PROTEINS: Structure, Function, and Genetics 6:240–248 (1989)

Engineering Subtilisin BPN′ for Site-Specific Proteolysis

Paul Carter,[1] Björn Nilsson,[2] John P. Burnier,[1] Daniel Burdick,[1] and James A. Wells[1]
[1]Department of Biomolecular Chemistry and [2]Department of Cardiovascular Research, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080

ABSTRACT **A combination of protein engineering and substrate optimization was used to create variants of the serine protease, subtilisin BPN′, which efficiently and specifically cleave a designed target sequence in a fusion protein. The broad substrate specificity of wild-type subtilisin BPN′ is greatly restricted by substitution of the catalytic histidine 64 with alanine (H64A) so that certain histidine-containing substrates are preferentially hydrolysed (Carter, P., Wells, J.A. Science 237:394–399, 1987). The catalytic efficiency, (k_{cat}/K_m), of this H64A variant was increased almost 20-fold by judicious choice of substrate and by installing three additional mutations which increase the activity of wild-type subtilisin. The most favorable substrate sequence identified was introduced as a linker in a fusion protein between a synthetic IgG binding domain of *Staphylococcus aureus* protein A and *Escherichia coli* alkaline phosphatase. The fusion protein (affinity purified on an IgG column) was cleaved by the prototype H64A enzyme and its improved variant, efficiently and exclusively at the target site, to liberate an alkaline phosphatase product of the expected size and N-terminal sequence. Several features of H64A variants of subtilisin make them attractive for site-specific proteolysis of fusion proteins: they have exquisite substrate specificity on the N-terminal side of the cleavage site and yet are broadly specific on the C-terminal side; they can be produced in large quantities and remain highly active even in the presence of detergents, reductants (modest concentrations), protease inhibitors, at high temperatures, or when specifically immobilized on a solid support.**

Key words: substrate-assisted catalysis, serine protease, fusion proteins, site-directed mutagenesis

INTRODUCTION

Site-specific proteolysis is a powerful and often essential tool for recovery of heterologous proteins expressed as larger fusion proteins (reviewed in ref. 1), for peptide mapping, and for analysis of protein structure or folding by dissection into functional domains[2] or separate folding units.[3] Proteolysis is preferable to chemical cleavage for recovery of functional proteins because chemical methods have limited specificities and usually require extreme conditions that can lead to unwanted side reactions and product heterogeneity.

Although a number of proteases have been used for site-specific proteolysis,[4–7] none appears to be ideally suited. The utility of these proteases is limited by their substrate specificities (leading to undesirable or incomplete cleavage products) and instabilities in detergents, reductants, or at high temperatures, which may be necessary conditions for solubilizing fusion proteins and making the target site accessible for hydrolysis. Furthermore, many of these proteases (especially mammalian blood-clotting enzymes[4]) are unavailable in large quantities and in highly purified forms so that they are free of other proteolytic activities. Subtilisin BPN′ most nearly satisfies all of these requirements for a site-specific protease. Although subtilisin has been used to obtain specific proteolytic fragments,[2,3] its substrate specificity is much too broad to be generally useful. Here, we have engineered subtilisin by rational design for site-specific proteolysis of an optimized target site in a fusion protein.

Extensive protein engineering studies of subtilisin BPN′ (reviewed in ref. 8) have identified several residues in the active site (Fig. 1) where amino acid substitutions lead to large changes in substrate specificity. Among these, the most dramatic narrowing of specificity occurred by replacing the catalytic histidine 64 with alanine (H64A) so that substrates providing the missing histidine side chain were preferentially hydrolyzed.[9] The H64A enzyme was also

Received April 10, 1989; accepted June 1, 1989.

Address reprint requests to Paul Carter or James A. Wells, Department of Biomolecular Chemistry, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080.

Björn Nilsson's present address is Department of Biochemistry, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

Abbreviations used: Me_2SO, dimethyl sulfoxide; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl fluoride; AP, *E. coli* alkaline phosphatase; Z, synthetic domain of *Staphylococcus aureus* protein A; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; *p*-nitroanilide (*p*na) substrates are represented: $sX_1X_2X_3X_4$-*p*na where s is a succinyl group and X_1 to X_4 are L-amino acids.

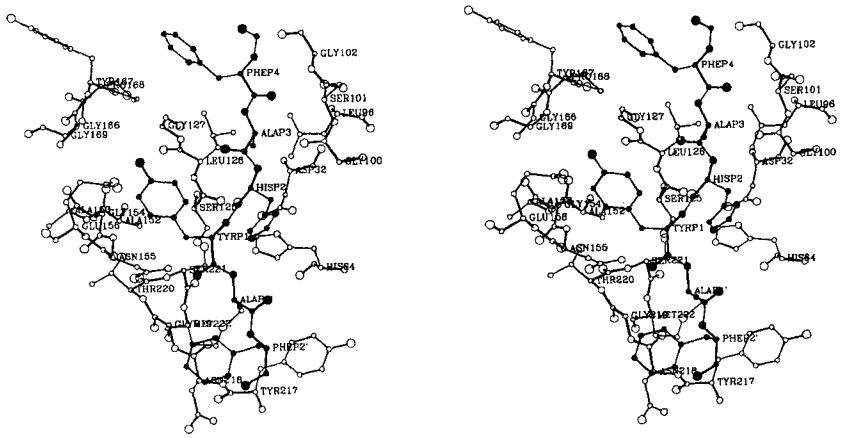

Fig. 1. Stereo view of a substrate model (filled atoms) L-Phe-L-Ala-L-His-L-Tyr-L-Ala-L-Phe bound to the active site of *Bacillus amyloliquefaciens* subtilisin BPN′ (open atoms). This model shows the superposition of the catalytic histidine (H64) with the substrate P2 histidine (taken from ref. 35). The substrate may be represented as:

$$NH_2\text{—Pn...P2—P1—}\underset{\displaystyle O}{\overset{\displaystyle H}{C\text{—}N}}\text{—P1′—P2′...Pn′—COOH}$$

where the scissile peptide bond is between the P1 and P1′ residues.[36]

attractive as a prototype site-specific protease because its specificity is directed toward histidine, which is the third least abundant residue in proteins.[10] Unfortunately, the catalytic efficiency of the H64A subtilisin towards the best His substrate originally tested (*N*-succinyl-L-Phe-L-Ala-L-His-L-Phe-*p*-nitroanilide; sFAHF-*p*na) was still 4000-fold below wild-type subtilisin BPN′, making the mutant enzyme virtually impractical as a processing enzyme.

