# Positive Selection Dictates the Choice between Kinetic and Thermodynamic Protein Folding and Stability in Subtilases

Ezhilkani Subbian, Yukihiro Yabuta, and Ujwal Shinde*

*Department of Biochemistry and Molecular Biology, Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239*

*Received July 27, 2004; Revised Manuscript Received September 1, 2004*

ABSTRACT: Subtilisin E (SbtE) is a member of the ubiquitous superfamily of serine proteases called subtilases and serves as a model for understanding propeptide-mediated protein folding mechanisms. Unlike most proteins that adopt thermodynamically stable conformations, the native state of SbtE is trapped into a kinetically stable conformation. While kinetic stability offers distinct functional advantages to the native state, the constraints that dictate the selection between kinetic and thermodynamic folding and stability remain unknown. Using highly conserved subtilases, we demonstrate that adaptive evolution of sequence dictates selection of folding pathways. Intracellular and extracellular serine proteases (ISPs and ESPs, respectively) constitute two subfamilies within the family of subtilases that have highly conserved sequences, structures, and catalytic activities. Our studies on the folding pathways of subtilisin E (SbtE), an ESP, and its homologue intracellular serine protease 1 (ISP1), an ISP, show that although topology, contact order, and hydrophobicity that drive protein folding reactions are conserved, ISP1 and SbtE fold through significantly different pathways and kinetics. While SbtE absolutely requires the propeptide to fold into a kinetically trapped conformer, ISP1 folds to a thermodynamically stable state more than 1 million times faster and independent of a propeptide. Furthermore, kinetics establish that ISP1 and SbtE fold through different intermediate states. An evolutionary analysis of folding constraints in subtilases suggests that observed differences in folding pathways may be mediated through positive selection of specific residues that map mostly onto the protein surface. Together, our results demonstrate that closely related subtilases can fold through distinct pathways and mechanisms, and suggest that fine sequence details can dictate the choice between kinetic and thermodynamic folding and stability.

The question of how proteins find their unique native structures by using information encoded within their amino acid sequences lies at the heart of molecular biology. Understanding mechanisms of folding and unfolding are of immense scientific significance because of their potential, in providing new approaches to combat protein-misfolding disorders and in facilitating the design of novel proteins that catalyze desired reactions. In addition, the protein folding problem is ostensibly the most fundamental example of biological self-assembly, and symbolizes the first step towards tackling one of the most arduous questions that can be addressed by contemporary science: How have complex biological systems evolved the ability to self-assemble into robust and predictable systems? The current literature suggests that each sequence folds through a unique energy landscape that is dictated by the intrinsic properties of the polypeptide, and by the extrinsic influence of the folding environment (*1*). The challenge now is to recognize how an energy landscape that is inherent to a given protein is defined by its amino acid sequence, and the extent to which external factors influence this process.

A popular approach to gaining insights into general protein folding principles and to determining what characteristics of this process are under evolutionary control is to compare folding of homologous proteins (*2*−*5*). Several elegant studies have demonstrated that homologous proteins fold through similar folding pathways and transition states (*2*, *4*, *6*−*8*). There is also a growing body of evidence which shows that the folding rates of proteins with diverse sequences strongly correlate with more global parameters such as contact order, topology, average hydrophobicity, and protein length (*9*). In essence, fine sequence details of the protein do not affect folding pathways and mechanisms (*5*, *10*).

While the current view of folding has emerged through studies based on proteins that fold into their global energy minima (thermodynamically stable states), there is a growing body of evidence which shows that several proteins can fold to kinetically stable local minima (*11*−*20*). While thermodynamically stable protein conformations are generally the norm, kinetically stable conformations may be selected for specific functional advantages. For example, kinetic stability provides mechanisms for protease longevity in α-lytic protease (*13*, *14*, *21*), for biological regulation in serpins (*17*, *18*), and for membrane insertion of viral glycoproteins (*20*). In the case of the broadly specific α-lytic protease, the uncoupling of the folding and unfolding pathways upon propeptide degradation enables the native state to be locked in a metastable state that confers low conformational entropy. This enhances the longevity of the protease by reducing its

* To whom correspondence should be addressed. E-mail: shindeu@ohsu.edu. Phone: (503) 494-8683. Fax: (503) 494-8393.

