Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 1219–1223, March 1987
Biochemistry

Designing substrate specificity by protein engineering of electrostatic interactions

(subtilisin/mutagenesis/ion-pair interactions)

JAMES A. WELLS[†‡], DAVID B. POWERS[†], RICHARD R. BOTT[†], THOMAS P. GRAYCAR[§], AND DAVID A. ESTELL[§]

[†]Department of Biocatalysis, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080; and [§]Research Department, Genencor, Inc., 180 Kimball Way, South San Francisco, CA 94080

Communicated by Frederic M. Richards, November 3, 1986 (received for review August 25, 1986)

ABSTRACT **Protein engineering of electrostatic interactions between charged substrates and complementary charged amino acids, at two different sites in the substrate binding cleft of the protease subtilisin BPN′, increases k_{cat}/K_m toward complementary charged substrates (up to 1900 times) and decreases k_{cat}/K_m toward similarly charged substrates. From kinetic analysis of 16 mutants of subtilisin and the wild type, the average free energies for enzyme–substrate ion-pair interactions at the two different sites are calculated to be -1.8 ± 0.5 and -2.3 ± 0.6 kcal/mol (1 cal = 4.18 J) [at 25°C in 0.1 M Tris·HCl (pH 8.6)]. The combined electrostatic effects are roughly additive. These studies demonstrate the feasibility for rational design of charged ligand binding sites in proteins by tailoring of electrostatic interactions.**

Protein engineering by *in vitro* mutagenesis of cloned genes (1) has been successfully applied to improve the thermal stability (2, 3) and oxidative stability (4–6) of various proteins and to produce enzymes with altered substrate specificities (7–10). Rational design of substrate specificity is especially significant because it can provide a general means for engineering biological macromolecules for useful chemical and biochemical reactions *in vitro* and *in vivo*. However, specificity is a challenging property to engineer predictably because it is determined by a complex and poorly understood mix of chemical binding forces that includes electrostatic interactions, steric, and hydrophobic effects (12). Enzyme–substrate interactions provide a useful means for isolating and evaluating chemical binding forces because the interactive surfaces can be modified both at the substrate by chemical synthesis and at the enzyme by protein engineering.

Although electrostatic interactions are a general binding force (13), they are difficult to predict quantitatively. Much of this uncertainty stems from the fact that the free energy, ΔG_e, is inversely proportional to the dielectric constant of the medium, which varies nonuniformly depending on shielding and charge effects. [For two complementary unit charges $\Delta G_e = -332/rD$, where the units for ΔG_e are in kcal/mol, r is in Å, and D is the dielectric constant (14).] Theoretical analyses reach different conclusions concerning the magnitude and constancy of the dielectric constant throughout a protein molecule (for reviews, see refs. 15–18). Furthermore, empirical estimates of specific electrostatic interactions are limited to a narrow data set provided by nature (18–22).

Here, electrostatic interactions are evaluated between charged peptide substrates and subtilisin, which is modified at two different sites using a cassette mutagenesis method (23) on the cloned subtilisin gene from *Bacillus amyloliquefaciens* (24). Large increases in k_{cat}/K_m toward charged substrates (up to 1900 times wild type) are produced by complementary charged mutations in the substrate binding cleft, and free energy values are determined for these electrostatic interactions.

MATERIALS AND METHODS

Materials. All restriction enzymes, T4 DNA kinase, and T4 DNA ligase were from New England Biolabs. DNA polymerase I large fragment (Klenow) was from Boehringer Mannheim. Oligodeoxyribonucleotides were provided by the Organic Chemistry Department (Genentech). Subtilisin substrates were synthesized by John Burnier (Genentech). *Escherichia coli* JM101 and MM294 were used for M13 infections and plasmid transformations, respectively.

Cassette Mutagenesis of the Subtilisin Gene. Position 156 and 166 single mutants, and position 156/166 double mutants were prepared by ligation of the three fragments. To produce fragment 1, a unique *Kpn* I site at codon 152 was introduced by site-directed mutagenesis (25) into the wild-type subtilisin sequence from pS4.5 (24) using a primer having the sequence 5′-TA-GTC-GTT-GC*G*-*GTA*-*CCC*-GGT-AAC-GAA-3′. (Asterisks indicate positions of mismatch with the wild-type sequence, and the mutant *Kpn* I site is italicized.) Enrichment for the mutant sequence was accomplished by restriction with *Kpn* I, purification of linear molecules, and self-ligation (26) to give pV152. To create a blunt end that terminated with codon 151, pV152 (≈1 μg) was digested with *Kpn* I and treated with 2 units of DNA polymerase I large fragment (Klenow fragment) plus 50 μM deoxynucleotide triphosphates (P-L Biochemicals) at 37°C for 30 min. The DNA was extracted with phenol/$CHCl_3$ (1:1, vol/vol), and DNA was precipitated with ethanol (28). DNA was digested with *Bam*HI, and the 4.6-kilobase piece (fragment 1) was purified by polyacrylamide gel electrophoresis (28).