One approach to increasing the catalytic efficiency of H64A subtilisin is to construct libraries of random mutants and to select (or screen) for mutants with higher activity. However, so far we have only achieved correct processing of the H64A variant in the presence of an active subtilisin BPN′,[9] which complicates the development of a selection scheme. As an alternative to random mutagenesis we have utilized the enormous kinetic data base of mutant and wild-type subtilisins (reviewed in ref. 8) to redesign the substrate binding site of H64A subtilisin BPN′. An attractive aspect of this approach is that mutants of the wild-type enzyme can be purified and assayed much more simply and rapidly than mutants of the H64A variant. Here we show that these mutations, combined with judicious choice of substrate, increase the catalytic efficiency of the H64A enzyme by almost 20-fold. We demonstrate that the prototype H64A enzyme and its most improved variant can be used for efficient site-specific proteolysis of a fusion protein at an optimized target linker.

MATERIALS AND METHODS

Materials

Oligodeoxyribonucleotides were synthesized by the Organic Chemistry Department at Genentech using hydrogen phosphonate chemistry[11] and purified by PAGE. All enzymes for DNA manipulations were from Bethesda Research Laboratories, except for T4 DNA ligase which was from New England Biolabs and *E. coli* DNA polymerase I large fragment (Klenow) which was from Boehringer Mannheim. Peptide and *p*-nitroanilide substrates were synthesized as described in ref. 12.

Construction, Expression, and Purification of Subtilisin BPN′ Mutants

The mutations S24C:H64A:E156S:G169A:Y217L (nomenclature described in Table I legend) in the cloned *Bacillus amyloliquefaciens* gene[13] were constructed by ligating three fragments: 0.75 kb *Eco*RI/*Pvu*I from S24C:H64A,[9] 0.75 kb *Pvu*I/*Bam*HI from E156S:G169A:Y217L,[14] and 6 kb *Eco*RI/*Bam*HI from pSS5 (B. Cunningham, D. Powers, and J.A.W., unpublished). The mutants S24C:H64A:G166A and S24C:H64A:E156S:G166A:G169A:Y217L were constructed by site-directed mutagenesis[15] using a 36-mer oligonucleotide (5′ GGTACCTCCGG-CTCGAGCAGCACAGTGGCCTACCCT 3′; *indicates mismatches) which also introduces a new *Xho*I site (underlined) to generate the G166A mutation using corresponding pSS5 templates. Mutant phagenids were verified by dideoxy sequencing[16] and then transformed into a protease deficient strain of *B. subtilis*, BG2036.[17] Mutant enzymes which contained H64 or H64A were expressed and purified as described in refs. 18 and 9, respectively. Purified enzymes were flash frozen in aliquots and stored at −70°C.

Kinetic Procedures

Mutant enzymes were assayed with substrates, *N*-succinyl-L-Phe-L-Ala-L-His-L-X-*p*-nitroanilide,

where X is either Phe or Tyr (sFAHF-*p*na and sFAHY-*p*na, respectively), against corresponding blanks in 1 ml 100 mM Tris-HCl at pH 8.60, 4% (v/v) dimethyl sulfoxide (Me_2SO) at (25 ± 0.2)°C with a Kontron Uvikon 860 spectrophotometer. Initial reaction rates were determined from the increase in absorbance at 410 nm on release of *p*-nitroaniline (ϵ_{410} = 8,480 M^{-1} cm^{-1}; ref. 19) and fitted to the Michaelis-Menten equation using a least-squares fit procedure.[20] Enzyme concentrations (determined spectrophotometrically; $\epsilon_{280}^{0.1\%}$ = 1.17, ref. 21) in the assays were 1–20 nM for H64 containing enzymes and 0.3–2 μM for H64A containing enzymes. The substrate concentrations were determined after total hydrolysis and corrected for background hydrolysis and were in the range of 0.1 K_m to 10 times K_m.

The S24C:H64A enzyme (1.9 μM) was assayed with the substrate sFAHF-*p*na (200 μM) as described above [except that the concentration of Me_2SO was 1% (v/v)] in the presence of varying amounts of KCl, NaCl, guanadine hydrochloride, urea, SDS, sodium deoxycholate, nonidet P-40, or Tween 20.

Digestion of Peptide Substrates

Peptide substrates (~0.6 mM) having the form *N*-succinyl-L-Phe-L-Ala-L-His-L-Tyr-L-[X]-L-Gly (where [X] represents the 20 common amino acids) were digested by S24C:H64A subtilisin (3.6 μM) in 1 ml 20 mM Tris-HCl at pH 8.0, 1 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 3.5% (v/v) Me_2SO, in the presence or absence of 2 M KCl at 37°C. (PMSF is added as a precaution because unlike wild-type subtilisin, or most other serine proteases, H64A containing variant enzymes are resistant to inhibition.[9]) At various times digests were applied to a C_{18} reverse-phase HPLC column (Waters) and eluted with a gradient of 0–40% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid. Elution of the substrate and *N*-succinyl-L-Phe-L-Ala-L-His-L-Tyr product fragment were monitored at 280 nm and quantified by peak integration and amino acid composition analysis. The relative (and absolute) cleavage rate for each peptide substrate was estimated from the initial rate of product formation in four to six successive time points.