10.1021/bi048397x CCC: $27.50 
Published on Web 10/21/2004

autoproteolytic susceptibility (*14*, *21*). Studies on plasminogen activator inhibitor 1 have shown that the active inhibitory form converts to a latent inactive form over a period of several hours. The latent form can be converted back to the active form by denaturation and renaturation. Thus, the active inhibitory form of the protein is not at its lowest-energy state but slowly converts to the latent form that appears to be more stable (*22*−*24*). In some serpins, this conversion from the active to the latent form occurs upon proteolysis (*17*, *18*, *25*). These provide unique methods for inhibition of serine protease activity. In the case of hemagglutinin, the trimeric envelope glycoprotein undergoes a dramatic loop to a coiled-coil transition upon exposure to low pH. The initially synthesized native state has a disordered loop region that appears to be separated from a more stable state by a kinetic barrier. The conformational change upon pH change facilitates membrane insertion (*20*, *26*, *27*). Hence, in the cases discussed above, the kinetically trapped native conformations enhance the biological functionality of proteins. Since the folding landscape is encoded in the polypeptide sequence, the above examples raise an important question of whether nature selects for specific folding pathways through fine sequence details. Agard and coworkers have compared folding of the kinetically trapped α-lytic protease with thermodynamically stable trypsin (*28*) to understand the functional basis of thermodynamic versus kinetic stability (*15*, *16*, *29*, *30*). However, although α-lytic protease and trypsin adopt very similar folds, their amino acid chains display insignificant sequence similarity (*21*). This prevents an analysis of how primary sequence evolution dictates selection between thermodynamic and kinetic stability because α-lytic protease and trypsin may have evolved via convergent evolution (*31*). The key to understanding how protein sequences have evolved to select between thermodynamic and kinetic stability is to identify protein subfamilies closely related by sequence and structure, but different in their folding mechanisms, such that one subfamily folds into thermodynamically stable native states, while the other adopts kinetically trapped conformers. This would facilitate a detailed analysis of the relation between primary sequences, native structures, folding pathways, and evolutionary constraints that dictate the choice between thermodynamic and kinetic stability.

Subtilases constitute a large ubiquitous superfamily of calcium-dependent serine proteases that span prokaryotes, eukaryotes, and archea (*11*, *13*, *32*). On the basis of their sequence homology, subtilases can be divided into at least six families, all of which have conserved catalytic domains (*32*). Although levels of overall sequence identity may vary between these families, they all adopt very similar three-dimensional structures. Most members of the subtilase superfamily are produced with dedicated propeptide domains located downstream from signal peptides. These propeptides are demonstrated to be essential for correct folding of their cognate protease domains in several prokaryotic and eukaryotic subtilases (*11*, *32*). Propeptide deletion variants of these subtilases are correctly localized but are unable to fold into their native states, suggesting that propeptides do more than merely assist signal peptides to correctly localize their protease domains. The contemporary understanding of propeptide-mediated folding mechanisms of subtilases has emerged from analyses of bacterial subtilisins (*11*, *13*).

In this paper, we investigate folding in the highly conserved subtilase superfamily (*11*, *32*). Like other subtilases, subtilisin E (SbtE)[1] is a calcium-dependent ESP that is secreted as a precursor (pro-SbtE) and is an established model for understanding propeptide-dependent folding (*11*−*13*, *32*−*39*). In absence of its propeptide (77-residue), SbtE (275-residue) folds into a stable, molten-globule intermediate that is separated from its kinetically trapped native state by a high-energy barrier (*13*, *33*). Spontaneous conversion to the native state is slow ($t_{1/2} \sim$ 1500 years) but is catalyzed $\sim 10^6$-fold by the propeptide, which lowers this high-energy barrier (*13*, *33*). Hence, propeptides function as dedicated intramolecular chaperones that are essential for efficient folding of their cognate protease domains to kinetically trapped native states (*11*, *21*, *22*, *40*). A database search of more than 500 homologues of SbtE identified several bacterial intracellular serine proteases that are highly conserved in sequence but lack classical propeptide domains. Through detailed analyses of amino acid and nucleotide sequences, protein structures, and folding kinetics, we demonstrate that intracellular serine protease 1 (ISP1), an ISP, and its homologue SbtE, an ESP from the same organism (*Bacillus subtilis*), can fold through significantly different pathways and kinetics. Despite displaying conserved topology, contact order, and hydrophobicity, SbtE absolutely requires assistance from its propeptide to fold into a kinetically trapped conformer, while ISP1 folds to a thermodynamically stable state more than 1 million times faster and independent of a propeptide. Analysis of the themodynamics and kinetics suggests that the folding transition state of ISP1 is different from that of SbtE. Interestingly, the differences between ESPs and ISPs appear to be localized mostly on the protein surfaces, with ISPs displaying a bias toward acidic amino acid residues (pI $<$5.5) and ESPs preferring basic amino acid residues (pI $>$8.0). In addition, ISPs lose the specificity and affinity for the propeptides of ESPs but continue to bind tightly with known subtilisin inhibitors. Evolutionary analysis suggests that the primary sequence of ISPs displays significant adaptive evolution of surface residues. Together, our results suggest that surface residues dramatically influence folding, and their positive selection can dictate folding pathways, mechanisms, and the choice between kinetic and thermodynamically stable folds in subtilases.