Position 166 mutants were prepared by cassette mutagenesis (23) as described (7). Fragment 3, containing the carboxyl-terminal portion of the subtilisin gene including the desired position 166 codon, was isolated as a 610-base-pair *Sac* I/*Bam*HI fragment. Fragment 2 was a duplex synthetic DNA cassette that properly restored the coding sequence except at codon 156. The top strand of the cassette was synthesized to contain a glutamine codon, and the complementary bottom strand coded for serine at position 156. Ligation of heterophosphorylated cassettes leads to a large and favorable bias for the phosphorylated over the nonphosphorylated oligonucleotide sequence in the final segregated plasmid product (23). To obtain glutamine-156, the top strand was phosphorylated, annealed to the nonphosphorylated bottom strand (serine-156), and ligated with fragments 1 and 3. Mutant sequences that were isolated after ligation and transformation were confirmed by restriction analysis and DNA sequencing (27). The double 156/166 mutants were prepared by ligation of the 4.6-kilobase *Sac* I to *Bam*HI fragment from the relevant position 156 mutant plasmid, and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

[‡]To whom reprint requests should be addressed.

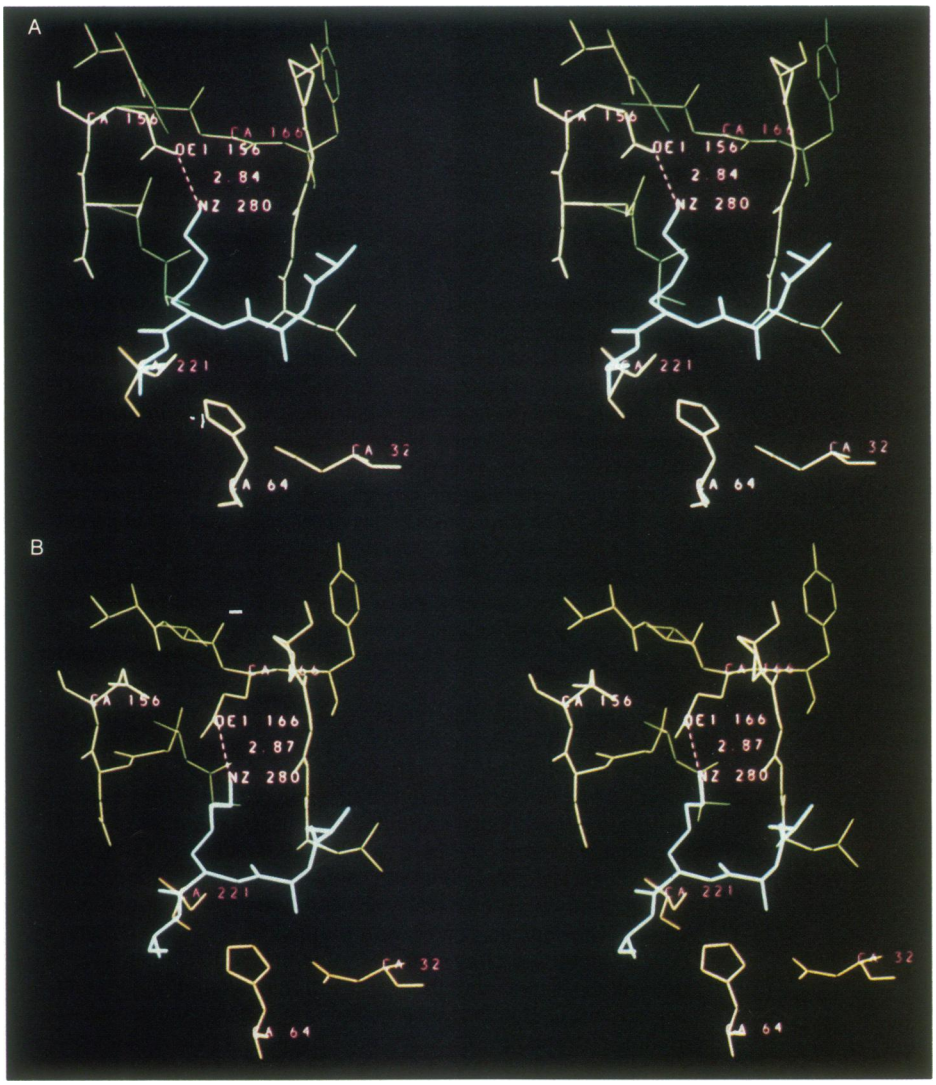

FIG. 1. Stereoview of the active site of subtilisin showing a lysine P1 substrate making potential ion pairs with the enzyme in two different ways. (*A*) A model ion pair (dashed line) of length 2.84 Å between a lysine P1 substrate (labeled NZ 280) and glutamate-156 (labeled OE1) in wild-type subtilisin. The catalytic triad of serine-221, histidine-64, and aspartate-32 is shown at the bottom (CA atoms labeled). The P1 binding cleft is composed of residues alanine-152 to glutamate-156 to the left (CA of glutamate-156 is labeled), methionine-124 to leucine-126 to the right, and threonine-164 to tyrosine-167 at the back (CA of glycine-166 is labeled). The substrate has the sequence L-Ala-L-Ala-L-Lys-L-Ala representing the P3-P2-P1 and P1′ residues, respectively. The substrate model is based on x-ray crystallographic data for a chloromethyl ketone peptide affinity analog having a P1 lysine residue bound to subtilisin BPN′ (32). The model was built from x-ray coordinates from a 1.8-Å structure (R.R.B., B. Katz, M. Ultsch, and T. Kossiakoff, unpublished results) and used an Evans and Sutherland PS300 graphics display. (*B*) A model ion pair of length 2.87 Å (dashed line) between the lysine P1 substrate (labeled NZ 280) and a glycine-166→glutamate mutant subtilisin (labeled OE1 166). The conformation of glutamate-156 has been slightly adjusted to avoid unfavorable steric interactions with glutamate-166.

the 0.6-kilobase *Sac* I to *Bam*HI fragment from the relevant 166 mutant plasmid. Expression and purification of *B. amyloliquefaciens* wild-type and mutant subtilisins to >95% homogeneity from *B. subtilis* culture supernatants were as described (4).