Construction, Expression, Purification, and Digestion of Z-AP Fusion Protein

A phagemid pZAP encoding the Z-AP fusion protein was constructed by ligating three fragments: 4.4 kb *Hin*dIII (filled-in)–*Nar*I from protein A phagemid vector pEZ,[22] 1.4 kb *Not*I (filled-in)–*Mlu*I from the *E. coli* alkaline phosphatase (AP) gene[23] engineered with sites for *Mlu*I and *Not*I (3′ to coding sequence) (J. A. Wells, unpublished data), and a synthetic cassette coding for a histidine-containing linker with *Mlu*I and *Nar*I compatible ends. The ligation mixture was transformed into *E. coli* JM101 and plated on LB plates containing the chromogenic substrate for AP (5-bromo-4-chloro-3-indolyl phosphate; 2 μg/ml). Several AP expressing clones (blue colonies) were verified by dideoxy sequence analysis.[16]

The Z-AP fusion was expressed in *E. coli* JM101 containing pZAP, purified from osmotically shocked cells by binding to IgG sepharose, and eluted with lithium diiodosalicylate.[24] The purified Z-AP fusion protein was desalted by gel filtration (PD10 disposable columns, Pharmacia) and then dialyzed at 4°C overnight against 2 liters 50 mM Tris-HCl at pH 8.0. Aliquots were flash frozen and stored at −70°C. Samples of the Z-AP fusion protein that were digested by mutant subtilisins were precipitated with 10% (w/v) trichloroacetic acid and analyzed by SDS-PAGE.[25] The AP digestion product (M_r = 47,000) was electroblotted on polyvinylidene difluoride membrane[26] and the N-terminus was sequenced directly.

RESULTS

Enhancing the Catalytic Efficiency of Subtilisin

The mutations G166A[27] and E156S:G169A:Y217L[14] which enhance the activity of wild-type subtilisin BPN′ toward *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (sAAPF-*p*na) were examined first singley and then in combination using the substrate sFAHF-*p*na (Table I). The catalytic efficiency (k_{cat}/K_m) of wild-type subtilisin with sFAHF-*p*na was increased 2- and 12-fold by the mutations G166A and E156S:G169A:Y217L, respectively, and by 19-fold by combining these enhanced activity mutants.

Wild-type subtilisin hydrolyses the Tyr P1 substrate sAAPY-*p*na more efficiently than the homologous Phe P1 substrate (sAAPF-*p*na).[27] Accordingly, both the wild-type and the E156:G169A:Y217L enzymes showed increases in k_{cat}/K_m toward sFAHY-*p*na compared to sFAHF-*p*na. As expected, the G166A enzyme, which has been previously shown to sterically exclude a sAAPY-*p*na substrate,[27] is similarly reduced on the sFAHY-*p*na substrate. This was also found for the mutant E156S:G166A:G169A:Y217L.

The combined effects of these binding site mutants were then evaluated in the context of the H64A mutation with Phe P1 and Tyr P1 substrates (Table I). A surface accessible thiol (S24C) that has no effect upon enzyme activity[9] was introduced into all H64A variant subtilisins to facilitate their purification. Numerous control studies[9,20] have demonstrated that the purified active site mutants are free of detectable contaminating protease activity. The G166A variant or the combination of the three mutations (E156S:G169A:Y217L) improved k_{cat}/K_m by about 4-fold each. When all four mutations were in-

TABLE I. Kinetic Analysis of Mutant Subtilisins Against *N*-Succinyl-L-Phe-L-Ala-L-His-L-X-*p*-nitroanilide Where X is Phe or Tyr, at pH 8.60 and (25 ± 0.2)°C*

Enzyme	sFAHF-*p*na k_{cat} (s^{-1})	sFAHF-*p*na K_m (μM)	sFAHF-*p*na k_{cat}/K_m ($s^{-1}\ M^{-1}$)	sFAHY-*p*na k_{cat} (s^{-1})	sFAHY-*p*na K_m (μM)	sFAHY-*p*na k_{cat}/K_m ($s^{-1}\ M^{-1}$)	P1 substrate preference Tyr/Phe†
WT	4.4 ± 0.5	17 ± 2	$(2.6 \pm 0.2)\times10^5$	5.1 ± 0.1	6.6 ± 0.6	$(7.8 \pm 0.5)\times10^5$	3.0
G166A	2.3 ± 0.1	4.6 ± 0.5	$(5.1 \pm 0.4)\times10^5$	4.1 ± 0.1	26 ± 2	$(1.6 \pm 0.1)\times10^5$	0.31
E156S:G169A: Y217L	$(5.9\pm0.3)\times10^1$	20 ± 3	$(3.0 \pm 0.3)\times10^6$	$(2.8 \pm 0.1)\times10^1$	6.3 ± 0.5	$(4.4 \pm 0.2)\times10^6$	1.5
E156S:G166A: G169A: Y217L	$(3.8 \pm 0.1)\times10^1$	7.6 ± 0.8	$(5.0 \pm 0.4)\times10^6$	$(3.0 \pm 0.1)\times10^1$	41 ± 4	$(7.5 \pm 0.6)\times10^5$	0.15
S24C:H64A	$(2.1 \pm 0.1)\times10^{-2}$	340 ± 30	$(6.2 \pm 0.4)\times10^1$	$(9.9 \pm 0.2)\times10^{-2}$	210 ± 10	$(4.8 \pm 0.1)\times10^2$	7.7
S24C:H64A: G166A	$(3.8 \pm 0.1)\times10^{-2}$	150 ± 10	$(2.6 \pm 0.1)\times10^2$	$(1.5 \pm 0.1)\times10^{-2}$	340 ± 50	$(4.2 \pm 0.5)\times10^1$	0.16
S24C:H64A: E156S:G169A: Y217L	$(4.5 \pm 0.2)\times10^{-2}$	200 ± 20	$(2.2 \pm 0.2)\times10^2$	$(1.5 \pm 0.1)\times10^{-1}$	150 ± 10	$(1.0 \pm 0.1)\times10^3$	4.5
S24C:H64A: E156S:G166A: G169A:Y217L	$(8.3 \pm 0.1)\times10^{-2}$	120 ± 10	$(6.7 \pm 0.2)\times10^2$	$(7.9 \pm 0.2)\times10^{-3}$	270 ± 10	$(2.9 \pm 0.1)\times10^1$	0.043

*Mutants are designated by the single letter code for the wild-type amino acid followed by the residue number and then the amino acid replacement. Multiple mutants are identified by listing the single mutant components separated by colons (for example the double mutant Ser-24→Cys, His-64→Ala is designated S24C:H64A).
†Calculated from the ratio of k_{cat}/K_m terms for the Tyr P1 and Phe P1 substrates.

troduced into the S24C:H64A enzyme, k_{cat}/K_m was increased by more than 11-fold. The results with the Tyr P1 substrate toward the S24C:H64A family of enzymes also parallel those with the wild-type family; the S24C:H64A:E156S:G169A:Y217L variant was the most efficient catalyst followed by the S24C:H64A enzyme. Variants containing the G166A mutation were much worse due to steric hindrance for the Tyr P1 substrate as expected. Thus, the best enzyme–substrate pair within this family is the pentamutant (S24C:H64A:E156S:G169A:Y217L) subtilisin with the sFAHY-*p*na substrate; this enzyme is nearly 20-fold improved over the prototype pair (S24C:H64A hydrolyzing the sFAHF-*p*na substrate).

