from the deletion (see Fig. 1). This, we predict, adds 15 "new" amino acid residues (K-R-R-R-R-W-V-F-Q-S-H-L-R-Y-L) to the protein product of this gene and could conceivably affect the properties of this moiety. Likewise, clone X9-3, which also has a diminished cytopathic potential, has an "extended" envelope resulting from the substitution of the last five amino acids of *env* with the last 153 amino acids of 3′-*orf*. However, since we do not know precisely the replication competence of virus derived from X9-3, it is possible that reduced replication is, in this instance, the reason for its reduced cytopathic potential. Further studies will be necessary to determine precisely the genetic and biochemical changes that account for the loss of cytopathogenicity in clones X10-1 and X9-3. The present results suggest that caution should be used in selecting candidate immunogens for vaccination derived from the HTLV-III envelope region, since they might themselves be cytopathic for susceptible cells. Our data also suggest that novel therapeutic approaches such as those being developed specifically to block *tat* function may be ineffective unless they are designed to halt completely the expression of HTLV-III in infected cells.

REFERENCES AND NOTES

1. F. Barré-Sinoussi *et al.*, *Science* **220**, 868 (1983); R. C. Gallo *et al.*, *ibid.* **224**, 503 (1984).
2. M. S. Gottlieb *et al.*, *N. Engl. J. Med.* **305**, 1424 (1981); M. Popovic and R. C. Gallo, in *Immune Regulation*, M. Feldman and N. A. Mitcheson, Eds. (Humana, Clifton, NJ, 1985), pp. 167–180; A. G. Dalgleish *et al.*, *Nature (London)* **312**, 763 (1984); D. Klatzmann *et al.*, *Science* **225**, 69 (1984).
3. A. G. Fisher, E. Collalti, L. Ratner, R. C. Gallo, F. Wong-Staal, *Nature (London)* **319**, 262 (1985).
4. L. Ratner *et al.*, *ibid.* **313**, 227 (1985); S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, M. Alizon, *Cell* **40**, 9 (1985); R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985).
5. J. G. Sodroski *et al.*, *Nature (London)* **321**, 412 (1986).
6. M. B. Feinberg, R. F. Jarrett, A. Aldovini, R. C. Gallo, F. Wong-Staal, *Cell*, in press.
7. S. K. Arya, C. Guo, S. F. Josephs, F. Wong-Staal, *Science* **229**, 69 (1985); J. G. Sodroski, R. Patarca, F. Wong-Staal, C. Rosen, W. A. Haseltine, *ibid.*, p. 74.
8. A. G. Fisher *et al.*, *Nature (London)* **320**, 367 (1986); A. Dayton *et al.*, *Cell* **44**, 941 (1986).
9. T. H. Lee *et al.*, *Science* **231**, 1546 (1986); G. Franchini *et al.*, in preparation.
10. M. Popovic *et al.*, *Science* **224**, 497 (1984).
11. M. E. Harper, L. M. Marselle, R. C. Gallo, F. Wong-Staal, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 772 (1986).
12. H. Mitsuya *et al.*, *ibid.* **82**, 7096 (1985).
13. L. Ratner *et al.*, *Nucleic Acids Res.* **13**, 8219 (1985); L. Ratner *et al.*, in preparation.
14. G. Franchini, personal communication.
15. M. Essex, personal communication.
16. L. J. Siegal *et al.*, *Virology* **148**, 226 (1986).
17. G. M. Shaw *et al.*, *Science* **226**, 1165 (1984); B. H. Hahn *et al.*, *Nature (London)* **312**, 166 (1984).
18. F. Veronese di Marzo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5199 (1985); *Science* **229**, 1402 (1985).
19. C. Queen and D. Baltimore, *Cell* **33**, 741 (1983).
20. T. Maniatis, E. F. Fitsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), pp. 23–233.
21. J. Chirgwin, A. Prybyla, R. McDonald, W. Rutter, *Biochemistry* **18**, 5294 (1979).

22. We thank J. Lemp for assistance in the preparation of virus stocks, M. Gonda and M. White for performing electron microscopy, and L. Simek, L. Stingl, and R. Zicht for technical assistance. The plasmid clone pHXB2hygro was a gift from M. Feinberg. This work was supported in part by a grant to A.G.F. from the Foundation for the Advancement of Education in Science in association with Dupont de Nemours Inc., and an American Cancer Society Institute grant to L.R.

14 April 1986; accepted 17 June 1986

Probing Steric and Hydrophobic Effects on Enzyme-Substrate Interactions by Protein Engineering

D. A. ESTELL, T. P. GRAYCAR, J. V. MILLER, D. B. POWERS, J. P. BURNIER, P. G. NG, J. A. WELLS*

Steric and hydrophobic effects on substrate specificity were probed by protein engineering of subtilisin. Subtilisin has broad peptidase specificity and contains a large hydrophobic substrate binding cleft. A conserved glycine (Gly^{166}), located at the bottom of the substrate binding cleft, was replaced by 12 nonionic amino acids by the cassette mutagenesis method. Mutant enzymes showed large changes in specificity toward substrates of increasing size and hydrophobicity. In general, the catalytic efficiency (k_{cat}/K_m) toward small hydrophobic substrates was increased (up to 16 times) by hydrophobic substitutions at position 166 in the binding cleft. Exceeding the optimal binding volume of the cleft (~160 Å^3), by enlarging either the substrate side chain or the side chain at position 166, evoked precipitous drops in catalytic efficiency (k_{cat}/K_m) (up to 5000 times) as a result of steric hindrance.

