Rao, S. T., & Rossmann, M. G. (1973) *J. Mol. Biol. 76*, 241–256.
Read, R. J. (1986) *Acta Crystallogr. A42*, 140–149.
Schreuder, H. A., Laan, J. M. van der, Thunnissen, M. M. G. M., Kalk, K. H., Swarte, M. B. A., Hol, W. G. J., & Drenth, J. (1987) in *Flavins and Flavoproteins* (Edmondson, D. E., & McCormick, D. B., Eds.) de Gruyter, New York.
Schreuder, H. A., Laan, J. M. van der, Hol, W. G. J., & Drenth, J. (1988) *J. Mol. Biol. 199*, 637–648.
Schulz, G. E., Schirmer, R. H., Sachsenheimer, W., & Pai, E. F. (1982) *J. Mol. Biol. 160*, 287–308.
Schwager, P., Bartels, K., & Jones, A. (1975) *J. Appl. Crystallogr. 8*, 275–280.
Sussman, J. L., Holbrook, S. R., Church, G. M., & Kim, S. H. (1977) *Acta Crystallogr. A33*, 800–804.
Sygush, J., Madsen, N. B., Kasvinski, P. J., & Fletterick, R. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4757–4761.
Weijer, W. J., Hofsteenge, J., Beintema, J. J., Wierenga, R. K., & Drenth, J. (1983) *Eur. J. Biochem. 133*, 109–118.
Wierenga, R. K. (1986) *Trans. Am. Crystallogr. Assoc. 22*, 49–62.
Wierenga, R. K., Jong, R. J. de, Kalk, K. H., Hol, W. G. J., & Drenth, J. (1979) *J. Mol. Biol. 131*, 55–73.
Wierenga, R. K., Kalk, K. H., Laan, J. M. van der, Drenth, J., Hofsteenge, J., Weijer, W. J., Jekel, P. J., Beintema, J. J., Müller, F., & Berkel, W. J. H. van (1982) in *Flavins and Flavoproteins* (Massey, V., & Williams, C. H., Jr., Eds.) pp 11–18, Elsevier North-Holland, New York.
Wierenga, R. K., Drenth, J., & Schultz, G. E. (1983) *J. Mol. Biol. 167*, 725–739.
Wierenga, R. K., Maeyer, M. C. H. De, & Hol, W. G. J. (1985) *Biochemistry 24*, 1346.

# Large Increases in General Stability for Subtilisin BPN′ through Incremental Changes in the Free Energy of Unfolding†

Michael W. Pantoliano,‡ Marc Whitlow, Jay F. Wood, Steven W. Dodd, Karl D. Hardman, Michele L. Rollence, and Philip N. Bryan*,§

*Genex Corporation, 16020 Industrial Drive, Gaithersburg, Maryland 20877*

*Received May 12, 1989; Revised Manuscript Received June 21, 1989*

ABSTRACT: Six individual amino acid substitutions at separate positions in the tertiary structure of subtilisin BPN′ (EC 3.4.21.14) were found to increase the stability of this enzyme, as judged by differential scanning calorimetry and decreased rates of thermal inactivation. These stabilizing changes, N218S, G169A, Y217K, M50F, Q206C, and N76D, were discovered through the use of five different investigative approaches: (1) random mutagenesis; (2) design of buried hydrophobic side groups; (3) design of electrostatic interactions at $Ca^{2+}$ binding sites; (4) sequence homology consensus; and (5) serendipity. Individually, the six amino acid substitutions increase the $\Delta G$ of unfolding between 0.3 and 1.3 kcal/mol at 58.5 °C. The combination of these six individual stabilizing mutations together into one subtilisin BPN′ molecule was found to result in approximately independent and additive increases in the $\Delta G$ of unfolding to give a net increase of 3.8 kcal/mol (58.5 °C). Thermodynamic stability was also shown to be related to resistance to irreversible inactivation, which included elevated temperatures (65 °C) or extreme alkalinity (pH 12.0). Under these denaturing conditions, the rate of inactivation of the combination variant is ~300 times slower than that of the wild-type subtilisin BPN′. A comparison of the 1.8-Å-resolution crystal structures of mutant and wild-type enzymes revealed only independent and localized structural changes around the site of the amino acid side group substitutions. Consequently, the discovery that each independent stabilizing mutation makes a nearly additive contribution to the $\Delta G$ of unfolding has made it possible to dramatically increase the stability of subtilisin BPN′ in a coherent step by step manner, consistent with the model of "localization of free energy changes" for protein folding.

Factors that contribute to the stability of proteins have been enumerated and their relative importance understood in terms of the free energy difference between the native, N, and unfolded, U, states (Schellman, 1955; Kauzmann, 1959; Tanford, 1962, 1970; Brandts, 1964). The total free energy change for the reaction N ⇔ U has been described (Tanford, 1970) by the relation

$$\Delta G = \Delta G_{conf} + \sum_i \Delta g_{i,int} + \sum_i \Delta g_{i,s} + \Delta W_{el} \qquad (1)$$

where $\Delta G_{conf}$ = configurational free energy (order/disorder term), $\Delta g_{i,int}$ = short-range interactions (H-bonds, van der Waals interactions, salt bridges, cofactor binding, etc.), $\Delta g_{i,s}$ = short-range interactions with solvent (hydrophobic effect, hydration of ions, etc.), and $\Delta W_{el}$ = long-range electrostatic interactions ($\alpha$-helix dipoles, etc.).

In this model for protein folding the symbol $\Delta g_i$ represents

† Presented in part at the Canbiocon 88 Meeting, Montreal, Canada, in April 1988, and also at the Upjohn/UCLA Symposium on Protein and Pharmaceutical Engineering, Park City, Utah, in January 1989. This research was sponsored in part by a grant from the Office of Naval Research.

* Corresponding author.
‡ Present address: E. I. du Pont de Nemours and Co., Central Research and Development, Building 328, Wilmington, DE 19880-0328.
§ Present address: Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, MD 20850.

0006-2960/89/0428-7205$01.50/0 

contributions to the corresponding $\Delta G$ that are assigned to a localized portion of a protein molecule (Tanford, 1970). That is, the total free energy of all short-range interactions within the ordered regions of a protein molecule is divided among those separate portions so that $g_{i,\text{int}}$ designates the free energy assigned to the $i$th part of the molecule. Thus the total $\Delta G$ between the states is expressed as the sum of contributions from all the physical and chemical factors that affect them.

Advances in molecular biology such as site-directed mutagenesis (Zoller & Smith, 1983) together with recent applications of random mutagenesis/genetic screening techniques (Alber & Wozniak, 1985; Shortle & Lin, 1985; Bryan et al., 1986a; Liao et al., 1986) have allowed further exploration of these physical and chemical forces that control protein folding. For example, it has been possible to rationally manipulate the $\Delta G_{\text{conf}}$ term to stabilize proteins by introducing disulfide linkages (Villafranca et al., 1983; Perry & Wetzel, 1984; Pabo & Suchanek, 1986; Pantoliano et al., 1987). Alternatively, the $\Delta G_{\text{conf}}$ term can be influenced to stabilize proteins by replacing Gly residues with Ala residues which have fewer degrees of freedom in the unfolded state (Matthews et al., 1987).

Similarly, the $\Delta g_{i,s}$ term of eq 1 has been logically explored for stability (Matsumura et al., 1988; Kellis et al., 1988) through changes in buried hydrophobic surface areas, although the majority of these changes were designed to be destabilizing. To date, the most successful examples of increased stabilization due to alterations in the short-range interactions in proteins, whether they involve solvent ($\Delta g_{i,s}$) or folded regions of the native structure ($\Delta g_{i,\text{int}}$), have come through the random mutagenesis/genetic screening approach (Bryan et al., 1986a; Das et al., 1989). The low success rate for designed stabilizing short-range interactions presumably reflects the unforgiving potential energy functions that are very sensitive to interatomic distances and geometry (Levitt, 1974; Brooks et al., 1983).