Kinetic Analysis. Purified subtilisins were assayed in 0.1 M Tris·HCl (pH 8.6) at 25°C as described (4, 7). Data from reaction progress curves were fit to the Michaelis–Menten equation by a nonlinear regression algorithm (30) and used to calculate the k_{cat}, K_m, and the product inhibition constant, K_i. In cases where product inhibition was observed, k_{cat} and K_m were also determined from initial rate data. Standard errors in k_{cat} and K_m for all values reported are below 10%.

RESULTS

Structural Considerations. X-ray crystal structures of subtilisin containing bound transition state analogues (31), peptide affinity labels (32, 33), and product inhibitors (34) show two possible modes of substrate binding depending on the P1 substrate side chain¶ (Fig. 1). For a phenylalanine P1 substrate, the side chain extends inward toward the α carbon of glycine-166, which is at the back of the P1 binding cleft. These structural findings have been corroborated by kinetic analysis of position 166 mutant proteins (7).

However, for a lysine P1 substrate, the side chain extends across the hydrophobic P1 binding cleft to form an ion pair with glutamate-156 at the entrance to the cleft (ref. 32; Fig. 1*A*). Model building shows that a glutamate side chain substituted at residue 166 can also form an ion pair with a lysine P1 substrate (Fig. 1*B*). Because side chains from either residue 156 or 166 have the potential to form an ion pair with a complementary charge at the P1 position of the substrate, we decided to make single- and double-mutant proteins and evaluate the effects on substrate specificity.

Changing Electrostatic Interactions Alters Substrate Specificity. From kinetic determination of the k_{cat}/K_m ratio (catalytic efficiency) one obtains the second-order rate constant for the conversion of a given substrate to product (12); differences in $\log(k_{cat}/K_m)$ provide an accurate measure of the lowering of the transition-state activation energy ($\Delta G_T^{\ddagger}$) (9, 12). To separate the contribution of electrostatics to substrate specificity from steric effects, mutant enzymes were analyzed with substrates that contained sterically similar side chains that differed in charge (e.g., glutamate versus glutamine, lysine versus methionine).

Mutations at position 156 and 166 produce changes in k_{cat}/K_m toward glutamate, glutamine, methionine, and lysine P1 substrates of up to 4000, 60, 200, and 80 times, respectively (Table 1). The wild-type enzyme is not the most catalytically efficient for any of the substrates tested, and two

¶Peptide substrate nomenclature can be represented as

$$NH_2\text{-Pn} \ldots \text{P2-P1-}\overset{\overset{O}{\|}}{C}\text{-}\overset{\overset{H}{|}}{N}\text{-P1}'\text{-P2}' \ldots \text{Pn}'\text{-COOH},$$

where the scissile peptide bond is between P1 and P1′.