BINDING SPECIFICITY IS A UBIQUITOUS feature of biological macromolecules; it is determined by chemical forces including hydrogen bonding and electrostatic, hydrophobic, and steric interactions (*1*, *2*). Substrate specificity studies of enzymes have been a traditional means of probing the relative importance of these binding forces. Although substrate analogs can be synthesized chemically, the production of complementary enzyme analogs has been extremely limited to natural variants (*3*) or chemically modified enzyme derivatives (*4*). Recently, through the technology of protein engineering, it has been possible to tailor a protein by site-directed mutagenesis of its cloned DNA sequence (*5*), and the mutant protein can be expressed in a heterologous host.

We have chosen subtilisin, a serine-class endopeptidase (M_r, 27,500), as a model to study the energetics of substrate specificity. The three-dimensional structure of subtilisin BPN′ (from *Bacillus amyloliquefaciens*) has been solved by x-ray crystallography to 2.5 Å resolution (*6*) and more recently to 1.8 Å resolution (*7*). Protein engineering has been applied to the cloned subtilisin gene (*8*) to create subtilisins with greater oxidative stability (*9*), to determine the importance of hydrogen bond formation in transition-state stabilization (*10*), to introduce disulfide bonds into subtilisin (*11*), and to alter the *p*H activity profile of the enzyme (*12*).

Our studies here are designed to probe steric and hydrophobic effects on substrate specificity of subtilisin. Steric and hydrophobic effects are among the most general and complicated of the chemical binding forces because they can embody the entire substrate. To analyze the contribution of these forces to substrate specificity, we have produced 12 noncharged mutations in the P1 (*13*) binding cleft of subtilisin (Fig. 1), and have determined the resulting specificities of mutant enzymes against P1 substrates of varying size and hydrophobicity. Mutant enzymes exhibit large changes in substrate specificity caused by a combination of steric hindrance and enhanced hydrophobic interactions.

To map the topology and chemical nature of the P1 binding cleft in subtilisin the kinetic parameters, k_{cat} (turnover number) and K_m (Michaelis constant), were determined for ten different noncharged substrates having the form succinyl-L-Ala-L-Ala-L-Pro-L-[X]-*p*-nitroanilide, where X is the P1 amino acid. The ratio of k_{cat}/K_m (also

D. A. Estell, T. P. Graycar, J. V. Miller, Research Department, Genencor, Inc., 180 Kimball Way, South San Francisco, CA 94080.
J. A. Wells and D. B. Powers, Department of Biocatalysis, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080.
J. P. Burnier and P. G. Ng, Department of Developmental Biology, Genentech, South San Francisco, CA 94080.

*To whom correspondence should be addressed.

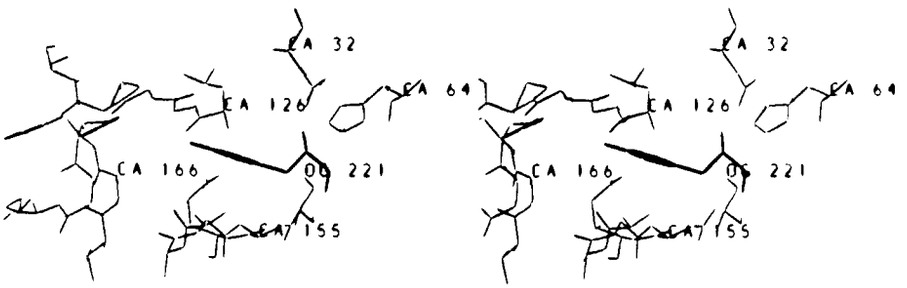

Fig. 1. Stereoview of the P1 binding site of subtilisin BPN′ from *B. amyloliquefaciens* showing a model of a bound peptide substrate containing Tyr at the P1 position in the E · S complex. The α-carbon atoms (CA) for His^{64} and Asp^{32} plus the γ-oxygen atom (OG) for Ser^{221} that comprise the catalytic triad are labeled. The P1 binding cleft is made up of residues from Leu^{126} (CA labeled) to Pro^{129} on the top, from Glu^{156} to Ala^{152} on the bottom (CA of Asn^{155} labeled), and from Gly^{166} (CA labeled) to Gly^{169} at the back. The model is based on crystallographic studies of bound peptide inhibitors (*18–20*) and was provided by R. Bott.

referred to as catalytic efficiency) is the apparent second-order rate constant for the conversion of free enzyme plus substrate (E+S) to enzyme plus products (E+P) (*1, 2*). The log (k_{cat}/K_m) is proportional to transition-state binding energy, $\Delta G_T^{\ddagger}$ (*14*). A plot of the log (k_{cat}/K_m) as a function of the hydrophobicity of the P1 side chain (Fig. 2) shows a strong linear relation (slope = −0.89 ± 0.07), with the exception of the glycine substrate which shows evidence for nonproductive binding. These data show that relative differences between transition-state binding energies can be accounted for by differences in P1 side chain hydrophobicity. When the transition-state binding energies are calculated for these substrates and plotted as a function of their respective side chain hydrophobicities, the slope of the line is 1.2 ± 0.09. A slope greater than unity, as is also the case for chymotrypsin (*2, 15*), suggests that the P1 binding cleft is more hydrophobic than ethanol or dioxane, solvents that were used to determine empirically the hydrophobicity of amino acids (*16*).