On the other hand, long-range electrostatic interactions have been found to be more easily amendable through logical molecular modeling approaches that are less sensitive to atomic geometry and, therefore, more forgiving of oversights. Examples of where $\Delta W_{\text{el}}$ has been successfully altered for the increased stabilization of proteins are found for $\alpha$-helix dipoles (Mitchinson & Baldwin, 1986; Nicholson et al., 1988) and $Ca^{2+}$ binding sites (Pantoliano et al., 1988).

We have continued our study of the physical and chemical factors that impart stability to proteins and have identified six new amino acid substitutions that independently *increase* the stability of subtilisin BPN′. This paper will focus on the physical characterization of the stability of each of these changes, with particular consideration of the dramatic stabilization acquired when all six mutations exist together. Moreover, a preliminary X-ray structural characterization of each stabilizing side-chain substitution will also be discussed.

## Materials and Methods

*Molecular Biology.* The subtilisin gene from *Bacillus amyloliquefaciens* (subtilisin BPN′) has been cloned, sequenced, and expressed at high levels from its natural promoter sequences in *Bacillus subtilis* (Wells et al., 1983; Vasantha et al., 1984). All mutant genes were recloned into a pUB110-based expression plasmid (Zoller & Smith, 1983; Bryan et al., 1986b) and used to transform *B. subtilis*. The *B. subtilis* strain used as the host contains a chromosomal deletion of its subtilisin gene and therefore produces no background wild-type (wt)[1] activity (Fahnestock & Fisher, 1987).

Random mutagenesis was employed for the identification of the stabilizing mutation, N218S, as previously described (Bryan et al., 1986a; Rollence et al., 1988).

*Molecular Modeling.* In addition to creating stabilizing mutations through random mutagenesis, an attempt was made to design stabilizing amino acid substitutions by rational approaches that exploit one or more of the terms in eq 1. Five different investigative approaches were undertaken, all of which successfully identified a stabilizing mutation, but *not* always for the reasons expected:

(*1*) *Stabilization through the Introduction of Buried Hydrophobic Side Chains.* In order to introduce buried methyl groups into the tertiary structure of subtilisin BPN′, we identified several Gly and Val residues that have low solvent accessibilities, as judged from a high-resolution X-ray structure of wt subtilisin BPN′ (Pantoliano et al., 1988). The residues were then prioritized by comparing primary sequences of homologous subtilisin-like enzymes (Meloun et al., 1985). Those residues that occur as Ala or Ile in related subtilisins in place of the wt Gly or Val, respectively, were given greater preference over conserved residues. This approach selected residue Gly 169 as an obvious candidate for site-directed mutagenesis; its solvent-accessible surface area is only 1%, as measured by water probe calculations (Richards, 1977), and four homologous subtilisins (amylosaccharicus, Carlsberg, DY, and thermitase) have Ala at this position.

(*2*) *Stabilization through Sequence Homology Consensus.* Similarly, one can compare the primary sequences of the four homologous subtilisins, as above, and look for positions in which subtilisin BPN′ is singularly divergent, without regard to any preconceived theory. This simple approach selected Met 50 as a candidate for site-directed mutagenesis since four homologous enzymes have Phe at this position. The assumption here is that the consensus amino acid at any one position has stood the test of evolution and could possibly be optimized for the task at hand.[2]

(*3*) *Stabilization by Influencing Binding Affinity at Calcium Ion Binding Sites.* It was previously reported (Pantoliano et al., 1988) that subtilisin BPN′ can be stabilized through engineered electrostatic interactions that incrementally increase the binding affinity at the $Ca^{2+}$ binding sites, as expected from the interrelation between thermostability and cofactor binding (Schellman, 1975, 1976). By use of Coulomb's law, $\Delta pK_a = 244Z_1Z_2/rD_{\text{eff}}$, as a rudimentary guide, a three-step algorithm was implemented to target amino acid residues in the vicinity of the $Ca^{2+}$ binding sites for mutagenesis to Asp or Glu. This selection process employed a high-resolution X-ray structure of wt subtilisin BPN′ (Pantoliano et al., 1988) and resulted in the choice of Asn 76 as a prime candidate for mutagenesis to Asp.[3]

[1] Abbreviations: $C_{\text{ex}}$, excess specific heat as measured by calorimetry; $C_{\text{max}}$, maximum excess specific heat; $\Delta H_{\text{cal}}$, calorimetric enthalpy for unfolding; $\Delta H_{\text{vH}}$, van't Hoff enthalpy for unfolding; Dns-BBA, *N*-dansyl-3-aminobenzeneboronic acid; DSC, differential scanning calorimetry; EDTA, disodium salt of ethylenediaminetetraacetic acid; [E], enzyme concentration; Hepes, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; Mes, 2-(*N*-morpholino)ethanesulfonic acid; rms, root mean square; sAAPFna, succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide; Tris, tris(hydroxymethyl)aminomethane; $t_{1/2}$, half-life for a kinetic experiment; wt, wild type. A shorthand for denoting amino acid substitutions employs the single letter amino acid code as follows (Knowles, 1987): Y217K denotes the change of Tyr 217 to Lys, and the variant enzyme containing this single change is denoted subtilisin Y217K.

[2] Subsequent to our initiation of this approach, a random mutagenesis study identified M50F as a stabilizing mutation (Cunningham & Wells, 1987).

(4) *Attempted Stabilization through the Introduction of Disulfide Linkages.* By use of a computer-based method that has been previously described (Pantoliano et al., 1987), two target residues, Gln 206 and Ala 216, were selected for mutagenesis to Cys residues for the introduction of a potentially stabilizing disulfide linkage. The Q206C/A216C changes *did* result in a stabilizing disulfide linkage; the unfolding transition for this variant was found to be 3.0 °C higher than that for the wt protein, as measured by DSC (unpublished results). However, analysis of the single-site variants where Q206C and A216C exist alone revealed that the Q206C mutation is even more stabilizing than the disulfide linkage (vide infra).

(5) *Introduction of Electrostatic Interactions at the Active Site.* Independent of the efforts to introduce stabilizing interactions into subtilisin BPN′, an attempt was made to introduce electrostatic interactions near the catalytic His 64 in order to shift the pH optimum for catalysis from near pH 9 toward neutral pH. After careful analysis of computer graphics simulations with the program FRODO (Jones, 1978), the Y217K mutation was chosen to introduce a positive charge within 7 Å of the catalytic His 64. This approach is similar to that described by others (Matthew, 1985; Russell & Fersht, 1987; Gilson & Honig, 1988).

*Differential Scanning Calorimetry (DSC).* DSC measurements were performed with a Hart 7707 DSC heat conduction scanning microcalorimeter consisting of two matched pairs of removable ampules and interfaced with an IBM personal computer. The temperature was increased from 20 to 105 °C at a scan rate of 60 °C/h for the majority of experiments, but at 16, 40, and 100 °C/h for a few selected studies. The solution mass of all protein and control solutions was near 0.70 g per ampule. Each experiment was comprised of four segments: (1) the first upward scan from 20 to 105 °C; (2) the first downward scan from 105 to 20 °C; (3) a second upward scan from 20 to 105 °C; and (4) a second downward scan to return to 20 °C. Since the thermal unfolding transitions of subtilisin BPN′ were previously observed to be irreversible (Takahashi & Sturtevant, 1981; Bryan et al., 1986a; Pantoliano et al., 1987) under the conditions of DSC scans, the power input from the second upward scan was subtracted from the first upward scan to obtain the excess power input for the unfolding transitions. The excess power thermal scans were converted to excess heat capacity vs $T$ scans by dividing by the scan rate (Schwarz & Kirchoff, 1988). The number of nanomoles of protein ranged from 70 to 100.

A sigmoidal base line, $C_{ps}$, was extrapolated under the thermal transitions by using a least-squares computer fit of the pre- and posttransitional base lines to linear equations in $T$ and $\alpha$, the fractional area under the transition curve at any particular $T$, as described by Schwarz (1988).[4] The calorimetric enthalpy of unfolding, $\Delta H_{cal}$, was determined from the area under the transition curve and the [E] in the ampule. The unfolding transition temperature, $T_m$, was measured as the temperature at $\alpha = 0.5$. The change in heat capacity upon unfolding, $\Delta C_p$, was determined from the difference between the extrapolated pre- and posttransitional base lines at $T_m$ and the [E] in the ampule.