## MATERIALS AND METHODS

*Sequence Analysis and Homology Modeling.* Protein sequences were aligned using ClustalW and analyzed in GeneDoc. Homology modeling of ISP1 was performed using Swiss Model, a program that predicts structures reliably (rmsd $<$ 2 Å) for sequences that are 50−60% identical (*41*). The structure was validated using ANOLEA, ProSAII, VADAR, and WHATCHECK and displayed using Pymol. Sites within the conserved hydrophobic core have a solvent

[1] Abbreviations: ESP, extracellular serine protease; ISP, intracellular serine protease; SbtE, subtilisin E; pro-SbtE, propeptide subtilisin E; ISP1, intracellular serine protease 1; HCI, hydrophobicity core index; CD, circular dichroism; GdnHCl, guanidine hydrochloride; *N*-succ-AAPF-*p*NA, *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide; LRT, likelihood ratio test; SSI, *Streptomyces* subtilisin inhibitor.

accessibility of <15% and a hydrophobicity core index (HCI) of >60%, where

$$\mathrm{HCI} = (\sum_{i=1}^{n} \mathrm{AA} \times \mathrm{Hyd}) \frac{N_{\mathrm{Hyd}}}{N} \quad (1)$$

where AA represents hydrophobic residues, Hyd is the hydrophobicity of AA, $N_{\mathrm{Hyd}}$ is the number of AA at site *i*, and *N* is the total number of residues at site *i*.

*Protein Expression and Purification.* ISP1 was amplified from the genomic DNA of *B. subtilis* (ATCC catalog no. 23857D, American Type Culture Collection, Manassas, VA) using PCR and cloned into pET11a. Site-directed mutagenesis was carried out as described previously (*35*). Nucleotide sequences were confirmed by sequencing both strands. Proteins were expressed in *Escherichia coli* BL21(DE3) and purified from the soluble fraction using techniques as described previously (*35*, *37*, *40*). Mature ISP1 purified from *E. coli* lacks its 17-residue N-terminus that does not affect its folding.

*Structural Characterization.* CD spectra were measured on an AVIV-215 spectrophotometer using a quartz cuvette with a path length of 1 mm (*40*). Protein samples (7−15 $\mu$M) were taken in renaturation buffer [50 mM Tris-HCl (pH 8.5), 0.5 M $(NH_4)_2SO_4$, and 1 mM $CaCl_2$]. The thermostability was monitored through the loss in ellipticity at 222 nm. For intrinsic tryptophan fluorescence (PTI fluorometer), samples were excited at 295 nm and emission spectra between 300 and 400 nm were obtained. Folded and unfolded samples were in renaturation buffer and 6 M guanidine hydrochloride (GdnHCl), pH 4.8, respectively. The fluorescence change upon folding or unfolding was used to monitor kinetics. Stopped-flow kinetics were measured on a Hi-Tech SF61-DX2 fluorometer using a $\lambda_{\mathrm{excitation}}$ of 295 nm and a 310 nm band-pass filter. For the Chevron analysis, folded and unfolded $S_{246}$A-ISP1 and Pro-SbtE (1−4 $\mu$M) were diluted (1:10 ratio) into renaturation buffer containing different concentrations of GdnHCl (0−5 M). Data were fitted to single- and double-exponential equations (GraphPad-Prism) to determine folding and unfolding rate constants (*40*). Chevron plots best fit a three-state equation (*8*, *42*, *43*). All experiments were performed at 23 °C and are averages of three to six independent repeats.

*Activity, Inhibition, and Calcium-Dependent Stability.* Activities of ISP1 and SbtE were measured at 405 nm using the chromogenic substrate *N*-succ-AAPF-*p*NA (*35*, *40*). The calcium dependence of ISP1 and SbtE was measured in Chelex-treated buffer [50 mM Tris-HCl (pH 8.5) and 0.5 M $(NH_4)_2SO_4$], to which various concentrations of calcium were added (*39*). Aliquots were removed as a function of time, and protease activity was measured to obtain decay rates ($k_{\mathrm{decay}}$). Slopes obtained from plots of $k_{\mathrm{decay}}$ versus [$Ca^{2+}$/protein] provide the relative calcium-dependent stability of ISP1 and SbtE. Inhibition constants ($K_{\mathrm{I}}$) were measured by adding proteases [ISP1 (7.6 nM) or SbtE (58.3 nM)] to renaturation buffer containing inhibitors [propeptide (0.05−35 $\mu$M), SSI (0.05−12.5 nM), or BSA (1.18−20 $\mu$M)] and 0.5 mM *N*-succ-AAPF-*p*NA (*35*). $K_{\mathrm{I}}$ was obtained from $EC_{50}$ plots of normalized rates of *p*-nitroanilide release versus inhibitor concentration (GraphPad-Prism).