For amide hydrolysis by subtilisin BPN′, k_{cat} can be interpreted as the acylation rate constant, and K_m approximates K_s (*17*), the dissociation constant for the Michaelis complex (E · S). The fact that the log k_{cat}, as well as log $1/K_m$, correlates with substrate hydrophobicity is consistent with proposals (*18, 19*) that during the acylation step the P1 side chain moves deeper into the hydrophobic cleft as the substrate advances from the Michaelis complex (E · S) to the tetrahedral transition-state complex (E · S≠). However, these data can also be interpreted as the hydrophobicity of the P1 side chain effecting the orientation, and thus the susceptibility of the scissile peptide bond to nucleophilic attack by the hydroxyl group of the catalytic Ser^{221}.

X-ray crystal structures of subtilisin containing covalently bound peptide inhibitors (*18*), product complexes (*19*), and transition-state analogs (*20*) help to delineate an extended peptide binding cleft. These data, as well as kinetic data (*21, 22*), show that subsites in the binding cleft are capable of interacting with substrate amino acid residues from P4 to P2′ (*13*). The P1 binding site is a large open cleft (Fig. 1). Our kinetic data, and that of others (*21, 22*), show that subtilisin exhibits a broad specificity although it prefers large P1 side chains.

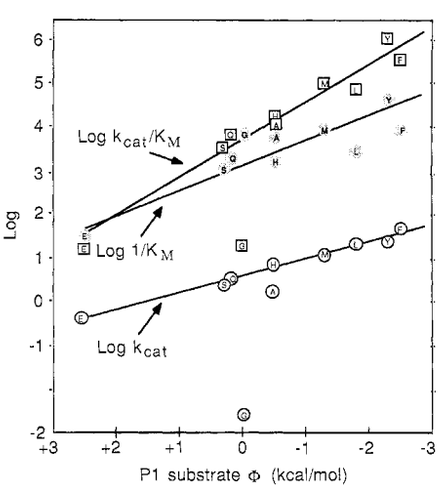

Fig. 2. Effect of the hydrophobicity of the P1 substrate side chain on the kinetic parameters of wild-type *B. amyloliquefaciens* subtilisin. Subtilisin was purified (*9*) from *B. subtilis* culture supernatants expressing the *B. amyloliquefaciens* subtilisin gene (*8*). Kinetic parameters, K_m (moles per liter) and k_{cat} (per second) were measured by a modified progress curve analysis (*9*) for tetrapeptide substrates having the form succinyl-L-Ala-L-Ala-L-Pro-L-[X]-*p*-nitroanilide (where X is the P1 amino acid). Errors in k_{cat} and K_m for all values reported are less than 5 percent. The hydrophobicity values represent the free energy for transfer for each amino acid measured from H_2O to ethanol or dioxane relative to glycine (*16*), except for the case of glutamate in which only a calculated hydrophobicity value is available (*28*). The single amino acid code is used to indicate each data point as follows: A, Ala; C, Cys; E, Glu; F, Phe; G, Gly; H, His; I, Ile; L, Leu; M, Met; P, Pro; Q, Gln; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. Based on linear least squares analysis omitting the Gly P1 substrate, lines have slopes and standard errors of −0.89 ± 0.07, −0.58 ± 0.10, and −0.40 ± 0.04 for log (k_{cat}/K_m) (squares), log ($1/K_m$) (shaded circles), and log (k_{cat}) (open circles), respectively.

The data above suggest that increasing the hydrophobicity of the P1 binding cleft by protein engineering may increase binding of hydrophobic P1 substrates. However, because the hydrophobicity of an amino acid is proportional to its surface area (*23*), increasing the hydrophobicity of the P1 cleft by amino acid substitutions may sterically hinder binding of larger substrates. Because of difficulties in predicting the relative importance of these two opposing effects, we elected to generate 12 noncharged mutations at Gly^{166} in the P1 binding cleft, and to determine the resulting specificities against noncharged substrates of varied size and hydrophobicity. Gly^{166} is located at the bottom of the P1 cleft (Fig. 1), and amino acid substitutions can project their side chains into the cleft to interact with the substrate.

To produce many mutations at position 166 efficiently, two silent restriction sites were introduced in the subtilisin gene to flank closely the target codon 166 (Fig. 3). Synthetic duplex DNA oligonucleotides (cassettes), containing the mutations of interest, were ligated into the gap between these restriction sites. The high efficiency of this cassette mutagenesis method (*24*) is attributed to the mutagenic event being a simple ligation as opposed to heteroduplex synthesis in M13 (*5*). Plasmids containing each mutant *B. amyloliquefaciens* subtilisin gene were expressed (*9, 11*) at roughly equivalent levels in a protease-deficient strain of *B. subtilis*.

To probe the change in substrate specificity caused by steric alterations in the P1 cleft, mutants at position 166 were analyzed kinetically with increasingly larger P1 substrates (Fig. 4). Values of log (k_{cat}/K_m) allow direct comparisons of transition-state binding energies between various enzyme-substrate pairs (*14*).