Survey of Conditions for Hydrolysis Assay for S24C:H64A Subtilisin

The prototype S24C:H64A enzyme was assayed under a variety of conditions to assess their use in digestion of protein substrates (Fig. 2). High concentrations of salt enhance the activity of the S24C:H64A enzyme (Fig. 2A), as has been shown for the wild-type enzyme.[28] The S24C:H64A variant retains 50% of its activity in the presence of 1 M urea or 0.5 M guanidine hydrochloride (Fig. 2B). The S24C:H64A variant is active in nonionic (Tween 20 and nonidet P-40) and ionic (SDS and sodium deoxycholate) detergents and at concentrations that are frequently used to solubilize and denature most proteins [e.g., 0.1% (w/v) SDS]. There was no detectable loss of activity during the kinetic runs (up to 30 minutes) under any of these conditions.

P1′ Specificity of S24C:H64A Subtilisin

Specificity determinants for subtilisin BPN′ extend for at least two residues on the C-terminal side of the scissile bond (P1′ and P2′, Fig. 1), which represent the first two residues of the protein of interest in a C-terminal fusion protein. The P1′ specificity of the prototype S24C:H64A enzyme was studied using the family of peptide substrates: *N*-succinyl-L-Phe-L-Ala-L-His-L-Tyr-L-[X]-L-Gly, where X represents the 20 common amino acids (Table II). The Phe-Ala-His-Tyr sequence was chosen from the most favorable *p*-nitroanilide substrate that we have identified for S24C:H64A subtilisin; glycine at P2′ was chosen to satisfy binding determinants between the main chain of the P2′ residue of the substrate and the enzyme. The relative rates of peptide cleavage were measured by the rate of formation of the product *N*-succinyl-L-Phe-L-Ala-L-His-L-Tyr. In every case hydrolysis occurred exclusively after the Tyr residue as expected. All of the P1′ substrates were hydrolyzed at rates within 7-fold of each other except for those containing Asp or Glu, which were cleaved slowly, or Pro or Ile, which were not cleaved at detectable rates. Hydrolysis of Asp or Glu P1′ substrates was stimulated 10-fold by addition of 2 M KCl and cleavage of other substrates tested was increased 1.5 to 3.5-fold. V_{max} for cleavage of sFAHY-*p*na (Table I) is roughly 5-fold greater than that for the most favorable P1′ peptide substrate, sFAHYRG (Table II). (The rate measured at ~0.6 mM sFAHYRG will be only slightly lower than that extrapolated to infinite substrate concentration.) The greater rate for the *p*-nitroanilide substrate com-

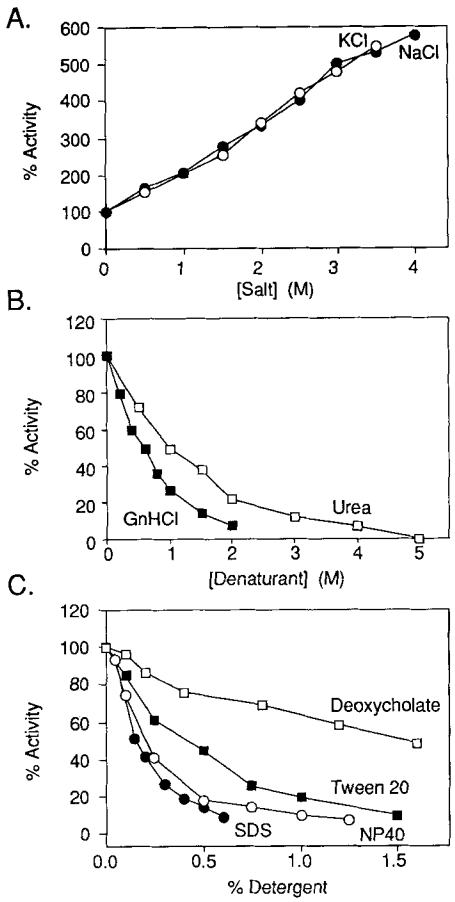

Fig. 2. Activity of S24C:H64A subtilisin BPN′ with 200 Ms-FAHF-*p*na at pH 8.60 and (25 ± 0.2)°C in the presence of **(A)** salts: KCl (○), NaCl (●); **(B)** denaturants: urea (□), guanidine hydrochloride (GuHCl, ■); **(C)** anionic detergents: SDS (●), sodium deoxycholate (□); nonionic detergents, nonidet P-40 (○) and Tween 20 (■).

pared to the hexapeptide may reflect the fact that the *p*-nitroanilide is a better leaving group than a dipeptide.

Specific Cleavage of a Model Fusion Protein

A model fusion protein was constructed (Fig. 3) that contains one synthetic (Z) domain of *Staphylococcus aureus* protein A,[22] followed by the optimized histidine-containing linker (Phe-Ala-His-Tyr) and *E. coli* alkaline phosphatase (AP). In an attempt to improve the accessibility of the site for cleavage, the target linker was preceded by the sequence Pro-Gly, where the glycine replaces a trypsin-sensitive lysine in protein A.[29] Furthermore, the N-terminus of AP is susceptible to proteolysis by both trypsin[30] (between Arg-11 and Ala-12) and by V8 protease[31] (between Glu-9 and Asn-10). AP was also an attractive marker protein because a similar fusion protein was expressed in high yield in the periplasmic space of *E. coli* with very little proteolytic degradation[24] and AP can be readily assayed using chromogenic substrates. The fusion protein was designed so that cleavage at the target site generates AP with an additional N-terminal Thr residue. This sequence

TABLE II. Digestion of Succinyl-L-Phe-L-Ala-L-His-L-Tyr-L-[X]-L-Gly (~0.6 mM) by S24C:H64A Subtilisin BPN′ (3.6 μM) at pH 8.0 and (37 ± 0.2)°C

P1′ residue	Relative cleavage rates	
	No KCl	+2 M KCl
Pro	<0.1	ND*
Ile	<0.1	<0.1
Asp	1	14
Glu	1	10
Met	6	ND
Phe	7	ND
Leu	7	ND
Gly	9	ND
Ser	10	19
Val	10	ND
Ala	17	ND
Tyr	21	27
Gln	23	81
Trp	24	ND
His	26	ND
Lys	30	45
Thr	35	ND
Asn	36	58
Cys	40	ND
Arg	43†	ND

*ND, not determined.
†This corresponds to an absolute rate of cleavage of 2 × 10^{-2} s^{-1}.