*Protein Purification and Characterization.* Wild-type subtilisin BPN′ and the variant enzymes were purified and verified for homogeneity essentially as previously described (Bryan et al., 1986b; Pantoliano et al., 1987). Assays were performed by monitoring the hydrolysis of succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (sAAPFna) as described by DelMar et al. (1979). The [E] was determined by using $E^{0.1\%} = 1.17$ at 280 nm (Ottesen & Svendsen, 1970), which compares reasonably well with $E^{0.1\%} = 1.20$ ($\epsilon = 33\,060$ $M^{-1}$ $cm^{-1}$) calculated by using $\epsilon = 1413$ $M^{-1}$ $cm^{-1}$ for Tyr (×10) and $\epsilon = 6310$ $M^{-1}$ $cm^{-1}$ for Trp (×3) at 280 nm (*Handbook of Chemistry and Physics*, 1974–1975). For variants that contain the Y217K change, the $E^{0.1\%}$ at 280 nm was calculated to be 1.15 (or 0.96 × wt) on the basis of the loss of one Tyr residue.

*X-ray Crystallography.* Large single crystal growth and X-ray diffraction data collection were performed essentially as previously reported (Bryan et al., 1986a; Howard et al., 1987; Pantoliano et al., 1988) except that it was not necessary to inactivate all the variants with diisopropyl fluorophosphate (DFP) in order to obtain suitable crystals.

## Results and Discussion

*Calorimetry.* The thermal unfolding transitions of variant and wt enzymes were studied by DSC. A typical excess heat capacity scan of the unfolding transition for wt subtilisin BPN′ in 10 mM EDTA, pH 8.0 (5.0 mM Tris-HCl, 50 mM KCl), is shown in Figure 1. This single symmetrical transition for wt subtilisin BPN′ was found to have a $T_m$ of 58.5 ± 0.1 °C and is believed to correspond to the form of the enzyme that binds no $Ca^{2+}$. Under the conditions of the DSC experiments, the pCa = 10.4, as estimated by the equation (Tsalkova & Privalov, 1985):

$$[Ca^{2+}]_{free} = [Ca]_0/([EDTA] - [Ca]_0)K_{Ca}$$

where $[Ca]_0$ is the concentration of $Ca^{2+}$ (same as [E] in this case since subtilisin BPN′ binds 1.0 Ca/mol as isolated from the fermentation broth; Pantoliano et al., 1988), [EDTA] = 10 mM, and $K_{Ca}$ is the $Ca^{2+}$ binding constant of EDTA, which was taken to be $10^{8.4}$ $M^{-1}$ at pH 8.0 (Schwarzenbach & Flaschka, 1965). The binding affinity of the Ca A site was estimated by performing DSC experiments as a function of $[Ca^{2+}]_{free}$ through manipulation of this 10 mM EDTA/Ca buffer. The $T_m$ of wt subtilisin BPN′ remains unchanged between pCa = 10.4 and 9.0, but at lower pCa (higher $[Ca^{2+}]_{free}$) the transition shifts to higher $T$ until pCa < 7.0, where the $T_m$ is near 70 °C, as was observed for the protein that binds 1.0 mol of Ca/mol of enzyme (Pantoliano et al., 1988). The nature of this shift in $T_m$ is similar to that observed for single domains of troponin C and calmodulin (Tsalkova & Privalov, 1985) except that the midpoint for the shift in $T_m$ for subtilisin occurs near a pCa of 8.0. Hence, our estimate for the apparent $pK_a$ for $Ca^{2+}$ binding to the Ca A site of subtilisin BPN′ is near 8.0 (at 65 °C and 50 mM Tris-HCl, pH 8.0, 50 mM KCl).

For the purpose of this study, however, it is sufficient to note that the $[Ca^{2+}]_{free}$ for all DSC scans reported here is well below the estimated $pK_a$ for $Ca^{2+}$ binding to the wild-type-like Ca A site. Therefore, the transitions observed in 10 mM EDTA are considered to be a measure of the intrinsic stability of the various forms of this enzyme and independent of metal ion mediated stabilization (Voordouw et al., 1976).

The amount of excess heat absorbed by a protein sample as $T$ is increased through a transition from the folded to unfolded state at constant pressure provides a direct measurement of the $\Delta H$ of unfolding (Privalov & Potekhin, 1986). With subtilisin, however, the autolysis that accompanies the unfolding process complicates this analysis (Takahashi &

[3] Subtilisin BPN′ contains two $Ca^{2+}$ binding sites, the high-affinity Ca A site (log $K_a \sim 8$), and the weak Ca B site (log $K_a \sim 2$). For a biophysical and crystallographic description of these sites, see Pantoliano et al. (1988).

[4] DSC curve fitting was done with the aid of a computer program written by F. Schwarz and W. Kirchhoff of the National Institute of Standards and Technology.

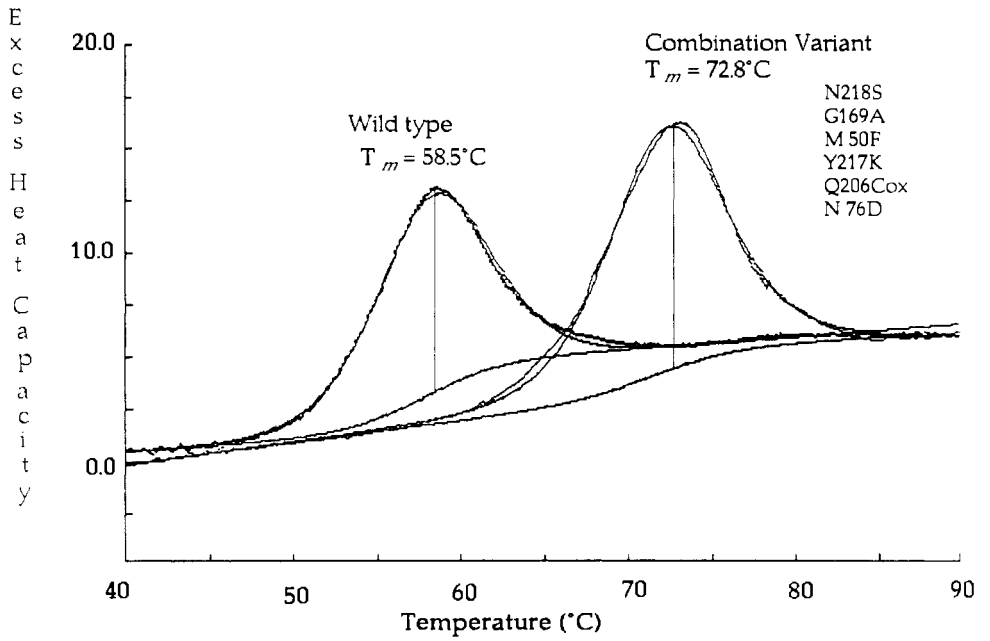


FIGURE 1: Differential scanning calorimetry. The calorimetric scans and a theoretical two-state fit are shown for wt ($T_m$ = 58.5 °C) and the combination variant ($T_m$ = 72.8 °C) of subtilisin BPN′. The experimentally measured excess heat capacity was fit, in each case, to a computer-derived theoretical curve that simulates a two-state unfolding process not involving association or dissociation (Schwarz & Kirchhoff, 1988). Excess heat capacity is in units of kcal/(deg mol).