As the volume of the side chain at position 166 increases, the substrate preference shifts from large to small P1 side chains (Fig. 4A). Enlarging the side chain at position 166 causes k_{cat}/K_m to decrease in proportion to the size of the P1 substrate side chain (from Gly^{166} through Trp^{166}, the k_{cat}/K_m for the Tyr substrate is decreased most and then followed in order by the Phe, Met, and Ala P1 substrates). Specific steric changes in the position 166 side chain, such as the presence of a β-hydroxyl group, β- or γ-aliphatic branching, cause large decreases in k_{cat}/K_m for larger P1 substrates (Fig. 4B).

The change in log (k_{cat}/K_m) for the P1

substrate Ala, Met, Phe, and Tyr as a function of the volume of the position 166 amino acid (*25*) is shown in Fig. 5. Catalytic efficiency for the Ala substrate (top panel) reaches a maximum for Ile[166], and for the Met substrate it reaches a maximum between Val[166] and Leu[166]. The Phe substrate shows a broad log (k_{cat}/K_m) peak but is optimal with Ala[166]. Here, the β-branched substitutions at position 166 substitutions form a line that is parallel to, but roughly 50 times lower in k_{cat}/K_m than side chains of similar size (Cys[166] compared to Thr[166], Leu[166] compared to Ile[166]). The Tyr substrate is most efficiently utilized by wild-type enzyme (Gly[166]), and there is a steady decrease as the side chain becomes larger at position 166. The β-branched and γ-branched substitutions form a parallel line below the other noncharged substitutions of similar molecular volume.

The optimal substitution at position 166 (Fig. 5) decreases in volume with increasing volume of the P1 substrate (Ile[166] with the Ala substrate, Leu[166] with the Met substrate, Ala[166] with the Phe substrate, and Gly[166] with the Tyr substrate). The combined volumes for these optimal pairs may approximate the volume for productive binding in the P1 cleft. For the optimal pairs, Gly[166] for the Tyr substrate, Ala[166] for the Phe substrate, Leu[166] for the Met substrate, Val[166] for the Met substrate, and Ile[166] for the Ala substrate, the combined volumes are 138, 163, 207, 181, and 129 Å^3, respectively. This gives an average combined side chain volume of 160 ± 30 Å^3 for productive binding in the transition state.

The effect of volume, in excess to the productive binding volume, on the drop in transition-state binding energy can be estimated from the Tyr substrate curve (Fig. 5, bottom panel) because these data, along with modeling studies (Fig. 1), suggest that any substitution beyond glycine causes steric repulsion. A best-fit line drawn to all the data gives a slope of −0.019 ± 0.003 Δlog (k_{cat}/K_m) per cubic angstrom, indicating a loss of about 2.6 kcal/mol in transition-state binding energy per 100 Å^3 of excess volume. (100 Å^3 is approximately the volume of a leucyl side chain.)

It is oversimplified to treat side chain interactions from the enzyme and substrate as simple steric packing volumes (*25*). For example, the side chains for Ser, Thr, Tyr, and Trp, and to a lesser extent Cys and Met, are likely to be hydrated in the absence of a complementary hydrogen bond interaction between the enzyme and substrate. Thus, the actual volumes for these side chains may be larger by the volume of a water molecule (~20 Å^3). In general, these observed effects are small because isosteric amino acid substitutions at position 166 that differ in hydrogen bond potential have nearly identical kinetic behavior (for example, compare Thr[166] with Val[166], or Tyr[166] with Phe[166]). There is a large difference in steric repulsion between the Phe and Tyr P1 substrates that may be partly accounted for by the added volume of hydration of the Tyr hydroxyl group. However, molecular modeling (Fig. 1) shows that, even in the absence of hydration, addition of a β carbon at position 166 can sterically hinder binding of a Tyr but not a Phe P1 substrate.

Steric repulsion is a reasonable explanation for the sharp declines in the catalytic efficiencies with increasing molecular volume. This is because the repulsion term for van der Waals potential energy function is extremely dependent on the distance between the two interacting groups (that is, $1/r^{12}$ where r is the separation distance) (*2*). The fact that all enzymes maintain activities that are greater than or equal to wild-type enzyme with the Ala P1 substrate would argue that the specificity changes resulting from P1 cleft mutations are not the result of large disruptions in enzyme structure. Indeed, electron density difference maps of a number of position 166 mutant proteins (*7*), as well as subtilisin disulfide mutants (*26*), show that, except for the side chain substitution, very small structural changes are produced. Molecular modeling supports the interpretations of steric repulsion (*27*); however, high-resolution x-ray crystallography containing bound transition-state substrate analogs will be necessary to validate these interpretations.