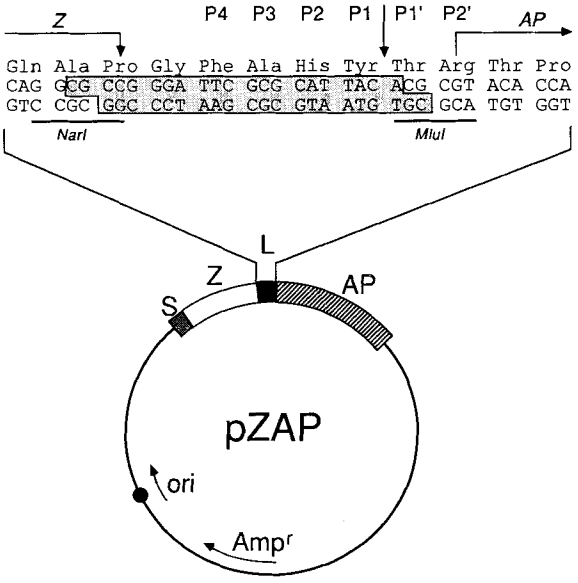

Fig. 3. Construction of phagemid pZAP encoding a fusion protein for the signal (S) and one synthetic domain for *Staphylococcus aureus* protein A (Z) followed by a histidine-containing linker (L) and then alkaline phosphatase (AP). The residues in the target site to be cleaved by the engineered subtilisin BPN′ variant are designated P4 through to P2′ (scissile bond indicated by large arrow).

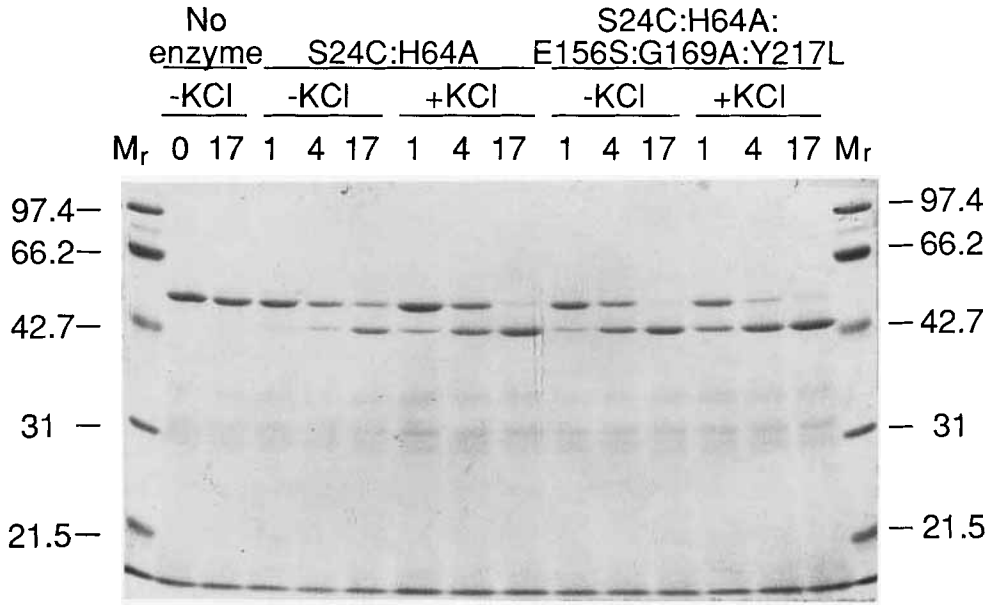

Fig. 4. Digestion of the Z-AP fusion protein with mutant subtilisins. Z-AP (200 pmol) was incubated without enzyme or with 10 pmol of either S24C:H64A or S24C:H64A:E156S:G169A:Y217L subtilisin variants in 100 μl 100 mM Tris-HCl at pH 8.60, 1 mM PMSF, 0.1% (v/v) Tween 20 in the presence (+) or absence (−) of 2 M KCl at 37°C for the times indicated (hr). Samples were analyzed by SDS-PAGE.[25] Molecular weight standards (M_r) have sizes as indicated ($\times 10^{-3}$).

(containing a *Mlu*I site) was particularly convenient to construct, and simplifies any subsequent manipulations to substitute the P1′ residue. The protein A derived fusion protein was efficiently purified by IgG affinity chromatography as previously described.[24]

Protease digestion experiments (Fig. 4) show that the prototype enzyme (S24C:H64A) and the most active pentamutant variant (S24C:H64A:E156S:G169A:Y217L) cleave the fusion protein (M_r = 54,000) efficiently and specifically to generate a protein with the expected electrophoretic mobility of AP ($M_r \cong$ 47,000). The protein A fragment ($M_r \cong$ 7,000) is too small to be resolved from the dye-front on this gel. N-terminal sequence analysis of the purified AP product in each case gave the sequence expected for cleavage at the designed target site (Thr-Arg-Thr-Pro-Glu-Met-Pro-). Digestion by the pentamutant was about 4-fold faster than by the prototype H64A variant, and in each case the cleavage rate was enhanced about 3-fold by 2 M KCl as was observed for many of the peptide substrates (Table II).

DISCUSSION

Enhancing the Activity of S24C:H64A Subtilisin

Mutations which increase the activity of wild-type subtilisin BPN′ were found to have approximately additive effects when combined.[14] Transferring these mutations to the S24C:H64A enzyme gave incremental improvements in activity, and so provided a useful strategy for enhancing catalytic efficiency. Furthermore the preference for Tyr over Phe at the P1 position for wild-type subtilisin was also found to hold for the S24C:H64A enzyme, suggesting that this is a useful approach for substrate optimization.