Table I: Experimental Thermodynamic Data for the Thermal Unfolding of Subtilisin BPN′ and the Combination Variant

| protein | $T_m$[a] (°C) | $\Delta H_{cal}$ (kcal/mol) | $\Delta H_{vH}(1)$[b] (kcal/mol) | $\Delta H_{vH}(1)/\Delta H_{cal}$ | $\Delta S$ [cal/(deg mol)] | $\Delta C_p$ [kcal/(deg mol)] | $C_{max}$ [kcal/(deg mol)] |
|---|---|---|---|---|---|---|---|
| wild type | 58.5 ± 0.1 | 88.3 ± 5 | 97 ± 4 | 1.10 ± 0.09 | 265 ± 15 | 4.8 ± 0.4 | 9.6 |
| combination variant | 72.8 ± 0.2 | 105 ± 0.2 | 102 ± 0.5 | 0.97 ± 0.01 | 304 ± 1 | 3.9 ± 0.6 | 11.2 |

[a] $T_m$ is the temperature at which the excess heat capacity reaches its maximum value, $C_{max}$. All thermodynamic values are for $T_m$, where $T_m$(K) = $T_m$(°C) + 273.15. Uncertainties represent two standard deviations for the data from three independent DSC scans. [b] $\Delta H_{vH}(1) = ART_m^2C_{max}/\Delta H_{cal}$, where $A$ = 4.00, the factor appropriate for a simple two-state process, N ⇔ U, not involving association or dissociation (Manly et al., 1985).

Sturtevant, 1981). In the studies reported here, an attempt to circumvent this problem was made by addition of the competitive inhibitor *N*-dansyl-3-aminobenzeneboronic acid (Dns-BBA), which has a $K_i$ = 2 μM at pH 8.0 (Philipp & Maripuri, 1981). In spite of these precautions, the reaction appears to be irreversible by the usual DSC criterion of the absence of an unfolding transition after a rescan of the denatured protein. It has been recently shown, however, that the DSC-measured transitions for the thermal denaturation of several proteins can be interpreted in terms of the van't Hoff equation, in spite of calorimetric irreversibility (Manly et al., 1985; Edge et al., 1985; Hu & Sturtevant, 1987). This has been explained in terms of an irreversible change of the unfolded state, with little heat effect, at temperatures above that of the calorimetric transition. If the thermal unfolding of subtilisin in the calorimeter can be represented as (Lumry & Eyring, 1954)

$$N \underset{k_2}{\overset{k_1}{\rightleftharpoons}} U \xrightarrow{k_3} I \qquad (2)$$

then the equilibrium between the folded and unfolded states will be approximated when $k_3$ is small relative to the time required to complete a calorimetric unfolding transition (Privalov & Potekhin, 1986). In this case, the unfolding transition should fit a two-state model and follow the dictates of equilibrium thermodynamics as expressed in the van't Hoff equation, $d \ln K/dT = \Delta H_{vH}/(RT^2)$. In this situation, $\Delta H_{vH}$, the van't Hoff enthalpy, or apparent enthalpy, should equal the calorimetric, $\Delta H_{cal}$, or true enthalpy.

Analysis of the experimental thermodynamic data for the thermal unfolding transition of subtilisin BPN′ reveals a reasonably close agreement between the $\Delta H_{cal}$ and $\Delta H_{vH}(1)$, which was calculated from the heat capacity data (see Table I). The agreement was found to be even closer for the variant enzyme which contains six stabilizing mutations. More convincing evidence for a two-state model to explain the thermal unfolding of subtilisin under the stated conditions was found upon application of least-squares curve fitting. Thermodynamic parameters derived from curve fitting of the experimental DSC scans are summarized in Table II. A computer program developed by Schwarz and Kirchhoff (1988) was used to fit a theoretical curve to the experimental data by varying the adjustable parameters $T_m$ and $\Delta H_{vH}(2)$ to minimize the standard deviation of the observed values of excess heat capacity, $C_{ex}$, from the calculated curve in much the same way as described by Takahashi and Sturtevant (1981). The fitted curves in Figure 1 are nearly superimposable on the experimental DSC scans and were obtained by assuming a simple two-state model for unfolding, N ⇔ U. Indeed, the $\Delta H_{vH}(2)$'s derived from the two-state model fit agree reasonably well with the calorimetrically derived $\Delta H_{cal}$'s (Table II).

Further support for this model was obtained by varying the scan rate. At faster scan rates (~100 °C/h) the $\Delta H_{cal}$ and $\Delta H_{vH}$(1 or 2) were found to be nearly identical with those found at 60 °C/h (Table I). But if the scan rate was slowed to ~16 °C/h, then the $\Delta H_{vH}/\Delta H_{cal}$ ratio was found to increase to >2.0, consistent with the model (eq 2). When the scan rate becomes slow enough so that $k_3$ is significant relative to the time required to complete the unfolding transition, then

Table II: Thermodynamic Data Derived from Curve Fitting of the Thermal Unfolding of Subtilisin BPN′ and the Combination Variant

| protein | $T_m$ (°C) | $\Delta H_{vH}(2)$[a] (kcal/mol) | $\Delta S$ [cal/(deg mol)] | SD[b] (% $C_{max}$) | $\Delta H_{vH}(2)/\Delta H_{cal}$ |
|---|---|---|---|---|---|
| wild type | 57.2 ± 0.1 | 94 ± 3 | 285 ± 9 | 1.6 ± 0.9 | 1.07 ± 0.09 |
| combination variant | 71.6 ± 0.2 | 101 ± 1.0 | 293 ± 3 | 2.9 ± 0.3 | 0.96 ± 0.01 |

[a] $\Delta H_{vH}(2)$ is obtained by minimization of the standard deviation of a theoretical curve that assumes a simple two-state model, N ⇔ U, from the observed experimental data. All thermodynamic parameters are at $T_m$, where $T_m(K) = T_m(°C) + 273.15$. Uncertainties represent two standard deviations. [b] SD, standard deviation of the fitted curve from the experimental data.

Table III: Free Energy Changes for Thermostable Mutations in Subtilisin BPN′

| mutation(s) | $T_m$[a] (°C) | $\Delta T_m$[b] (°C) | $\Delta\Delta G$[c] (kcal/mol) | $\Delta\Delta G$ (sum)[d] |
|---|---|---|---|---|
| wild type | 58.5 ± 0.1 | 0.0 | 0.0 | |
| N218S | 62.5 ± 0.2 | 4.0 | 1.07 ± 0.12 | |
| G169A | 59.6 ± 0.2 | 1.1 | 0.30 ± 0.07 | |
| M50F | 60.3 ± 0.2 | 1.8 | 0.48 ± 0.08 | |
| Y217K | 61.2 ± 0.2 | 2.7 | 0.72 ± 0.09 | |
| Q206Cox | 63.2 ± 0.2 | 4.7 | 1.25 ± 0.12 | |
| N76D | 60.2 ± 0.2 | 1.7 | 0.45 ± 0.08 | |
| in combination | | | | |
| N218S | 72.8 ± 0.2 | 14.3 | 3.80 ± 0.27 | 4.27 ± 0.25 |
| G169A | | | (4.35 ± 0.08)[e] | |
| M50F | | | | |
| Y217K | | | | |
| Q206Cox | | | | |
| N76D | | | | |

[a] All proteins were dissolved in 50 mM Tris-HCl (pH 8.0), 50 mM KCl, 10 mM EDTA, and 2 mM Dns-BBA. [b] Increase in $T_m$ relative to the wt enzyme. [c] $\Delta\Delta G$ is relative to that of wild type at 58.8 °C and is calculated from the relation $\Delta\Delta G = \Delta T_m(\Delta S_{wt})$ (Becktel & Schellman, 1987); $\Delta S_{wt}$ = 265 ± 15 cal/(deg mol) (Table I). Alternatively, the Gibbs–Helmholtz equation, $\Delta G°(T_2)/T_2 - \Delta G°(T_1)/T_1 = \Delta H(T_2^{-1} - T_1^{-1})$, may be used to obtain $\Delta\Delta G$ relative to the wild-type protein at $T_m$. The results are nearly identical. [d] Represents the sum of the individual $\Delta\Delta G$'s measured separately in column 4. [e] Value in parentheses was calculated by using $\Delta S_{mut}$ = 304 ± 1 cal/(deg mol) (Table I).

the reaction U → I begins to diminish the concentration of participating molecules in the equilibrium N ⇔ U, leading to a sharpening of the observed thermal unfolding curve (Privalov & Potekhin, 1986). Under these circumstances the thermal denaturation can be interpreted in terms of a kinetic process, N $\xrightarrow{k}$ I. Plots of ln $k$ versus $1/T$ [where $k = vC_{ex}/(Q_t - Q)$; $v$ = scan rate (°C/min), $C_{ex}$ = excess heat capacity, $Q_t$ = total heat of the process, and $Q$ = heat evolved at any given $T$] were found to be linear at this slow scan rate of 16 °C/min. However, attempts to fit the data collected at 60 and 100 °C/h to this kinetic model proved unproductive. Similarly, attempts to fit the experimental DSC data of Figure 1 to various kinetically controlled models of unfolding as recently applied to thermolysin (Sanchez-Ruiz et al., 1988) also proved to be unproductive for subtilisin BPN′ and its variants under the conditions employed. The experimental evidence, therefore, strongly suggests that the thermal unfolding of subtilisin BPN′ and its variants follows equilibrium thermodynamics as embodied by the van't Hoff equation (Privalov & Potekhin, 1986) during the course of these DSC experiments.