Substantial increases in k_{cat}/K_m occur with enlargement of the position 166 side chain, except for the Tyr P1 substrate (Fig. 5). For example, k_{cat}/K_m increases in progressing from Gly[166] to Ile[166] for the Ala substrate (net of ten times), from Gly[166] to Leu[166] for the Met substrate (net of ten

```
                                                                  166
1. Wild-type amino acid sequence:     Thr Ser Gly Ser Ser Ser Thr Val Gly Tyr Pro Gly

2. Wild-type DNA sequence:        5'-ACT TCC GGC AGC TCA AGC ACA GTG GGC TAC CCT GGT-3'
                                  3'-TGA AGG CCG TCG AGT TCG TGT CAC CCG ATG GGA CCA-5'

3. pΔ166 DNA sequence:                  *                                       *
                                  5'-ACT TCC GGG AGC TCA A................C CCG GGT-3'
                                  3'-TGA AGG CCC TCG AGT T................G GGC CCA-3'
                                            Sac I                           Xma I

4. pΔ166 cut with Sac I and Xma I:      *                                       *
                                  5'-ACT TCC GGG AGC T                      pCCG GGT-3'
                                  3'-TGA AGG CCCp                             CA-5'

5. Cut pΔ166 ligated with               *                          ***          *
   duplex DNA cassette pools:     5'-ACT TCC GGG AGC TCA AGC ACA GTG NNN TAC CCG GGT-3'
                                  3'-TGA AGG CCC TCG AGT TCG TGT CAC NNN ATG GGC CCA-5'
```

Fig. 3. Strategy for construction of mutant codons at position 166 of *B. amyloliquefaciens* subtilisin by cassette mutagenesis (*24*). The wild-type DNA sequence (line 2) was altered by site-directed mutagenesis in M13 (*5*), to introduce a 13-bp deletion (dotted line) and Sac I and Xma I sites (shaded sequences) that closely flank codon 166. The subtilisin gene fragment was subcloned back into the *Escherichia coli–B. subtilis* shuttle plasmid, pBS42 (*33*), giving the plasmid pΔ166 (line 3). pΔ166 was cut open with Sac I and Xma I, and gapped linear molecules were purified (line 4). Pools of single-stranded oligonucleotides were annealed to give duplex DNA cassettes (shaded sequences in line 5) that were ligated into gapped pΔ166. This construction restored the coding sequence except over position 166 (NNN in line 5). Asterisks denote sequence changes from the wild-type sequence. Mutant sequences were confirmed by dideoxy sequencing in plasmid vehicles (*34*) and plasmids were transformed into the *B. subtilis* strain, BG2036 (*35*).

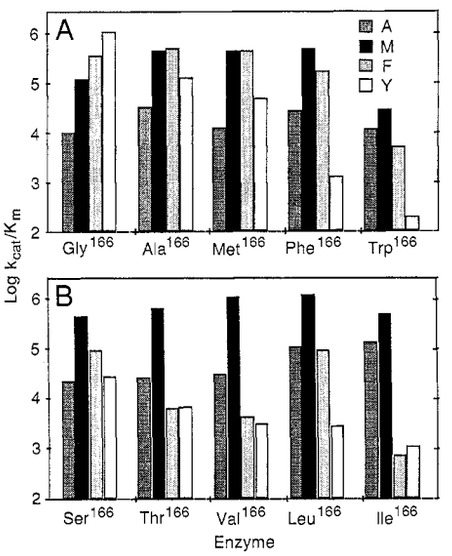

Fig. 4. Effect of position 166 side chain substitutions on P1 substrate specificity. Mutant proteins were purified and kinetic parameters were analyzed (*9*) with substrates having the form succinyl-L-Ala-L-Ala-L-Pro-L-[X]-*p*-nitroanilide, where X is Ala, Met, Phe, or Tyr indicated above each bar as A, M, F, or Y, respectively. Each bar represents the log (k_{cat}/K_m). (A) Position 166 variant subtilisins containing nonbranched alkyl and aromatic side chain substitutions arranged in order of increasing molecular volume (*25*). (B) A series of variant enzymes progressing through β- and γ-branched aliphatic side chain substitutions of increasing molecular volume.

times), and from Gly^{166} to Ala^{166} for the Phe substrate (net of two times). The increases in k_{cat}/K_m cannot be entirely explained by the attractive terms in the van der Waals potential energy function because of the weak nature and strong distance dependence ($1/r^6$) of these attractive forces (*2, 28*). For example, Levitt (*28*) has calculated that the van der Waals attraction between two methionyl residues would produce a maximal interaction energy of roughly −0.2 kcal/mol. This energy would translate to only a 1.4-fold increase in k_{cat}/K_m (*14*). Furthermore, hydrogen bonding or electrostatic interactions cannot account for increases in k_{cat}/K_m that result from interactions between aliphatic or aromatic side chains on the enzyme and P1 substrate.

The increases of catalytic efficiency caused by side chain substitutions at position 166 are better accounted for by increases in the hydrophobicity of the P1 cleft. The increase in k_{cat}/K_m observed for the Ala and Met substrates with increasing size of position 166 side chain would be expected because hydrophobicity is roughly proportional to side chain surface area (*23*). Moreover, a hydrophobic effect can be seen when comparing k_{cat}/K_m for isosteric substitutions that differ in hydrophobicity, such as Ser^{166} and Cys^{166} (Fig. 5). Cysteine is considerably more hydrophobic than serine (−1.0 compared to +0.3 kcal/mol) (*16, 28*). The difference in hydrophobicity correlates with the observation that Cys^{166} becomes more catalytically efficient relative to Ser^{166} as the hydrophobicity of the substrate increases (that is, Ala < Met < Tyr < Phe). Steric hindrance cannot explain these differences because serine is considerably smaller than cysteine (99 compared to 118 Å^3) (*25*).