Table 1. Kinetics of positions 156/166 subtilisins determined for different P1 substrates

Residue position		Net	P1 substrate, log k_{cat}/K_m (log $1/K_m$)			
156	166	charge	Glu	Gln	Met	Lys
Glu	Asp	−2	ND	3.02 (2.56)	3.81 (2.93)	4.21 (3.18)
Glu	Glu	−2	ND	3.06 (2.91)	3.86 (3.28)	4.48 (3.69)
Glu	Asn	−1	1.62 (2.22)	3.85 (3.14)	5.02 (3.97)	4.25 (3.07)
Glu	Gln	−1	1.20 (2.12)	4.36 (3.64)	5.54 (4.52)	4.10 (3.15)
Gln	Asp	−1	1.30 (1.79)	3.40 (3.08)	5.03 (3.98)	4.41 (3.22)
Ser	Asp	−1	1.23 (2.13)	3.41 (3.09)	4.67 (3.68)	4.24 (3.07)
Glu	Met	−1	1.20 (2.3)	3.89 (3.19)	5.64 (4.83)	4.70 (3.89)
Glu	Ala	−1	ND	4.34 (3.55)	5.65 (4.46)	4.90 (3.24)
Glu	Gly (wt)	−1	1.54 (2.29)	3.95 (3.43)	5.15 (4.04)	4.60 (3.13)
Gln	Gly	0	2.79 (2.98)	4.71 (4.17)	5.48 (4.32)	3.03 (2.40)
Ser	Gly	0	2.59 (2.92)	4.38 (3.79)	5.77 (4.73)	3.37 (2.70)
Gln	Asn	0	2.04 (2.72)	4.51 (3.76)	5.95 (4.86)	3.75 (2.74)
Ser	Asn	0	1.91 (2.78)	4.57 (3.82)	5.72 (4.64)	3.68 (2.80)
Glu	Arg	0	2.91 (3.30)	4.26 (3.50)	5.32 (4.22)	3.19 (3.06)
Glu	Lys	0	4.09 (4.25)	4.70 (3.88)	6.15 (4.45)	4.23 (2.93)
Gln	Lys	+1	4.82 (4.66)	4.64 (3.68)	5.97 (4.68)	3.23 (2.75)
Ser	Lys	+1	4.21 (4.40)	4.84 (3.94)	6.16 (4.90)	3.73 (2.84)
			Maximum difference: log k_{cat}/K_m (log $1/K_m$)			
			3.6 (2.9)	1.8 (1.4)	2.3 (2.0)	−1.9 (−1.5)

Codon numbers are defined from the start of the mature protease sequence (24). Net charge in the P1 binding site is defined as the sum of charges from positions 156 and 166 at pH 8.6. Values for k_{cat} (s^{-1}) and K_m (M) were measured in 0.1 M Tris·HCl (pH 8.6) at 25°C against P1 substrates having the form succinyl-L-Ala-L-Ala-L-Pro-L-(Xaa)-*p*-nitroanilide, where Xaa is the indicated P1 amino acid. Values for log $1/K_m$ are shown in parentheses. Because values for Glu-156/Asp-166 and Glu-156/Glu-166 with the glutamate P1 substrate are too small to determine accurately, the maximum difference is limited to a charge range of +1 to −1. ND, not determined; wt, wild type.

of the mutants (i.e., glutamate-156/alanine-166 and glutamate-156/methionine-166) are better than wild type on all the substrates shown. The changes in k_{cat}/K_m are dominated by changes in the $1/K_m$ term. Because $1/K_m$ is approximately equal to $1/K_s$, the enzyme–substrate association constant,[‖] the mutations primarily alter substrate binding (i.e., E + S→ E·S). This assumes that, like the wild type, the acylation step for the mutant enzymes is the rate-determining step.

Changes in substrate preference that arise from changes in the net charge in the P1 binding site show trends that are best attributed to electrostatic effects (Fig. 2). As the P1 binding cleft becomes more positively charged, the average catalytic efficiency increases much more for the glutamate P1 substrate than for its neutral and isosteric P1 homolog, glutamine (Fig. 2*A*). In contrast, as the P1 site becomes more positively charged, the catalytic efficiency toward the lysine P1 substrate decreases and diverges from its neutral and steric homolog, methionine (Fig. 2*B*). The parallel upward trend seen with increasing positive charge for both neutral P1 substrates, methionine and glutamine, may result in part from the fact that all the substrates are succinylated on their amino-terminal end and thus carry at least one negative charge.

The effect of changing the P1 binding site charge on substrate preference can be estimated from the differences between the slopes for the charged and neutral isosteric P1 substrates (Fig. 2). The increase in log(k_{cat}/K_m) for charged compared to neutral isosteric substrates is additive, and roughly a factor of 10 in k_{cat}/K_m for each complementary charge present in the P1 binding site on the enzyme. When inverting the charge on the substrate (i.e., P1 lysine to glutamate) the k_{cat}/K_m increases 100 times per complementary enzyme charge.

Ion Pair Interaction at Residues 156 and 166. The free energy of electrostatic interactions depends on the distance between the charges and the local dielectric constant of the medium. To dissect these structural and microenvironmental effects, the energies involved in specific ion pairs between enzyme and substrate need to be evaluated. In addition to those modeled in Fig. 1, reasonable ion pairs can be built between a lysine P1 substrate and aspartate at position 166, and between a glutamate P1 substrate and a lysine at position 166 (not shown). Although only the ion pair between a lysine P1 substrate and glutamate-156 has been confirmed by x-ray crystallography (32), all models are sterically reasonable and have favorable torsion angles.