Although the results on wild-type subtilisin are qualitatively similar to those with the S24C:H64A enzyme, there are significant quantitative differences. For example, the G166A mutation gave a 4-fold improvement toward the Phe substrate in the S24C:H64A enzyme yet only about a 2-fold improvement in the wild-type enzyme. Conversely, the three mutations (E156S:G169A:Y217L) had a larger effect upon wild-type than upon the S24C:H64A enzyme. Furthermore, partitioning of the increase in catalytic efficiency into the k_{cat} and K_m terms were significantly different (Table I). These differences probably reflect subtle changes in substrate binding and/or catalytic mechanism between wild-type and the H64A variant enzymes.

Mutants which increase the activity of subtilisin BPN′ can mediate their effects in the absence of a functional catalytic triad. The four mutations E156S:G166A:G169A:Y217L together enhance the activity of the wild-type subtilisin with sFAHF-*p*na by 19-fold. These same mutations increase the activity of the S24C:H64A enzyme with His P2 (sFAHF-*p*na; Table I) and *des*-histidine (sAAPF-*p*na; P. C., unpublished results) substrates, by 11-fold and 10-fold, respectively. This suggests that these mutants act by directly stabilizing the transi-

TABLE III. Substrate Specificity of S24C:H64A Variant of Subtilisin BPN'

P4	P3	P2	P1	P1'	P2'	Substrate	Reference
				A. Efficiently Cleaved Substrates			
F	A	H	Y	*p*na			Table I
F	A	H	Y	[X]*	G		Table II
F	A	H	Y	T	R	Z-AP fusion protein	Figure 4
I	N	H	Y	R	M	Inhibition β-chain (residues 61–80)	9
F	A	H	F	*p*na			Table I, 9
M	E	H	F	R	W	ACTH (residues 1–10)	9
				B. Substrates not Detectably Cleaved†			
T	L	H	L	V	L	Ubiquitin (residues 62–76)	9
Y	E	H	F	E	N	BOP gene product (residues 68–86)	9
N	Q	H	L	C^{SO3H}	G	Bovine insulin B chain (oxidized)	9
G	S	H	L	V	E	Bovine insulin B chain (oxidized)	9
R	G	H	S	P	F	Inhibin β-chain (residues 61–80)	9

*[X] all common L-amino acids except Pro and Ile.
†Only small peptide substrates where the potential cleavage site is presumably accessible have been included.

tion state rather than by indirectly modulating the performance of the catalytic triad.

Substrate Specificity of S24C:H64A Subtilisin

Determination of the substrate specificity of H64A subtilisin variants is important in assessing their utility in site-specific proteolysis. Specificity determinants for subtilisin extend over at least six residues (Fig. 1); four on the N-terminal side of the scissile bond and two on the C-terminal side. Substrate specificity data for the prototype S24C:H64A variant are summarized in Table III. A histidine residue is apparently necessary but not sufficient for efficient polypeptide hydrolysis by the S24C:H64A enzyme. The P1 residue is also important in determining the efficiency of "substrate-assisted catalysis" by H64A variants, as shown by the qualitatively similar effects of mutants in the P1 pocket upon the activity of wild-type and S24C:H64A subtilisin toward Phe P1 and Tyr P1 substrates (Table I). Furthermore, all substrates identified to date that are efficiently cleaved by the S24C:H64A enzyme have Phe, Met, or Ile at the P4 position. This is consistent with the known preference of wild-type enzyme for hydrophobic residues at P4 and the hydrophobic nature of the P4 binding site.[32] For wild-type subtilisin (and probably for H64A variants) the side chain of the P3 residue is oriented away from the enzyme toward solvent, and consequently there is broad specificity at this subsite.

The S24C:H64A enzyme has very broad substrate specificity on the C-terminal side of the cleavage site (P1' and P2'). This is desirable since these residues would represent the first two residues of the protein of interest linked to the C-terminus of the fusion protein. All residues at P1' (apart form Ile and Pro) allow efficient substrate hydrolysis (Table II). From proteolysis of protein and synthetic peptide substrates, sequences containing at least Trp, Arg, Met, or Gly at P2' can be cleaved efficiently (Table IIIA). It is likely that the Pro P2' is unfavorable for H64A subtilisin as it is for the wild-type enzyme. X-ray crystallography shows that the main chain amide nitrogen and carbonyl oxygen of the P2' residue make hydrogen bonds with the main chain carbonyl oxygen and amide nitrogen of Asn-218, respectively.[32]

Naturally Occurring Cleavage Sites for S24C:H64A Subtilisin Are Rare

The major problem in achieving site-specific proteolysis is that digestion may not be limited to the designed target sequence. Even factor X_a, which has a very narrow substrate specificity, occasionally cleaves at other sites besides the Ile-Glu-Gly-Arg target sequence.[4] Digestion of the Z-AP fusion protein by S24C:H64A subtilisin and the pentamutant is restricted entirely to the target sequence, even though there are 12 other histidine residues present.[23] Seven of these histidines are surrounded by unfavorable P4, P1, or P1' residues and 4 others are unavailable because the histidine is coordinated to Zn^{2+} in native AP.[33] Aside from the target sequence, only His-87 in AP is in the context of a favorable amino acid sequence (Y T H^{87} Y A L). However this site, and all of the other histidine residues present in AP, are at least partially buried in the three-dimensional structure[33] making them unavailable for hydrolysis. In contrast to S24C:H64A and its variants, the wild-type enzyme rapidly cleaves the Z-AP fusion at two sites within four residues of the designed target followed by degradation of the AP product (not shown). This suggests that the region containing the target is highly accessible.

Assessment of the frequency of naturally occurring sites for S24C:H64A subtilisin rests upon attempting digestion of a large number of protein substrates. We were unable to cleave nine other

globular proteins (hen egg white lysozyme, horse cytochrome *c*, horseradish peroxidase, bovine pancreatic ribonuclease, spinach ferredoxin, bovine catalase, bovine serum albumin, human serum albumin, and human tissue-type plasminogen activator) which collectively contain more than 80 histidine sites, using the S24C:H64A enzyme under similar (native) conditions as described in Figure 4. In contrast, all of the proteins tested were digested at many sites by wild-type subtilisin. In some cases it may be necessary to denature the fusion protein in order to make the target site accessible for cleavage. Digestion of human serum albumin (contains 16 histidine residues) after reduction and carboxymethylation gave rise to limited proteolysis by the S24C:H64A enzyme (not shown) at a rate <100-fold that for cleavage of the Z-AP fusion protein.