In this regard, it was possible to use the DSC data for the six stabilizing mutations to determine the contribution that each amino acid substitution makes toward the total free energy change for thermal unfolding, either separately or in combination. The $T_m$ and $\Delta\Delta G$ for all single and combination mutations are summarized in Table III. Comparison of the $\Delta\Delta G$ of unfolding for individual mutations with that of the combination variant reveals that the free energy changes associated with each individual change accrue in a nearly additive fashion when they are combined into the same molecule.

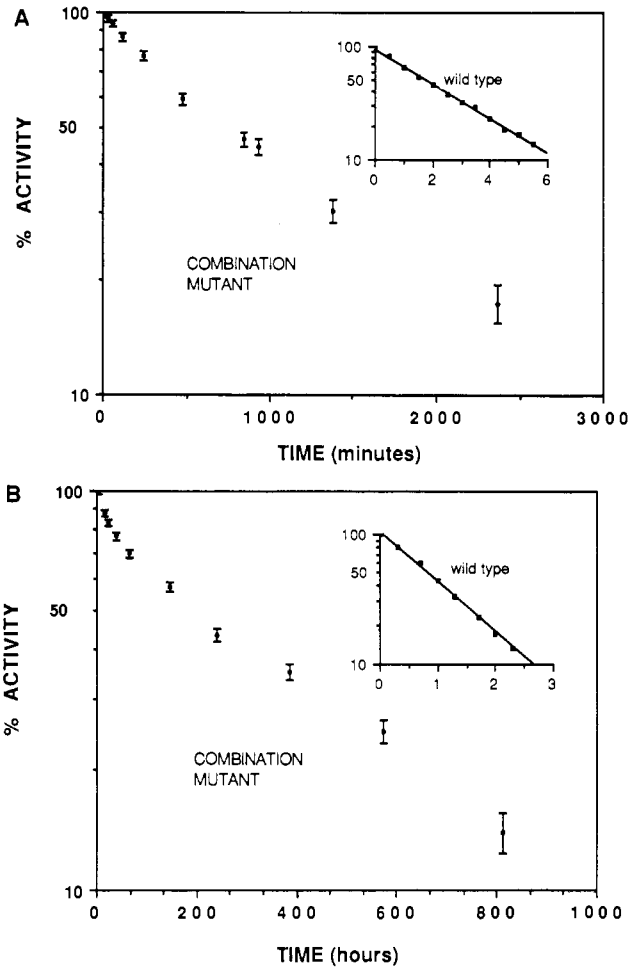


FIGURE 2: Kinetics of inactivation. The loss of initial activity as a function of time is plotted for wt subtilisin BPN′ and the combination variant under the following conditions: (Panel A) $T$ = 65 °C, pH = 8.0 (50 mM Tris-HCl, 50 mM KCl). Initial [E] = 100 μg/mL (3.6 μM). (Panel B) $T$ = 25 °C, pH = 12.0 (100 mM potassium phosphate). Initial [E] = 100 μg/mL (3.6 μM). The wt data are plotted in the inset graphs on an expanded time scale. Each data point is the average of three independent determinations.

*Kinetics of Irreversible Inactivation.* The mechanism by which a protease such as subtilisin is irreversibly inactivated may be complex, possibly involving not only unfolding but also aggregation, autodigestion, and chemical alteration of certain amino acids (deamidation). The extent to which thermodynamic stability protects subtilisin BPN′ from irreversible inactivation was determined by measuring the rates of inactivation under a variety of extreme denaturing conditions that included elevated temperature (65 °C) and extreme alkalinity (pH 12, 25 °C). The results at 65 °C for wt subtilisin and the variant containing six mutations are shown in Figure 2A. The results for these same proteins at pH 12 (25 °C) are shown in Figure 2B.

At [E] = 100 μg/mL (3.6 μM), the rate of inactivation of wt subtilisin BPN′ appears to be pseudo first order with regard to [E] over four $t_{1/2}$'s, either at elevated temperature (65 °C,

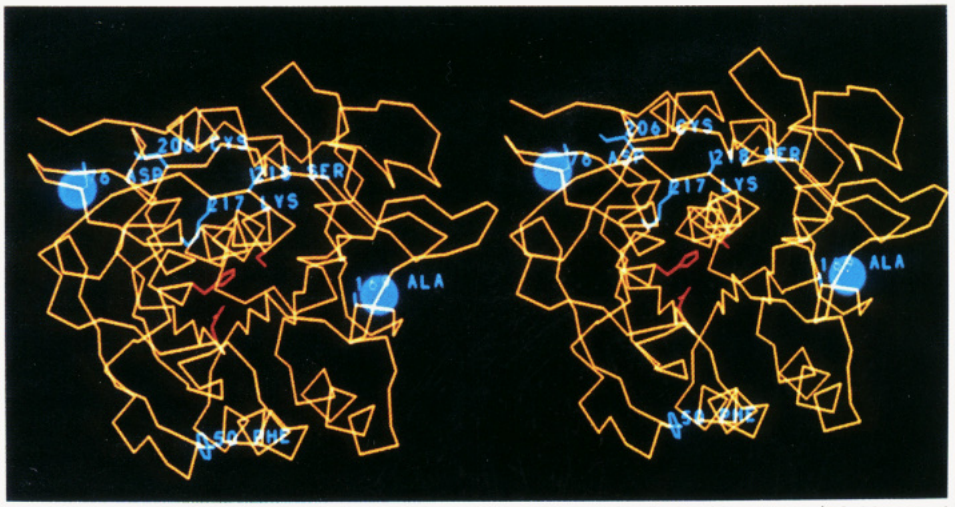


FIGURE 3: X-ray crystal structure of the combination variant of subtilisin BPN′. The $C_\alpha$ backbone trace of this protein is shown in yellow. The six amino acid mutations are depicted in blue. The van der Waals radii of the two calcium ions are outlined as blue dot surfaces. The high-affinity Ca A site is on the left, and the weak Ca B site is on the right. The residues that comprise the catalytic triad, His 64, Ser 221, and Asp 32 (Kraut, 1971), are shown in red. The structure was determined at 1.8-Å resolution and refined to $R = 0.15$. $R = \sum|F_o - F_c|/\sum F_o$, where $F_o$ and $F_c$ are the observed and calculated structure factor amplitudes, respectively.

pH 8.0) or at extreme pH (pH 12, 25 °C). Under these same conditions, the loss of activity of the combination mutant is not strictly first order, suggesting that the inactivation is not completely determined by the rate of unfolding. Nevertheless, the rate of inactivation of the thermodynamically stable variant is ~300 times slower than that of the wt enzyme under these conditions. Resistance to irreversible inactivation, therefore, appears to correlate with increased thermodynamic stability under many different conditions of extreme denaturation. In addition, it was also observed that the combination variant undergoes inactivation in anhydrous organic solvents [dimethylformamide (DMF)] at rates that are ~50 times slower than that of the wt enzyme (Wong, 1989; Chen et al., 1989).