To estimate the apparent electrostatic free energy, ΔG_e, involved in a putative ion pair between glutamate-156 and

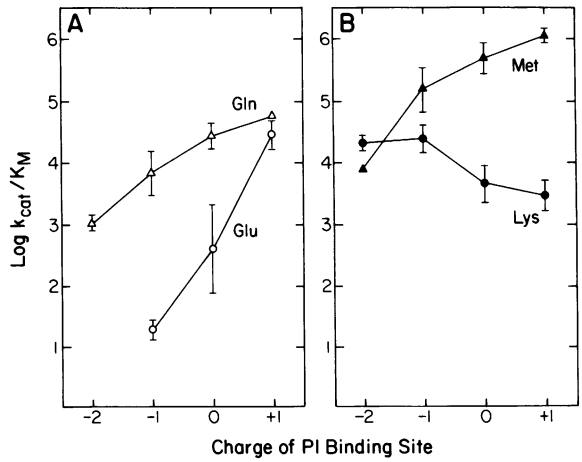

FIG. 2. Effect of charge in the P1 binding site due to substitutions at positions 156 and 166 on log(k_{cat}/K_m) for the P1 substrate: glutamine (△) and glutamate (○) (*A*); methionine (▲) and lysine (●) (*B*). Each data point represents the average values, and error bars represent the SD of log(k_{cat}/K_m) against each P1 substrate for enzymes having the same charge in the P1 binding site (data from Table 1).

[‖]The measured rate for enzyme acylation versus deacylation is 1:33 for wild-type subtilisin against succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (26). Under these circumstances, $k_2 \cong k_{cat}$ and $K_m \cong K_s$ (35).

Table 2. Effect of a negative charge at residue 156 on the preference for lysine over methionine P1 substrates

Enzyme 1	Enzyme 2	Change in substrate preference $\Delta\Delta\log(k_{cat}/K_m)$ (enzyme 1 − enzyme 2)
Glu-156/Asp-166	**Gln-156/Asp-166**	**1.02**
Glu-156/Asp-166	Ser-156/Asp-166	0.83
Glu-156/Asn-166	**Gln-156/Asn-166**	**1.43**
Glu-156/Asn-166	Gln-156/Gly-166	1.68
Glu-156/Asn-166	Ser-156/Gly-166	1.63
Glu-156/Asn-166	Ser-156/Asn-166	1.27
Glu-156/Gly-166	**Gln-156/Gly-166**	**1.90**
Glu-156/Gly-166	Gln-156/Asn-166	1.65
Glu-156/Gly-166	Ser-156/Gly-166	1.85
Glu-156/Gly-166	Ser-156/Asn-166	1.45
Glu-156/Lys-166	**Gln-156/Lys-166**	**0.82**
Glu-156/Lys-166	Ser-156/Lys-166	0.61
	Average $\Delta\Delta\log(k_{cat}/K_m)$ = 1.35 ± 0.43	

Enzyme 1 contains Glu-156 and enzyme 2 contains a neutral residue at position 156. The charge of the residue at position 166 is fixed for both enzymes 1 and 2. Enzyme comparisons shown in boldface type have the additional constraint that they contain an identical residue at position 166. Data from Table 1 were used to compute the difference in $\log(k_{cat}/K_m)$ (i.e., the substrate preference) between lysine and methionine P1 substrates for enzymes 1 and 2. Substrate $\text{preference}_{\text{Lys-Met}} = (\log k_{cat}/K_m)_{\text{Lys}} - (\log k_{cat}/K_m)_{\text{Met}}$. The change in substrate preference is $(\Delta\log k_{cat}/K_m)_{\text{Glu-156}} - (\Delta\log k_{cat}/K_m)_{\text{Ser (or Gln)-156}}$.

Table 3. Effect of a charge change at residue 166 on the preference for a complementary charge at the P1 position of the substrate