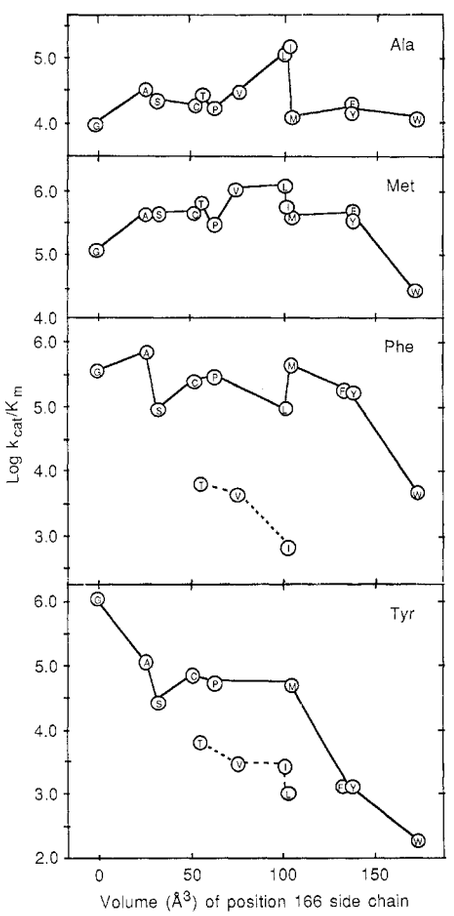

Fig. 5. Effect of position 166 side chain molecular volume (*25*) on the log(k_{cat}/K_m) for the P1 substrates: Ala (top), Met, Phe, and Tyr (bottom). Measurements of k_{cat} and K_m were made as in Fig. 4 and the single-letter amino acid code is used to indicate the data points for each position 166 substitution.

We find hydrophobicity a useful descriptive quantity. However, a molecular understanding of hydrophobicity, and its relation to the energetics of water structure and van der Waals attractive energy, remains unclear. For example, the specificity differences observed between Ser^{166} and Cys^{166} may be more precisely interpreted as differences in water structure caused by the greater hydrogen bonding potential of serine versus cysteine (*29*).

The Ile^{166} mutation illustrates particularly well that large changes in specificity can be produced by a single mutation that alters the structure and hydrophobicity of the P1 binding cleft (Fig. 6). Progressing through the small hydrophobic substrates, a maximal specificity improvement over wild-type occurs for the Val substrate (k_{cat}/K_m increases 16 times). As the size of the substrate side chain increases, these enhancements shrink to near unity (that is, Leu and His substrates). Against large aromatic substances of increasing size the Ile^{166} enzyme becomes less efficient and is more than 1000 times less effective for the Tyr substrate than is Gly^{166}. We attribute the increase in catalytic efficiency toward the small hydrophobic substrates for Ile^{166} compared to Gly^{166} to the greater hydrophobicity of isoleucine, −1.8 kcal/mol compared to 0 kcal/mol for glycine (*16*). We interpret the decrease in catalytic efficiency toward the very large substrates for Ile^{166} versus Gly^{166} as steric repulsion.

The specificity differences between Gly^{166} and Ile^{166} are similar to the specificity differences between chymotrypsin and the evolutionary relative elastase (*3, 15*). In elastase, bulky amino acid substitutions at the top of the P1 binding pocket block access to the P1 binding site for large hydrophobic substrates that are preferred by chymotrypsin. In addition, the catalytic efficiencies toward small hydrophobic substrates are greater for elastase than for chymotrypsin as we observe for Ile^{166} compared to Gly^{166} in subtilisin.

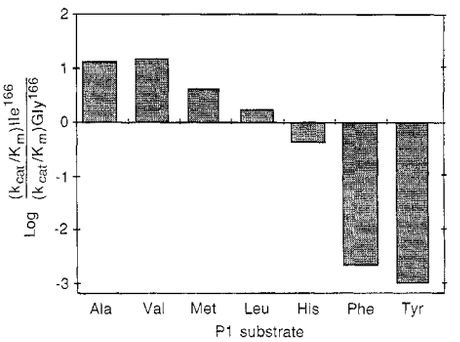

Fig. 6. Substrate specificity differences between Ile^{166} and wild-type (Gly^{166}) against a series of aliphatic and aromatic substrates. The P1 amino acid for each substrate is indicated below each bar. Each bar represents the difference in log (k_{cat}/K_m) for Ile^{166} minus Gly^{166} subtilisin.

A mutation of Met^{398} → Leu, believed to be the P1′ binding site of carboxypeptidase Y (*30*), was found to increase the catalytic efficiency toward large P1′ substrates (as high as 1.4 times) and decrease catalytic efficiency toward smaller P1′ substrates (as low as 3.6 times). Because Leu is both smaller and more hydrophobic than Met, the steric and hydrophobic effects could not be assessed separately.

It is necessary to evaluate many enzyme-substrate pairs to begin to quantify independently steric and hydrophobic effects on substrate specificity. Although the hydrophobic effect is substantial, it is considerably outweighed by steric repulsion. Furthermore, if we had chosen only β- or γ-branched or non-β-branched substitutions, we would not have separated the steric hindrance due specifically to β- and γ-branching from total side chain volume.