*X-ray Crystal Structures.* Structural characterization of these six stabilizing amino acid substitutions was possible after single crystals of the various derivatives of subtilisin BPN′ containing these changes were grown and then subjected to high-resolution X-ray diffraction techniques. One such X-ray structure was that of the combination variant which contains all six changes simultaneously and was solved to 1.8-Å resolution and refined to an $R = 0.15$ (see Figure 3). A preliminary description of each stabilizing change and the short-range interactions and contacts it makes with neighboring atoms in the tertiary structure and/or with solvent is outlined below. A more detailed crystallographic analysis will appear elsewhere.

(*1*) *N218S.* The first stabilizing mutation N218S was discovered through random mutagenesis and appears to improve two or three H-bonds in the vicinity of a β-bulge as previously described (Bryan et al., 1986a). A detailed structural comparison of the N218S variant and wt revealed that these two enzymes are superimposable to 0.07 Å for $C_\alpha$ positions and 0.10 Å for all atoms. The only region deviating from these values by more than 2-fold was in the immediate vicinity of residue 218 (Figure 4). Moreover, this highly localized change in the X-ray structure for subtilisin N218S was also superimposable on the X-ray structure for the combination variant. The results plotted in Figure 4 are for the N218S mutation alone because this mutation causes the largest observed structural movements of any of the six mutations and should therefore reflect the upper limit of atomic displacement for the entire group.

(*2*) *Y217K.* The structural explanation for increased stability achieved with the Y217K mutation is also believed to involve improved H-bonding energetics. The Y217K mutation was found to result in a H-bond donation from the $\epsilon$-$NH_3$ group of Lys to the $O^\gamma$ of Ser 63 ($N^\epsilon$–$O^\gamma$ distance = 2.8 Å). This new H-bond replaces an already existing H-bond between Ser 63 and the phenolic hydroxyl group of Tyr 217 in the wt structure (Tyr 217 $O^\eta$–$O^\gamma$ Ser 63 distance = 2.55 Å). The latter H-bond distance of the wt protein is shorter than that expected for an optimum H-bond (Baker & Hubbard, 1984; Creighton, 1984). Presumably, the less rigid Lys can more easily satisfy the geometric restraints of an ideal H-bond and yield a more stable interaction. Further exploration of the mechanism of stabilization at this site will be possible after other mutations are attempted.

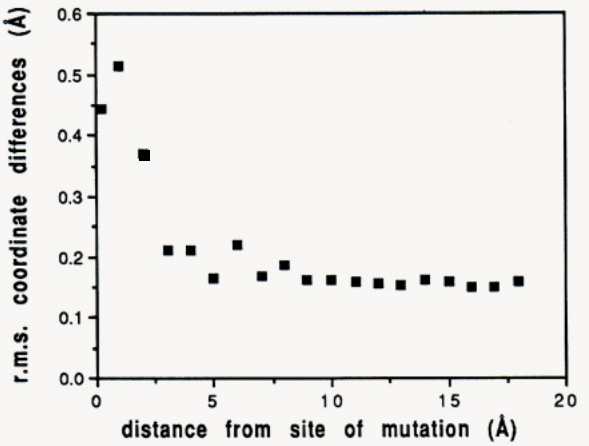


FIGURE 4: Atomic displacement as a function of distance from site of mutation. The rms coordinate differences between the positions of $C_\alpha$'s in the wt structure and the N218S mutant structure are plotted as a function of distance from the site of amino acid substitution. $C_\alpha$'s are grouped into shells of 1.0-Å thickness radiating from the $C_\alpha$ of 218.

This mutation was not designed to improve stability but to shift the $pK_a$ of His 64 and yield an enzyme with a lower pH optimum. This change does in fact shift the pH optimum of subtilisin BPN′ ~1.0 log unit toward lower pH as expected (unpublished results), but the stabilization of this change is a bonus and a clear case of serendipity.

(*3*) *Q206Cox.* The cysteine at position 206 is of special

interest because it is spontaneously oxidized during fermentation. The major species recovered upon purification appeared to be cysteine persulfide, Cys–S–SH or Cys–S–S$^-$, a form of cysteine sometimes found in naturally occurring proteins (Finazzi Agro et al., 1972; Ploegman et al., 1979; Briggs & Fee, 1978). The persulfide nature of this residue was deduced from the large extra electron density adjacent to the $S^\gamma$ atom of Cys 206 and interpreted as an extra sulfur atom, $S^\delta$. This interpretation is consistent with some of the chemical properties of this mutation. For example, when the Q206C variant is purified in the presence of 10 mM DTT, the reduced free thiol form of cysteine is obtained and observed to be slightly less stable than the wt enzyme. The oxidized form of the Cys 206 mutant (Q206Cox) that is isolated from the fermentation broth, however, is significantly more stable than the wt Gln (Table III) or any of the several natural amino acid substitutions examined to date. In addition, subtilisin BPN′ variants containing the Q206Cox change have a low-intensity absorption band near 330 nm, similar to those observed for persulfide groups in other proteins (Finazzi Agro et al., 1972; Briggs & Fee, 1978).

The stabilization acquired with the persulfide form appears to be partly due to a preferential van der Waals packing around the extra sulfur atom. Favorable van der Waals contacts (average interatomic distance of ~4 Å) are made with eight neighboring atoms: $C^\beta$ 214, C 214, N 215, $C^\alpha$ 215, C 215, O 215, N 216, and N 207. Apparently, the more polarizable extra sulfur atom can derive a greater free energy benefit from these interactions than can the $C^\delta$ atom of the wt Gln (or the H atom of the reduced Cys form obtained in the presence of 10 mM DTT) as expected from the Slater–Kirkwood equation for estimating dispersion energies between atoms (Fersht, 1985). The protein nearest neighbors to the extra sulfur atom comprise a shallow niche just below the protein–solvent interface, so that a hydrophobic contribution ($\Delta g_{i,s}$) to stability cannot be ruled out at this writing. These results illustrate how site-directed mutagenesis can be combined with chemical modification to improve a property of a protein through the introduction of an unusual amino acid, as was anticipated by Knowles (1987).

(*4*) *G169A*. The G169A mutation was shown to introduce a methyl group into a hydrophobic cavity buried below the protein–solvent interface without any other changes in the protein structure. A hydrophobic contribution ($\Delta g_{i,s}$) is, therefore, the most likely explanation for the 0.3 kcal/mol increase in $\Delta G$ of unfolding acquired by this mutation, although a $\Delta G_{conf}$ effect could also explain the data (Matthews et al., 1987).

(*5*) *M50F*. The stabilizing effect of the M50F mutation can also be explained in terms of a hydrophobic effect since only 10% of the surface area of the side chain of Met 50 is accessible to solvent, as measured through water probe calculations (Richards, 1977). The introduction of the phenyl group fills this shallow hydrophobic niche just below the surface of the protein to a greater extent than does the straight-chain aliphatic side chain of Met. Because of the shallow nature of this hydrophobic niche the phenyl group is also slightly exposed to solvent, so that the full benefit of burying Phe vs Met is not expected ($\Delta\Delta G_{tr}$ = 1.2 kcal/mol; Nozaki & Tanford, 1971). The observed 0.5 kcal/mol increase in stability is therefore consistent with the partial burying of a more hydrophobic side chain in the tertiary structure of the folded protein.

(*6*) *N76D*. The N76D mutation was the most conservative change since it is nearly indistinguishable from the wt structure. The increase in stability gained through this mutation can be described by a combination of two effects. The intended effect was to increase the binding affinity at the Ca A site by introducing an attractive electrostatic interaction 6.1 Å from the location of the bound $Ca^{2+}$. The $\Delta pK_a$ ($K_a$ = $Ca^{2+}$ association constant) expected for this mutation is in the range of 0.6–1.0 ($I$ = 0.1) depending on what $D_{eff}$ is operative in this region of the protein (Gilson & Honig, 1988; Pantoliano et al., 1988). This magnitude of increased binding affinity allows the Ca A site to compete more effectively with EDTA for $Ca^{2+}$ under the conditions of these experiments. The results for $Ca^{2+}$ binding will be described in another report.