Enzyme 1	Enzyme 2	P1 substrates compared	Change in substrate preference
Glu-156/Asp-166	**Glu-156/Asn-166**	**Lys–Met**	**1.17**
Glu-156/Asp-166	Glu-156/Gly-166	Lys–Met	0.95
Glu-156/Asp-166	Glu-156/Ala-166	Lys–Met	1.15
Glu-156/Asp-166	Glu-156/Met-166	Lys–Met	1.34
Glu-156/Asp-166	Glu-156/Gln-166	Lys–Met	1.88
Glu-156/Glu-166	**Glu-156/Gln-166**	**Lys–Met**	**2.06**
Glu-156/Glu-166	Glu-156/Gly-166	Lys–Met	1.15
Glu-156/Glu-166	Glu-156/Ala-166	Lys–Met	1.37
Glu-156/Glu-166	Glu-156/Asn-166	Lys–Met	1.39
Glu-156/Glu-166	Glu-156/Met-166	Lys–Met	1.56
Gln-156/Asp-166	**Gln-156/Asn-166**	**Lys–Met**	**1.58**
Gln-156/Asp-166	Gln-156/Gly-166	Lys–Met	1.61
Gln-156/Asp-166	Ser-156/Gly-166	Lys–Met	1.78
Gln-156/Asp-166	Ser-156/Asn-166	Lys–Met	1.40
Ser-156/Asp-166	**Ser-156/Asn-166**	**Lys–Met**	**1.61**
Ser-156/Asp-166	Ser-156/Gly-166	Lys–Met	1.97
Ser-156/Asp-166	Glu-156/Gly-166	Lys–Met	1.75
Ser-156/Asp-166	Gln-156/Asn-166	Lys–Met	1.77
Glu-156/Lys-166	**Glu-156/Met-166**	**Glu–Gln**	**2.06**
Glu-156/Lys-166	Glu-156/Gly-166	Glu–Gln	1.80
Glu-156/Lys-166	Glu-156/Asn-166	Glu–Gln	1.60
Glu-156/Lys-166	Glu-156/Gln-166	Glu–Gln	2.53
Gln-156/Lys-166	**Gln-156/Gly-166**	**Glu–Gln**	**2.10**
Gln-156/Lys-166	Gln-156/Asn-166	Glu–Gln	2.65
Gln-156/Lys-166	Ser-156/Gly-166	Glu–Gln	1.95
Gln-156/Lys-166	Ser-156/Asn-166	Glu–Gln	2.72
Ser-156/Lys-166	**Ser-156/Gly-166**	**Glu–Gln**	**1.16**
Ser-156/Lys-166	Ser-156/Asn-166	Glu–Gln	2.03
Ser-156/Lys-166	Gln-156/Gly-166	Glu–Gln	1.29
Ser-156/Lys-166	Gln-156/Asn-166	Glu–Gln	1.84
		Average $\Delta\Delta\log(k_{cat}/K_m)$ = 1.71 ± 0.45	

Analogous to Table 2 except enzyme 1 contains a charged residue at position 166 and enzyme 2 contains a neutral residue at position 166. The charge of the residue at position 156 is fixed, and enzyme comparisons shown in boldface type are isosteric. The substrate preference (i.e., $\Delta\Delta\log k_{cat}/K_m$ for enzyme 1 − enzyme 2) is calculated (see Table 2) for indicated P1 substrates whose charge is complementary with the charge at position 166.

the lysine P1 substrate, the substrate preference, $\Delta log(k_{cat}/K_m)$ between lysine and methionine P1 substrates (Table 1) was calculated for the relevant enzymes. The $[\Delta\log(k_{cat}/K_m)]_{\text{Lys-Met}}$ isolates the substrate binding effects resulting from the ε-ammonium group of lysine. The change in the substrate preference $[\Delta\Delta\log(k_{cat}/K_m)]$ between enzymes containing glutamate-156 and a neutral residue at position 156 (keeping the charge at 166 fixed) reflects the apparent electrostatic effect of glutamate-156 (Table 2). The ΔG_e at 25°C, calculated from the average $\Delta\Delta\log(k_{cat}/K_m)$, is −1.8 ± 0.5 kcal/mol (9, 12). Closely isosteric enzyme comparisons (boldface type in Table 2) give very similar values to ones that only control for the charge at positions 156 and 166. The presence of a charged residue at position 166 significantly lowers the effect of the charge at position 156 [i.e., average $\Delta\Delta\log(k_{cat}/K_m)$ is equal to 0.82 ± 0.17 with aspartate or lysine at position 166 compared to 1.61 ± 0.21 with a neutral residue at position 166].

A similar analysis was used to determine the apparent electrostatic effect resulting from a charge at residue 166 and a complementary charge in the P1 position of the substrate (Table 3). The average change in substrate preference is greater for a charged residue at position 166 than at position 156 (1.71 ± 0.45 versus 1.35 ± 0.43). The ΔG_e for apparent electrostatic interaction with residue 166 is −2.3 ± 0.6 kcal/mol. A glutamate residue at position 156 marginally lowers the effect of a charge at position 166 [i.e., $\Delta\Delta\log(k_{cat}/K_m)$ was 1.57 ± 0.44 with glutamate-156 compared to 1.79 ± 0.55 without glutamate-156]. The change in substrate preference is comparable, although slightly greater, for an interaction of lysine-166 with the glutamate substrate compared to aspartate or glutamate-166 with the lysine substrate and yields average $\Delta\Delta\log(k_{cat}/K_m)$ values of 2.01 ± 0.47 and 1.53 ± 0.31, respectively. Closely isosteric enzyme comparisons (boldface type in Table 3) give similar values to those only controlled for the charge change at position 166. Enzymes that have glutamate-156 show a reduction in lysine to methionine P1 substrate preference as the size of the neutral residue at position 166 increases (e.g., compare glutamate-156/glycine-166, glutamate-156/alanine-166, glutamate-156/asparagine-166, glutamate-156/glutamine-166, and glutamate-156/methionine-166).