Mutations in the P1 binding cleft of subtilisin produce large changes in specificity for the P1 amino acid, and produce some mutant enzymes with increased catalytic efficiencies toward particular substrates. In addition, the k_{cat}/K_m for some of the optimal enzyme-substrate pairs (for example, Leu^{166} to Met substrate) even exceeds the k_{cat}/K_m for the best wild-type case (that is, Gly^{166} to Tyr substrate). These findings contrast results for mutations in trypsin (*31*), where the relative substrate preference toward arginine compared to lysine substrates was changed by a factor of 20 by a large, and differential drop in catalytic efficiency of the mutant relative to the wild-type enzyme of 200- and 4000-fold. The subtilisin mutants are significant because they show that specificity can be dramatically changed while maintaining, or even enhancing, catalytic efficiency.

It may appear surprising that Gly^{166} is conserved in five known extracellular *Bacillus* subtilisins (*8, 22, 32*), and yet Gly^{166} can

be freely substituted to produce highly functional mutant proteases with generally more narrowed substrate specificities. However, a broadly specific subtilisin (Gly^{166}) is probably more biologically adaptive because, unlike many other microbes that secrete a myriad of proteases with narrowed and different specificities, *Bacillus* species secrete only two major extracellular proteases, subtilisin and neutral protease.

Engineered proteases tailored to have highly restricted specificities should be very useful in protein chemistry, synthetic chemistry, and industry. In addition, they contribute to the database needed for protein design.

REFERENCES AND NOTES

1. W. P. Jencks, in *Catalysis in Chemistry and Enzymology* (McGraw-Hill, New York, 1969), pp. 321–436.
2. A. Fersht, *Enzyme Structure and Mechanism* (Freeman, San Francisco, ed. 2, 1985), pp. 293–369.
3. J. Kraut, *Annu. Rev. Biochem.* **46**, 331 (1977).
4. E. T. Kaiser, D. S. Lawrence, S. E. Rokita, *ibid.* **54**, 565 (1985).
5. M. Smith, *Annu. Rev. Genet.* **19**, 423 (1985).
6. C. S. Wright, R. A. Alden, J. Kraut, *Nature (London)* **221**, 235 (1969); J. Drenth, W. G. J. Hol, J. Jansonius, R. Kockoek, *Eur. J. Biochem.* **26**, 177 (1972).
7. R. R. Bott *et al.*, in preparation.
8. J. A. Wells, E. Ferrari, D. J. Henner, D. A. Estell, E. Y. Chen, *Nucleic Acids Res.* **11**, 7911 (1983).
9. D. A. Estell, T. P. Graycar, J. A. Wells, *J. Biol. Chem.* **260**, 6518 (1985).
10. J. A. Wells, B. C. Cunningham, T. P. Graycar, D. A. Estell, *Philos. Trans. R. Soc. London Ser. A* **317**, 415 (1986); P. Bryan, M. W. Pantoliano, S. G. Quill, H. Y. Hsiao, T. Poulos, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3743 (1986).
11. J. A. Wells and D. B. Powers, *J. Biol. Chem.* **261**, 6564 (1986).
12. P. G. Thomas, A. J. Russell, A. R. Fersht, *Nature (London)* **318**, 375 (1985); D. A. Estell, T. P. Graycar, D. B. Powers, J. A. Wells, unpublished results.
13. Peptide substrate nomenclature can be diagrammed as

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

where the scissile peptide bond is between P1 and P1′ [I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.* **27**, 157 (1967)].
14. According to transition state theory [see equation 12.4 in reference (*2*)], the free energy difference between the free enzyme plus substrate (E + S) and the transition state complex ($E \cdot S^{\ddagger}$) can be calculated from Eq. 1,

$$\Delta G_T^{\ddagger} = -RT \ln k_{cat}/K_m + RT \ln kT/h \quad (1)$$

in which k_{cat} is the turnover number, K_m is the Michaelis constant, R is the gas constant, T is the temperature, k is Boltzmann's constant, and h is Planck's constant. Specificity differences are expressed quantitatively as differences between transition state binding energies ($\Delta\Delta G_T^{\ddagger}$) and can be calculated from Eq. 2.

$$\Delta\Delta G_T^{\ddagger} = -RT \ln (k_{cat}/K_m)_A/(k_{cat}/K_m)_B \quad (2)$$