In the absence of improved $Ca^{2+}$ binding, however, the N76D mutation is expected to enhance an already existing H-bond between the $O^{\delta 1}$ atom of Asn 76 and the $N^{\epsilon 2}$ atom of Gln 2 (Asp 76 $O^{\delta 1}$–$N^{\epsilon 2}$ Gln 2 distance is 2.9 Å). Asp makes a better H-bond acceptor than Asn because of its formal negative charge, as was previously demonstrated for the N218D mutation (Bryan et al., 1986a).

Each of these mutations needs to be further explored through additional efforts in mutagenesis, biophysics, and X-ray crystallography to more fully understand the physical and chemical forces involved in the observed increased stability. Information concerning changes in solvation of either the native or denatured state upon mutagenesis (Shortle et al., 1988; Gao et al., 1989) should also be assembled to support or refute any of the rationalizations given above for these six stabilizing mutations. Additional support for the structure/stability correlations described above will be forthcoming.

## Conclusions

The six stabilizing mutations reported here appear to contribute to the $\Delta G$ of unfolding in a variety of ways that include improved H-bonding, van der Waals contacts, burying of hydrophobic side groups, and long-range electrostatic interactions at $Ca^{2+}$ binding sites. Each mutation was found to significantly increase the $\Delta G$ of unfolding and increase resistance to irreversible thermal inactivation. Combining these individual stabilizing mutations together into one protein resulted in a 14.3 °C increase in the midpoint of the thermal unfolding transition and an ~300-fold decrease in the rate of thermal inactivation under a variety of conditions.[5] Calorimetric data demonstrate that independent, incremental increases in the free energy of unfolding resulting from combining six individual stabilizing mutations yielded approximately additive increases in overall stability. Similar results for combinations of two or three mutations have also been recently reported (Matsumura et al., 1986; Hecht et al., 1986; Pantoliano et al., 1988; Rollence et al., 1988; Nicholson et al., 1988; Stearman et al., 1988).

The crystallographic comparisons reported here demonstrate that on a structural level the effects of these mutations are independent and highly localized at least as far as can be determined by the 1.8-Å X-ray crystal structures. Even the two substitutions at residues 217 and 218 do not result in any concerted change in the peptide backbone. Because of the subtlety of the structural changes associated with each of these modifications, the basic structure of even the most altered (and

[5] Most of the stabilizing mutations result in improvements in catalytic parameters for the hydrolysis of sAAPFna. The combination variant has a specific activity of 290 units/mg compared to 82 units/mg for the wt enzyme. Relative proteinase activity for the hydrolysis of azocasein, however, is only about 82% that of wt. Most of the loss of proteinase activity can be attributed to the N218S mutation, which by itself is 25% less active than wt. Clearly, stabilizing mutations do not necessarily adversely affect catalytic activity.

stable) combination variant is still very similar to that of the wt protein. These results are consistent with other recent X-ray crystallographic studies of stability mutations in proteins (Bryan et al., 1986a; Matthews et al., 1987; Alber et al., 1987; Pantoliano et al., 1988; Matsumura et al., 1988; Nicholson et al., 1988).

This structural isolation of independent mutational events is presumably the physical explanation for the nearly additive incremental assembly of the free energy changes for thermal unfolding ($\Delta\Delta G$) that result from each change. These observations are instructive since they provide strong experimental support for the model of "localization of free energy changes" for protein folding (Tanford, 1970), as exemplified by eq 1. Consequently, this revelation has made it possible to dramatically increase the stability of subtilisin BPN′ in a coherent step by step fashion.

## ACKNOWLEDGMENTS

We thank F. Schwarz and W. Kirchoff for lending us a copy of their computer program for DSC curve fitting. We also thank S. Pulford, J. Nagle, D. Filpula, and L. Moser-Dinterman for oligonucleotide syntheses, DNA sequencing, and fermentation support. We gratefully acknowledge T. Poulos for many helpful discussions during the course of this research and R. Salemme, E. Jackson, and Vasantha N. for their foresight in the initiation of the project.

## REFERENCES

Alber, T., & Wozniak, J. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 747–750.

Alber, T., Sun, D., Wilson, K., Wozniak, J. A., Cook, S. P., & Matthews, B. W. (1987) *Nature 330*, 41–46.

Baker, E. N., & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol. 44*, 97–179.

Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859–1877.

Brandts, J. F. (1964) *J. Am. Chem. Soc. 86*, 4302–4314.

Briggs, R. G., & Fee, J. A. (1978) *Biochim. Biophys. Acta 537*, 100–109.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Bryan, P. N., Rollence, M. L., Pantoliano, M. W., Wood, J., Finzel, B. C., Gilliland, G. L., Howard, A. J., & Poulos, T. L. (1986a) *Proteins: Struct., Funct., Genet. 1*, 326–334.

Bryan, P., Pantoliano, M. W., Quill, S. G., Hsiao, H. Y., & Poulos, T. (1986b) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3743–3745.

Chen, S.-T., Hennen, W. J., Bibbs, J. A., Wang, Y.-F., Liu, J. L.-C., Wong, C.-H., Pantoliano, M. W., Whitlow, M., & Bryan, P. N. (1989) *J. Am. Chem. Soc.* (submitted for publication).

Creighton, T. E. (1984) in *Proteins: Structures and Molecular Properties*, Freeman, New York.

Cunningham, B. C., & Wells, J. A. (1987) *Protein Eng. 1*, 319–325.

Das, G., Hickey, D. R., McLendon, D., McLendon, G., & Sherman, F. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 496–499.

DelMar, E., Largman, C., Brodrick, J., & Geokas, M. (1979) *Anal. Biochem. 99*, 316–320.

Edge, V., Allwell, N. M., & Sturtevant, J. M. (1985) *Biochemistry 24*, 5899–5906.

Fahnestock, S. R., & Fisher, K. E. (1987) *Appl. Environ. Microbiol. 53*, 379–384.

Fersht, A. R. (1985) in *Enzyme Structure & Mechanism*, 2nd ed., Freeman, New York.

Finazzi Agro, A., Federici, G., Giovagnoli, C., Cannella, C., & Cavallini, D. (1972) *Eur. J. Biochem. 28*, 89–93.

Gao, J., Kuczera, K., Tidor, B., & Karplus, M. (1989) *Science 244*, 1069–1072.

Gilson, M. K., & Honig, B. H. (1988) *Proteins: Struct., Funct., Genet. 3*, 32–52.

*Handbook of Chemistry and Physics* (1974–1975) 55th ed., Chemical Rubber Publishing Co., Cleveland, OH.

Hecht, M., Sturtevant, J., & Sauer, R. (1986) *Proteins: Struct., Funct., Genet. 1*, 43–46.

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H., & Salemme, F. R. (1987) *J. Appl. Crystallogr. 20*, 383–387.

Hu, C. Q., & Sturtevant, J. M. (1987) *Biochemistry 26*, 178–182.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268.

Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1–63.

Kellis, J. T., Nyberg, K., Sali, D., & Fersht, A. R. (1988) *Nature 333*, 784–786.

Knowles, J. R. (1987) *Science 236*, 1252–1258.

Kraut, J. (1971) *Enzymes (3rd Ed.) 3*, 547–560.

Levitt, M. (1974) *J. Mol. Biol. 82*, 393–420.

Liao, H., McKenzie, T., & Hageman, R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 576–580.

Lumry, R., & Eyring, H. (1954) *J. Phys. Chem. 58*, 110.

Manly, S. P., Matthews, K. S., & Sturtevant, J. M. (1985) *Biochemistry 24*, 3842–3846.

Matsumura, M., Yasumura, S., & Aiba, S. (1986) *Nature 323*, 356–358.

Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–410.

Matthew, J. B. (1985) *Annu. Rev. Biophys. Biophys. Chem. 14*, 387–417.