Based upon the natural abundance of histidine in proteins (2%; ref. 10) and good P1 residues, Tyr, Phe, Leu, and Met (collectively about 22%), the frequency of good cleavage sites that only satisfy the P2 and P1 dominant sequence requirement is ~0.5%. Thus, although other histidines may be present in the product protein, very few are likely to contain satisfactory determinants at P4, P1, and P1′ as well as being accessible for hydrolysis by H64A variant subtilisins.

Nevertheless, if cleavage does occur at a significant rate at a site additional to the target site, it may be possible to overcome this by judicious choice of H64A variant and of target sequence. For example, if the offending site has Tyr P1, then one could use the S24C:H64A:E156S:G166A:G169A:Y217L variant in combination with a Phe P1 linker. This variant favors Phe over Tyr at P1 by 23-fold and retains high catalytic efficiency for Phe P1 (Table I).

Comparison of H64A Subtilisin Variants With an Ideal Site-Specific Protease

Several proteases have been used successfully for site-specific proteolysis, including factor X_a (ref. 4 and references cited therein), enteropeptidase,[5] collagenase,[6] and thrombin.[7] However, each of these enzymes is deficient by one or more of the criteria discussed below, and none has yet been improved by rational protein design. Specific cleavage of a peptide substrate using a catalytic antibody has recently been demonstrated by ingeniously inducing a metal binding site in the antibody.[34] However, the turnover number estimated for this antibody peptidase ($\sim 6 \times 10^{-4}$ s^{-1}) is 250-fold below that of the pentamutant subtilisin with sFAHY-*p*na, (Table I).

The S24C:H64A subtilisin variant satisfies most of the properties that we consider to be desirable for an ideal site-specific protease. It is exquisitely specific on the N-terminal side of the cleavage site and yet broadly specific on the C-terminal side to allow specific cleavage of a target linker. This enzyme can be recovered free of detectable protease contaminants in high yields (>30 mg/liter in shake flasks). It resists a variety of protease inhibitors (including PMSF, EDTA, leupeptin, and pepstatin) which permits their use during digestion to inactivate protease contaminants that may be present in fusion protein preparations (not shown). This enzyme, like wild-type subtilisin,[18] is fully active in reductants or 0.1% (w/v) SDS or 0.1% (v/v) Tween 20 and is moderately active in denaturants (20% activity is retained in 2 M urea and 10% activity in 2 M guanidine hydrochloride against sFAHF-*p*na, Fig. 2B) that may be required to solubilize the fusion protein or to make the target site accessible for cleavage. Activity of the pentamutant (S24C:H64A:E156S:G169A:Y217L) is enhanced about 4-fold compared with the prototype H64A enzyme with the Z-AP fusion protein containing the optimized target linker. The activity of each enzyme is enhanced ~3-fold in the presence of 2 M KCl and a further 2-fold by performing the digests at 50°C instead of at 37°C (data not shown). The pentamutant has been irreversibly immobilized on a solid support via the S24C residue and retains the ability to cleave the Z-AP fusion, albeit at rates several fold slower than for the free enzyme in solution (P. C., unpublished results). This type of protease column eliminates the need to purify the protease away from the cleaved products and facilitates recycling of the protease. A final advantage of a protease such as subtilisin BPN′, which is amenable to rational design, is that there is an extensive structural and functional data base[8] that can be utilized as need be for further modification of specificity determinants at the P4, P1, P1′, and P2′ binding sites. Even though the H64A family of subtilisin variants has many of the properties we consider important for site-specific proteolysis, assessment of their general usefulness rests upon testing a large number of different fusion proteins. Nonetheless, four different fusion proteins evaluated to date (including Z-AP) have been efficiently and specifically cleaved with the pentamutant subtilisin (P. C., unpublished results).

CONCLUSION

We have demonstrated the ability to make a number of small but incremental increases in the catalytic efficiency of a highly specific but weakly active protease by enzyme and substrate design. This strategy may prove useful in improving the catalytic efficiency of catalytic antibodies or other imperfect enzymes.

ACKNOWLEDGMENTS

We thank Byron Nevins for help with amino acid composition, Bill Henzel and Henry Rodriguez for N-terminal analysis, Mark Vasser, Parkash Jhurani, and Peter Ng for synthesizing oligonucleotides, Martin Struble for purification of peptide sub-

strates, Colin Watanabe for help with data base searching, Dr. Tony Kossiakoff for critical reading of the manuscript, and Carol Morita for preparing Figures 2–4.

NOTES ADDED IN THE PROOF

1. We have recently found that k_{cat}/K_m for the S24C:H64A:E156S:G169A:Y217L enzyme with the substrate sAAHY-*p*na ($(4.0 \pm 0.1) \times 10^3\ s^{-1}\ M^{-1}$ is 4-fold greater then for the substrate sFAHY-*p*na described in this study, and 65-fold greater than for the prototype enzyme (S24C:H64A) with the original substrate (sFAHF-*p*na). The sequence AAHY is therefore currently being tested as a target linker in fusion proteins.

2. In a few cases, slow cleavage has been detected with a histidine residue present at the P1′ (rather than P2) position of the substrate (P.C., J.A.W., R. Vandlen, K. Miller and S. Braxton, unpublished results). Modelling studies show that a P1′ histidine can occupy a relatively favorable hydrogen-bonding position with respect to the catalytic Ser221 (but not Asp32). Further studies are underway to investigate His P1′ assisted catalysis.