A and B represent either two different substrates assayed against the same enzyme or two mutant enzymes assayed against the same substrate.
15. J. W. Harper *et al.*, *Biochemistry* **23**, 2995 (1984).
16. Y. Nozaki and C. Tanford, *J. Biol. Chem.* **246**, 2211 (1971); C. Tanford, *Science* **200**, 1012 (1978).
17. The relative rates for hydrolysis of *p*-nitroanilide versus *p*-nitrophenyl ester substrates of succinyl-L-Ala-L-Ala-L-Pro-L-Phe is 1:33 (T. Graycar and D. Estell, unpublished). Because the rate of acylation (that is, k_2) is 33 times slower than the rate of deacylation (that is, k_3), k_{cat} nearly equals k_2, and K_m approximates the dissociation constant (K_s) for the E · S complex [H. Gutfreund and J. M. Sturtevant, *Biochem. J.* **63**, 656 (1956)].
18. J. D. Robertus, R. A. Alden, J. J. Birktoft, J. Kraut, J. C. Powers, P. E. Wilcox, *Biochemistry* **11**, 2439 (1972).
19. J. D. Robertus, J. Kraut, R. A. Alden, J. J. Birktoft, *ibid.*, p. 4293.
20. D. A. Matthews, R. A. Alden, J. J. Birktoft, S. T. Freer, J. Kraut, *J. Biol. Chem.* **250**, 7120 (1975); T. L. Poulos, R. A. Alden, S. T. Freer, J. J. Birktoft, J. Kraut, *ibid.* **251**, 1097 (1976).
21. M. Philipp and M. L. Bender, *Mol. Cell. Biochem.* **51**, 5 (1983).
22. F. S. Markland and E. L. Smith, *The Enzymes*, P. D. Boyer, Ed. (Academic Press, New York, 1971), vol. 3, pp. 561–608.
23. G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, M. H. Zehfus, *Science* **229**, 834 (1985); J. A. Reynolds, D. B. Gilbert, C. Tanford, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2925 (1974).
24. J. A. Wells, M. Vasser, D. B. Powers, *Gene* **34**, 315 (1985).
25. The values for volumes of side chains are taken as the average packing volumes determined from x-ray crystal structures of proteins; C. Chothia, *Annu. Rev. Biochem.* **53**, 537 (1984).
26. B. A. Katz and A. A. Kossiakoff, in preparation.
27. R. R. Bott, unpublished results.
28. M. Levitt, *J. Mol. Biol.* **104**, 59 (1976).
29. I. C. Paul, in *Chemistry of the −SH Group*, S. Patai, Ed. (Wiley-Interscience, New York, 1974), pp. 111–149.
30. J. R. Winther, M. C. Kielland-Brandt, K. Breddam, *Carlsberg Res. Commun.* **50**, 273 (1985).
31. C. S. Craik *et al.*, *Science* **228**, 291 (1985).
32. P. Nedkov, W. Oberthur, G. Braunitzer, *Hoppe Seyler's Z. Physiol. Chem.* **364**, 1537 (1983); M. L. Stahl and E. Ferrari, *J. Bacteriol.* **158**, 411 (1984).
33. L. Band and D. J. Henner, *DNA* **3**, 17 (1984).
34. E. Y. Chen and P. H. Seeberg, *ibid.* **4**, 165 (1985).
35. M. Yang, E. Ferrari, D. J. Henner, *J. Bacteriol.* **160**, 15 (1984).

36. We thank R. Bott for help with molecular graphics; and R. Bott and B. Katz for making crystallographic data available prior to publication; the organic chemistry department at Genentech for oligonucleotide synthesis; T. Kossiakoff, I. D. Kuntz, J. Rubin, and R. M. Stroud for critical advice on the manuscript; and W. Anstine for preparation and graphics of this manuscript. Support provided by Genencor.

18 February 1986; accepted 5 June 1986

A Mouse Homeo Box Gene Is Expressed in Spermatocytes and Embryos

MICHAEL R. RUBIN, LESLIE E. TOTH, MAYURI D. PATEL, PETER D'EUSTACHIO, M. CHI NGUYEN-HUU

The MH-3 gene, which contains a homeo box that is expressed specifically in the adult testis, was identified and mapped to mouse chromosome 6. By means of in situ hybridization with adult testis sections and Northern blot hybridization with testis RNA from prepuberal mice and from *Sl/Sl*^{d} mutant mice, it was demonstrated that this gene is expressed in male germ cells during late meiosis. In the embryo, MH-3 transcripts were present at day 11.5 post coitum, a stage in mouse development when gonadal differentiation has not yet occurred. The MH-3 gene may have functions in spermatogenesis and embryogenesis.

AMONG THE GENES THAT CONTROL development in the fruit fly, the homeotic and segmentation genes seem to control the identity, polarity, and number of body segments (*1–4*). A characteristic component of such genes in the Antennapedia, Bithorax, and Engrailed complexes is a 180–base pair sequence called the homeo box (*5–9*). The sequence of the carboxyl half of the homeo box shows homology to the helix-turn-helix domain of several DNA binding proteins that control developmental processes in bacteria and yeast, such as the λ *cro* and *c*I proteins and the yeast mating type a1 and α2 proteins (*10–13*). It has been suggested that the homeo box proteins exert their function by regulating in *trans* batteries of genes involved in cell differentiation (*10*, *11*).

Homeo box sequences have also been identified and isolated from the DNA of frogs, mice, and humans (*14–23*). The understanding of the function of mammalian homeo box–containing genes requires a detailed knowledge of where, when, and how these genes are expressed. The differential expression of several homeo box–containing genes in frog embyros, mouse embryos, and mouse adult tissues has been recently reported (*15, 18, 21–23*). Expression of homeo box–containing genes has also been found in frog oocytes (*16*) and during the differentiation of several mouse teratocarcinoma cell lines (*20–23*). We report here the identification and chromosomal assignment of a new mouse homeo box gene, MH-3, and the analysis of its expression in the adult and in the embryo.

A large complementary DNA (cDNA) library from adult mouse testis was constructed in the vector λgt10 (*24*). To ask whether homeo box sequences are expressed in the adult testis, we screened this library with mixtures of the homeo box probes Hu-1, Hu-2 (*17*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*) (*5*), *engrailed* (*en*) (*7*), and

M. R. Rubin, L. E. Toth, M. D. Patel, M. C. Nguyen-Huu, Departments of Microbiology and Urology, Columbia University, College of Physicians and Surgeons, New York, NY 10032.
P. D'Eustachio, Department of Biochemistry, New York University Medical School, New York, NY 10016.