DISCUSSION

We attribute the general changes in substrate specificity resulting from charged amino acid substitutions at residues 156 and 166 in the P1 binding site to electrostatic effects. This is supported by the data showing that charged substitutions substantially increase the catalytic efficiency toward complementary charged P1 substrates (up to 1900 times) and decrease it toward similarly charged P1 substrates (Fig. 2). Furthermore, the effects on k_{cat}/K_m resulting from the combination of charged substitutions at residues 156 and 166 are roughly additive. Preliminary experiments show that P1 substrate preference for lysine over methionine decreases with increasing ionic strength for glutamate-156/aspartate-166 and glutamate-156/glycine-166, as expected for electrostatic effects. The differences in substrate preference are not dominated by steric repulsion because the kinetic effects best correlate with the charge and not the size of the interacting amino acid side chains. Moreover, reasonable molecular models can be built for enzyme–substrate ion pairs proposed

at positions 156 and 166; one such ion pair (at position 156) has been confirmed by x-ray crystallography (32).

In principle, increasing the hydrophilic character of the substrate binding site by charged amino acid substitutions may also increase the preference for hydrophilic over hydrophobic substrates. However, we observe only a slight increase in the preference for the glutamine P1 substrate over the methionine P1 substrate as the binding site becomes more negatively charged. In general, hydrophobic amino acid substitutions at residue 166 increase k_{cat}/K_m for extremely hydrophobic substrates only by a factor of 2–10 (7); the hydrophobic effects appear smaller than the electrostatic effects on substrate specificity.

Energetics of Specific Enzyme–Substrate Ion Pair Interactions. The variation in the estimates for ΔG_e for specific electrostatic interactions at residues 156 and 166 of -1.8 ± 0.5 and -2.3 ± 0.6 kcal/mol, respectively, emphasize the need to analyze many mutant enzymes to develop a statistically significant data set. Much of the variation can be explained by additional electrostatic or steric effects. For example, the presence of a charge at position 166 appears to buffer the effect of a charge at position 156 (Table 2). In addition, enzymes containing glutamate-156 show a reduction in lysine to methionine P1 substrate preference as the size of the neutral residue at position 166 increases (Table 3). Furthermore, the ΔG_e estimated for a lysine or glutamate P1 substrate with a complementary charge at position 166 are comparable but marginally different (-2.1 ± 0.4 and -2.7 ± 0.7 kcal/mol, respectively; Table 3). These minor trends tend to be confined to a narrow data set and we avoid generalizing from them.

It is unlikely that variation in ΔG_e results from large alterations in the overall structure of the P1 binding site, because high resolution x-ray crystal structures of 20 mutant subtilisins show little structural change except for the side-chain substitution and small local perturbations (R.R.B., M. Ultsch, B. Katz, and T. Kossiakoff, unpublished data). Furthermore, because estimates of ΔG_e are calculated from the preference for charged over sterically similar and neutral P1 substrates, variation arising from electrostatic effects on substrate binding outside the P1 binding site should be minimized. However, further x-ray crystallography will be necessary to define the precise mode(s) of substrate binding.

The average ΔG_e values for electrostatic interactions between substrate and enzyme at positions 156 and 166 of -1.8 and -2.3 kcal/mol, respectively, are below values estimated for more buried ion pairs in chymotrypsin (19) and phenylalanine tRNA synthetase (20) (-2.9 and -2.7 kcal/mol, respectively, under similar ionic strength and temperature conditions). Differences in ΔG_e values can be accounted for by differences in ion pair separations, formal charges, side-chain torsional energies, or local dielectric constants. Fersht and coworkers have measured the effect of a charge change in subtilisin at positions 99 (11) and 156 (A. J. Russell, P. G. Thomas, and A. R. Fersht, personal communication) on the pKa of histidine-64 located further than 10 Å away. From the shift in the pKa of histidine-64 caused by a charge at position 156 they estimate an apparent dielectric constant of 67 ± 9, under identical conditions as described here. Assuming identical ion pairs separated by 3 Å for a unit charge on the substrate and a complementary unit charge on the enzyme at position 156 or 166, we calculate effective dielectric constants of 61 ± 19 and 48 ± 12, respectively [in 0.1 M Tris·HCl (pH 8.6) at 25°C]. The apparent agreement between the effective dielectric constants at position 156 calculated from these two independent experimental lines may only be fortuitous; no adequate formalism exists for interpretation or prediction of a dielectric constant, which is a macroscopic quantity, at atomic resolution.