Matthews, B., Nicholson, H., & Becktel, W. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6663.

Meloun, B., Baudys, M., Kostka, V., Hausdorf, G., Frommel, C., & Hohne, W. E. (1985) *FEBS Lett. 183*, 195–200.

Mitchinson, C., & Baldwin, R. L. (1986) *Proteins: Struct., Funct., Genet. 1*, 23–33.

Nicholson, H., Becktel, W. J., & Matthews, B. W. (1988) *Nature 336*, 651–656.

Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246*, 2211–2217.

Ottesen, M., & Svendsen, I. (1970) *Methods Enzymol. 19*, 199–215.

Pabo, C. O., & Suchanek, E. G. (1986) *Biochemistry 25*, 5987–5991.

Pantoliano, M. W., Ladner, R. C., Bryan, P. N., Rollence, M. L., Wood, J. F., & Poulos, T. L. (1987) *Biochemistry 26*, 2077–2082.

Pantoliano, M. W., Whitlow, M., Wood, J. F., Rollence, M. L., Finzel, B. C., Gilliland, G., Poulos, T. L., & Bryan, P. N. (1988) *Biochemistry 27*, 8311–8317.

Perry, L. J., & Wetzel, R. (1984) *Science 226*, 555–557.

Philipp, M., & Maripuri, S. (1981) *FEBS Lett. 133*, 36–38.

Ploegman, J. H., Drent, G., Kalk, K. H., & Hol, W. G. J. (1979) *J. Mol. Biol. 127*, 149.

Privalov, P. L., & Potekhin, S. A. (1986) *Methods Enzymol. 131*, 4–51.

Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151–176.

Rollence, M. L., Filpula, D., Pantoliano, M. W., & Bryan, P. N. (1988) *CRC Crit. Rev. Biotechnol. 8*, 217–224.

Russell, A. J., & Fersht, A. R. (1987) *Nature 328*, 496–500.
Sanchez-Ruiz, J. M., Lopez-Lacomba, J. L., Cortijo, M., & Mateo, P. L. (1988) *Biochemistry 27*, 1648–1652.
Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg 29*, 223.
Schellman, J. A. (1975) *Biopolymers 14*, 999–1018.
Schellman, J. A. (1976) *Biopolymers 15*, 999–1000.
Schwarz, F. P. (1988) *Biochemistry 27*, 8429–8436.
Schwarz, F. P., & Kirchhoff, W. H. (1988) *Thermochim. Acta 128*, 267–295.
Schwarzenbach, G., & Flaschka, H. (1965) *Die Komplexometrische Titration*, Ferdinand Enke Verlag, Stuttgart.
Shortle, D., & Lin, B. (1985) *Genetics 110*, 539–555.
Shortle, D., Meeker, A. K., & Freire, E. (1988) *Biochemistry 27*, 4761–4768.
Stearman, R. S., Frankel, A. D., Freire, E., Liu, B., & Pabo, C. O. (1988) *Biochemistry 27*, 7571–7574.
Takahashi, K., & Sturtevant, J. M. (1981) *Biochemistry 20*, 6185–6190.
Tanford, C. (1962) *J. Am. Chem. Soc. 84*, 4240–4247.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1–95.
Tsalkova, T. N., & Privalov, P. L. (1985) *J. Mol. Biol. 181*, 533–544.
Vasantha, N., Thompson, L. D., Rhodes, C., Banner, C., Nagle, J., & Filpula, D. (1984) *J. Bacteriol. 159*, 811–819.
Villafranca, J. E., Howell, E. E., Voet, D. H., Strobel, M. S., Ogden, R. C., Abelson, J. N., & Kraut, J. (1983) *Science 222*, 782–788.
Voordouw, G., Milo, C., & Roche, R. S. (1976) *Biochemistry 15*, 3716–3724.
Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., & Chen, E. Y. (1983) *Nucleic Acids Res. 11*, 7911–7925.
Wong, C. H. (1989) *Science 244*, 1145–1152.
Zoller, M. J., & Smith, M. (1983) *Methods Enzymol. 100*, 468–500.

# Vitamin $B_{12}$ Monocarboxylic Acids: Unambiguous Isomer Assignments by Modern Two-Dimensional NMR Spectroscopy†

Thomas G. Pagano and Luigi G. Marzilli*
*Department of Chemistry, Emory University, Atlanta, Georgia 30322*
*Received January 12, 1989; Revised Manuscript Received April 7, 1989*

ABSTRACT: The three cyanocobalaminmonocarboxylic acid isomers known to be produced by the mild acid hydrolysis of the *b*-, *d*-, and *e*-propionamide side chains of vitamin $B_{12}$ have been unambiguously assigned by modern 2D NMR methods. Previously, structural assignments had been made by less definitive NMR methods, and both X-ray and neutron diffraction studies had failed to locate unambiguously the position of the carboxyl group. The b and e isomers were structurally assigned in this study, on the basis of the assignment of the $^{13}C$ NMR signal of the carboxyl group from HMBC ($^1H$-detected heteronuclear multiple-bond correlation) spectra. The carboxyl group resonances exhibited the greatest changes in chemical shift between the protonated (pH 2) and deprotonated (pH >7) forms of the acids. The d isomer was assigned by difference. Since the HMBC experiments required the assignments of side-chain $CH_2$ signals, homonuclear Hartmann–Hahn, 2D homonuclear correlation, 2D nuclear Overhauser effect, $^1H$-detected heteronuclear multiple quantum coherence, and HMBC spectroscopies were used to assign completely the $^1H$ and $^{13}C$ NMR spectra of the b and e isomers at pH ~7. By comparison with the $^{13}C$ NMR spectra of the b and e isomers, nearly one-fourth of the resonances of the $^{13}C$ NMR spectrum of vitamin $B_{12}$ have been reassigned. The sites of incorporation of $^{13}C$-labeled precursors in $B_{12}$ biosynthesis found in previous studies have been verified by a comparison of $^{13}C$ assignments. The results of studies using cobalamins modified at the *b*-, *d*-, and *e*-propionamide side chains in which the incorrect structural assignments were used (before 1980), particularly studies of $B_{12}$-dependent enzymes, require reinterpretation using the correct structural assignments.

An essential early step in the catalytic cycle of $B_{12}$-dependent enzymes is the homolysis of the Co–C bond of coenzyme $B_{12}$ (AdoCbl)[1] (Dolphin, 1982). Although several hypotheses have been proposed to account for the roles the enzymes play in promoting homolysis (Finke et al., 1984; Halpern, 1985; Bresciani-Pahor et al., 1985), the exact mechanism is unknown. There is general agreement that an enzyme-induced distortion of the coenzyme leads to the Co–C bond cleavage.

The six amide groups around the periphery of the corrin ring in the coenzyme often play essential roles in these hypotheses. These amide groups can interact with the enzymes (Toraya et al., 1979). Since the molecular masses of the holoenzymes range from 76 to 560 kDa (Dolphin, 1982), it is very difficult to obtain structural information from direct techniques such as NMR spectroscopy, and no X-ray structures of a $B_{12}$ enzyme have been reported. Therefore, indirect methods such as the use of coenzyme $B_{12}$ analogues modified at the pro-

† This work was supported by NIH Grant GM 29225.
* Author to whom correspondence should be addressed.

[1] Abbreviations: AdoCbl, (5′-deoxyadenosyl)cobalamin; CNCbl, cyanocobalamin; NMR, nuclear magnetic resonance; 2D, two dimensional; 1D, one dimensional; COSY, homonuclear shift correlation spectroscopy; HOHAHA, homonuclear Hartmann–Hahn; NOESY, nuclear Overhauser effect spectroscopy; NOE, nuclear Overhauser effect; HMBC, $^1H$-detected multiple-bond heteronuclear multiple quantum coherence; HMQC, $^1H$-detected heteronuclear multiple quantum coherence; TSP, sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d*$_4$; ALA, δ-aminolevulinic acid; PBG, porphobilinogen.

0006-2960/89/0428-7213$01.50/0 