REFERENCES

1. Uhlén, M., Moks, T., Methods Enzymol., 1989 (in press).
2. Jacobsen, H., Klenow, H., Overgaard-Hansen, K. The N-terminal amino-acid sequences of DNA polymerase I from *Escherichia coli* and of the large and small fragments obtained by limited proteolysis. Eur. J. Biochem. 45:623–627, 1974.
3. Richards, F.M., Vithayathil, P.J. The preparation of subtilisin-modified ribonuclease and the separation of the peptide and protein components. J. Biol. Chem. 234:1459–1465, 1959.
4. Nagai, K., Thøgersen, H.C. Synthesis and sequence-specific proteolysis of hybrid proteins produced in *Escherichia coli*. Methods Enzymol. 153:461–481, 1987.
5. Craik, C.S., Largman, C., Fletcher, T., Roczniak, S., Barr, P.J., Fletterick, R., Rutter, W.J. Redesigning trypsin: Alteration of substrate specificity. Science 228:291–297, 1985.
6. Germino, J., Bastia, D. Rapid purification of a cloned gene product by genetic fusion and site-specific proteolysis. Proc. Natl. Acad. Sci. U.S.A. 81:4692–4696, 1984.
7. Reinach, F.C., Nagai, K., Kendrick-Jones, J. Site-directed mutagenesis of the regulatory light-chain Ca^{2+}/Mg^{2+} binding site and its role in hybrid myosins. Nature (London) 322:80–83, 1986.
8. Wells, J.A., Estell, D.A. Subtilisin-an enzyme designed to be engineered. Trends Biochem. Sci. 13:291–297, 1988.
9. Carter, P., Wells, J.A. Engineering enzyme specificity by "substrate-assisted catalysis." Science 237:394–399, 1987.
10. Klapper, M.H. The independent distribution of amino acid neighbor pairs in polypeptides. Biochem. Biophys. Res. Commun. 78:1018–1024, 1977.
11. Froehler, B.C., Ng, P.G., Matteucci, M.D. Synthesis of DNA via deoxynucleoside H-phosphonate intermediates. Nucleic Acids Res. 14:5399–5407, 1986.
12. Barany, G., Merrifield, R.B. Solid-phase peptide synthesis. In: The Peptides, Analysis, Synthesis, Biology. Special Methods in Peptide Synthesis, Part A. Vol. 2. E. Gross, J. Meienhofer (eds.). New York: Academic Press, 1979: 3–254.
13. Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A., Chen, E.Y. Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. Nucleic Acids Res. 11:7911–7925, 1983.
14. Wells, J.A., Cunningham, B.C., Graycar, T.P., Estell, D.A. Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering. Proc. Natl. Acad. Sci. U.S.A. 84:5167–5171, 1987.
15. Carter, P., Bedouelle, H., Winter, G. Improved oligonucleotide site-directed mutagenesis using M13 vectors. Nucleic Acids Res. 13:4431–4443, 1985.
16. Sanger, F., Nicklen, S., Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74:5463–5467, 1977.
17. Yang, M.Y., Ferrari, E., Henner, D.J. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an *in vitro*-derived deletion mutation. J. Bacteriol. 160:15–21, 1984.
18. Wells, J.A., Powers, D.B. *In vivo* formation and stability of engineered disulfide bonds in subtilisin. J. Biol. Chem. 261:6564–6570, 1986.
19. DelMar, E.G., Largman, C., Brodrick, J.W., Geokas, M.C. A sensitive new substrate for chymotrypsin. Anal. Biochem. 99:316–320, 1979.
20. Carter, P., Wells, J.A. Dissecting the catalytic triad of a serine protease. Nature (London) 332:564–568, 1988.
21. Matsubara, H., Kaspar, C.B., Brown, D.M., Smith, E.L. Subtilisin BPN′. I Physical properties and amino acid composition. J. Biol. Chem. 240:1125–1130, 1965.
22. Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T.A., Uhlén, M. A synthetic IgG-binding domain based on staphylococcal protein A. Protein Engineering 1:107–113, 1987.
23. Chang, C.N., Kuang, W.-J., Chen, E.Y. Nucleotide sequence of the alkaline phosphatase gene of *Escherichia coli*. Gene 44:121–125, 1986.
24. Nilsson, B., Abrahmsén, L., Uhlén, M. Immobilization and purification of enzymes with staphylococcal protein A gene fusion vectors. EMBO J. 4:1075–1080, 1985.
25. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685, 1970.
26. Matsudaira, P. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. 262:10035–10038, 1987.
27. Estell, D.A., Graycar, T.P., Miller, J.V., Powers, D.B., Burnier, J.P., Ng, P.G., Wells, J.A. Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering. Science 233:659–663, 1986.
28. Ottesen, M., Svendsen, I. The effects of salts on the enzymatic activity of subtilisin type Carlsberg and type Novo. Compt. Rend. Trav. Lab. Carlsberg 38:369–383, 1970.
29. Sjödahl, J. Structural studies on the four repetitive Fc-binding regions in protein A from *Staphylococcus aureus*. Eur. J. Biochem. 78:471–490, 1977.
30. Roberts, C.H., Chlebowski, J.F. Trypsin modification of *Escherichia coli* alkaline phosphatase. J. Biol. Chem. 259: 729–733, 1984.
31. Tyler-Cross, R., Roberts, C.H., Chlebowski, J.F. Proteolytic modification of *Escherichia coli* alkaline phosphatase. J. Biol. Chem. 264:4523–4528, 1989.
32. Robertus, J.D., Kraut, J., Alden, R.A., Birktoft, J.J. Subtilisin; a stereochemical mechanism involving transition-state stabilization. Biochemistry 11:4293–4303, 1972.
33. Sowadski, J.M., Handschumacher, M.D., Krishna Murthy, H.M., Foster, B.A., Wyckoff, H.W. Refined structure of alkaline phosphatase form *Escherichia coli* at 2.8 Å resolution. J. Mol. Biol. 186:417–433, 1985.
34. Iverson, B.L., Lerner, R.A. Sequence-specific peptide cleavage catalyzed by an antibody. Science 243:1184–1188, 1989.
35. Wells, J.A., Cunningham, B.C., Graycar, T.P., Estell, D.A., Carter, P. eds. On the evolution of specificity and catalysis in subtilisin. In: Evolution of Catalytic Function. Cold Spring Harbor Symposia on Quantitative Biology. Vol. LII. Cold Spring Harbor, New York: The Cold Spring Harbor Laboratory, 1987: 647–652.
36. Schechter, I., Berger, A. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27: 157–162, 1967.