uhlén, M. (1992) *Curr. Opin. Struct. Biol. 2*, 569.

Perona, J. J., & Craik, C. S. (1995) *Protein Sci. 4*, 337−360.

Russell, A. J., & Fersht, A. R. (1987) *Nature 328*, 496−500.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.

Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.

Siezen, R. J., de Vos, W. M., Leunissen, A. M., & Dijkstra, B. W. (1991) *Prot. Eng. 4*, 719-737.

Smeekens, S. P. (1993) *Bio/Technology 11*, 182−186.

Steiner, D. F., Smeekens, S. P., Ohagi, S., & Chan, S. J. (1992) *J. Biol. Chem. 267*, 23435−23438.

Stennicke, H. R., Ujje, H. M., Christensen, U., Remington, S. J., & Breddam (1994) *Protein Eng. 7*, 911−916.

Uhlen, M., & Moks, T.(1990) *Methods Enzymol. 185*, 129−143.

Wells, J. A. (1990) *Biochemistry 29*, 8509−8516.

Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., & Chen, E. Y. (1983) *Nucleic Acids Res. 11*, 7911−7929.

Wells, J. A., Cunningham, B. C., Graycar, T. P., & Estell, D. A. (1986) *Philos. Trans. R. Soc. London, Ser. A 317*, 415−423.

Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P.and Estell, D. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1219−1223.

Wise, R. J., Baar, P. J., Wong, P. A., Kiefer, M. C., Brake, A. J., & Kaufman, R. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9378−9382.

Wright, C. S., Alden, R. A., & Kraut, J. (1969) *Nature 221*, 235−242.

Yang, M. Y., Ferrari, E., & Henner, D. J. (1984) *J. Bacteriol. 160*, 15−21.

BI951041O