In practice, the charged side-chain substitutions have altered the substrate specificity of subtilisin from being chymotrypsin-like to being more trypsin-like (i.e., glutamate-156/glutamate-166) or more V-8 protease-like (i.e., glutamine-156/lysine-166). The fact that wild-type enzyme can be improved in k_{cat}/K_m toward any of the P1 substrates tested is encouraging to goals for tailoring biological specificity while maintaining protein function. These studies demonstrate the feasibility of engineering substrate specificity by altering electrostatic interactions and contribute to the data base required for evaluation of electrostatic theories in proteins and for rational design of ligand binding sites.

We thank Parkash Jhurani, Peter Ng, and Mark Vasser for oligonucleotide synthesis; John Burnier and Martin Struble for peptide synthesis; and Tony Kossiakoff, Steve Shire, and Colin Mitchinson for critical advice on this manuscript. We also thank Alan Russell, Paul Thomas, and Alan Fersht for exchanging results prior to publication. Funding for this work was provided by Genencor, Inc.

1. Smith, M. (1985) *Annu. Rev. Genet.* **19,** 423–462.
2. Villafranca, J. E., Howell, E. E., Voet, D. M., Strobel, M. S., Ogden, R. C., Abelson, J. N. & Kraut, J. (1983) *Science* **222,** 782–788.
3. Perry, L. J. & Wetzel, R. (1984) *Science* **226,** 555–557.
4. Estell, D. A., Graycar, T. P. & Wells, J. A. (1985) *J. Biol. Chem.* **260,** 6518–6521.
5. Rosenberg, S., Barr, P. J., Najarilan, R. C. & Hallewell, R. A. (1984) *Nature (London)* **312,** 77–80.
6. Courtney, M., Jallat, S., Tessier, L. H., Benavente, A., Crystal, R. G. & Lecocq, J. P. (1985) *Nature (London)* **313,** 149–151.
7. 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.
8. Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R. & Rutter, W. J. (1985) *Science* **228,** 291–297.
9. Fersht, A. R., Shi, J. P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y. & Winter, G. (1985) *Nature (London)* **314,** 235–238.
10. Winther, J. R., Kielland-Brandt, M. C. & Breddam, K. (1985) *Carlsberg Res. Commun.* **50,** 273–284.
11. Thomas, P. G., Russell, A. J. & Fersht, A. R. (1985) *Nature (London)* **318,** 375–376.
12. Fersht, A. R. (1985) in *Enzyme Structure and Mechanism* (Freeman, San Francisco), 2nd Ed., pp. 293–369.
13. Perutz, M. F. (1978) *Science* **201,** 1187–1191.
14. Bjerrum, N. (1923) *Z. Physik. Chem. Stoecliom. Verwandtschaftsl.* **106,** 219.
15. Matthew, J. B. (1985) *Annu. Rev. Biophys. Biophys. Chem.* **14,** 387–417.
16. Warshel, A. & Russell, S. T. (1984) *Q. Rev. Biophys.* **17,** 283–422.
17. Honig, B. H., Hubbell, W. L. & Flewelling, R. F. (1986) *Annu. Rev. Biophys. Chem.* **15,** 163–193.
18. Matthew, J. B., Gurd, F. R. N., Garcia-Moreno, B., March, K. L. & Shire, S. J. (1985) *Crit. Rev. Biochem.* **18,** 91–197.
19. Fersht, A. R. (1972) *J. Mol. Biol.* **64,** 497–509.
20. Mulivor, R. & Rappaport, K. P. (1973) *J. Mol. Biol.* **76,** 123–134.
21. Mehler, E. L. & Eichele, G. (1984) *Biochemistry* **23,** 3887–3891.
22. Rees, D. C. (1980) *J. Mol. Biol.* **141,** 323–326.
23. Wells, J. A., Vasser, M. & Powers, D. B. (1985) *Gene* **34,** 315–323.
24. Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A. & Chen, E. Y. (1983) *Nucleic Acids Res.* **11,** 7911–7925.
25. Adelman, J. P., Hayflick, J. S., Vasser, M. & Seeburg, P. H. (1983) *DNA* **2,** 183–193.
26. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. (1986) *Philos. Trans. R. Soc. London Ser. A* **317,** 415–423.
27. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
28. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
29. Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27,** 157.
30. Marquardt, D. W. (1963) *J. Soc. Ind. Appl. Math.* **11,** 431–441.
31. Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T. & Kraut, J. (1975) *J. Biol. Chem.* **250,** 7120–7126.
32. Poulos, T. L., Alden, R. A., Freer, S. T., Birktoft, J. J. & Kraut, J. (1976) *J. Biol. Chem.* **251,** 1097–1103.
33. Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C. & Wilcox, P. E. (1972) *Biochemistry* **11,** 2439–2449.
34. Robertus, J. D., Kraut, J., Arden, R. A. & Birktoft, J. J. (1972) *Biochemistry* **11,** 4293–4303.
35. Gutfreund, H. & Sturtevant, J. M. (1956) *Biochem. J.* **63,** 656.