*Kinetics and Equilibrium Folding and Unfolding.* Pro-SbtE folding and SbtE folding were monitored as described previously (*36*, *40*). Purified ISP1 was completely denatured (loss of structure and activity), after precipitation [55% $(NH_4)_2SO_4$] and resuspension in 6 M GdnHCl (pH 4.8). Denatured ISP1 (22 $\mu$M) was rapidly diluted (50-fold) into renaturation buffer, and the recovered activity was assayed at various times after folding initiation. Figure 4c represents maximum activity recovered after complete folding. Equilibrium folding or unfolding was monitored (CD at 222 nm) to obtain free energy differences and *m* values for each transition as described previously (*8*, *33*, *42*−*44*). The transitions are represented by at least three states (*43*), the native (N), intermediate (I), and unfolded states (U) as follows:

$$\mathrm{N} \overset{K_{\mathrm{NI}}}{\rightleftharpoons} \mathrm{I} \overset{K_{\mathrm{IU}}}{\rightleftharpoons} \mathrm{U} \quad (2)$$

The observed ellipticity [$A_{\mathrm{obs}}(c)$] at any concentration of the denaturant is given by the sum of the contributions from the three states as (*42*, *45*)

$$A_{\mathrm{obs}}(c) = \frac{A_{\mathrm{N}} + A_{\mathrm{I}} \exp[-(\Delta G_{\mathrm{NI}}^{\mathrm{H_2O}} - m_{\mathrm{NI}}c)/RT] + A_{\mathrm{U}} \exp[-(\Delta G_{\mathrm{NU}}^{\mathrm{H_2O}} - m_{\mathrm{NU}}c)/RT]}{1 + \exp[-(\Delta G_{\mathrm{NI}}^{\mathrm{H_2O}} - m_{\mathrm{NI}}c)/RT] + \exp[-(\Delta G_{\mathrm{NU}}^{\mathrm{H_2O}} - m_{\mathrm{NU}}c)/RT]} \quad (3)$$

where $f_{\mathrm{N}}(c)$, $f_{\mathrm{I}}(c)$, and $f_{\mathrm{U}}(c)$ are the fractions of the three states at a GdnHCl concentration of *c* ($f_{\mathrm{N}} + f_{\mathrm{I}} + f_{\mathrm{U}} = 1$) and $A_{\mathrm{N}}$, $A_{\mathrm{I}}$, and $A_{\mathrm{U}}$ are the ellipticity values of the pure N, I, and U states, respectively. The $f_{\mathrm{N}}$, $f_{\mathrm{I}}$, and $f_{\mathrm{U}}$ terms are related to the equilibrium constants, $K_{\mathrm{NI}}$ and $K_{\mathrm{NU}}$, of the unfolding transitions from N to I and from N to U, respectively. $\Delta G_{\mathrm{NI}}^{\mathrm{H_2O}}$ and $\Delta G_{\mathrm{NU}}^{\mathrm{H_2O}}$ are $\Delta G_{\mathrm{NI}}$ and $\Delta G_{\mathrm{NU}}$ at 0 M GdnHCl, respectively, and $m_{\mathrm{NI}}c$ and $m_{\mathrm{NU}}c$ represent the dependence of the respective free energy changes on *c*. The data were fitted using Prism Graphpad.

*Statistical Analysis of Phylogeny.* Protein sequences of ESPs and ISPs (SwissProt, Hobacgen, and Merops databases) were aligned using ClustalW and used to obtain nucleotide alignments. DNAML (PHYLIP package) was used to estimate tree topology (Figure 6a). Positive selection ($\omega > 1$) was tested using several codon-based likelihood models (codeml) in PAML (*46*). Likelihood-ratio tests (LRTs) were used to establish best fits by comparing general models with their null models (*47*−*49*). The null distribution of the LRT statistic ($2\Delta l$, where $\Delta l$ equals the difference between log-likelihood scores of the two models) can be approximated using the $\chi^2$ distribution, with the degree of freedom being the difference in the number of free parameters between the two models. Briefly, the site models were M0 (one $\omega$ for all sites), M1 (allows $\omega = 1$ and $\omega = 0$), M2 ($\omega = 1$, $\omega = 0$, and $\omega > 1$), M3 (discrete model with three site classes for $\omega$ estimated from data), M7 ($\beta$ distribution for $\omega$ from 0 and 1), and M8 ($\beta$ distribution from 0 to 1, and independent $\omega$ estimated from data). The branch site models used were model A and model B (*47*−*49*). Branch site models test for adaptive evolution along a specific lineage, which is termed the foreground branch. The basic branch site model allows for four site classes. The first two classes ($\omega_0$ and $\omega_1$) are uniform over the entire phylogeny. The other two classes allow for some sites with $\omega_0$ and $\omega_1$ to change to positive

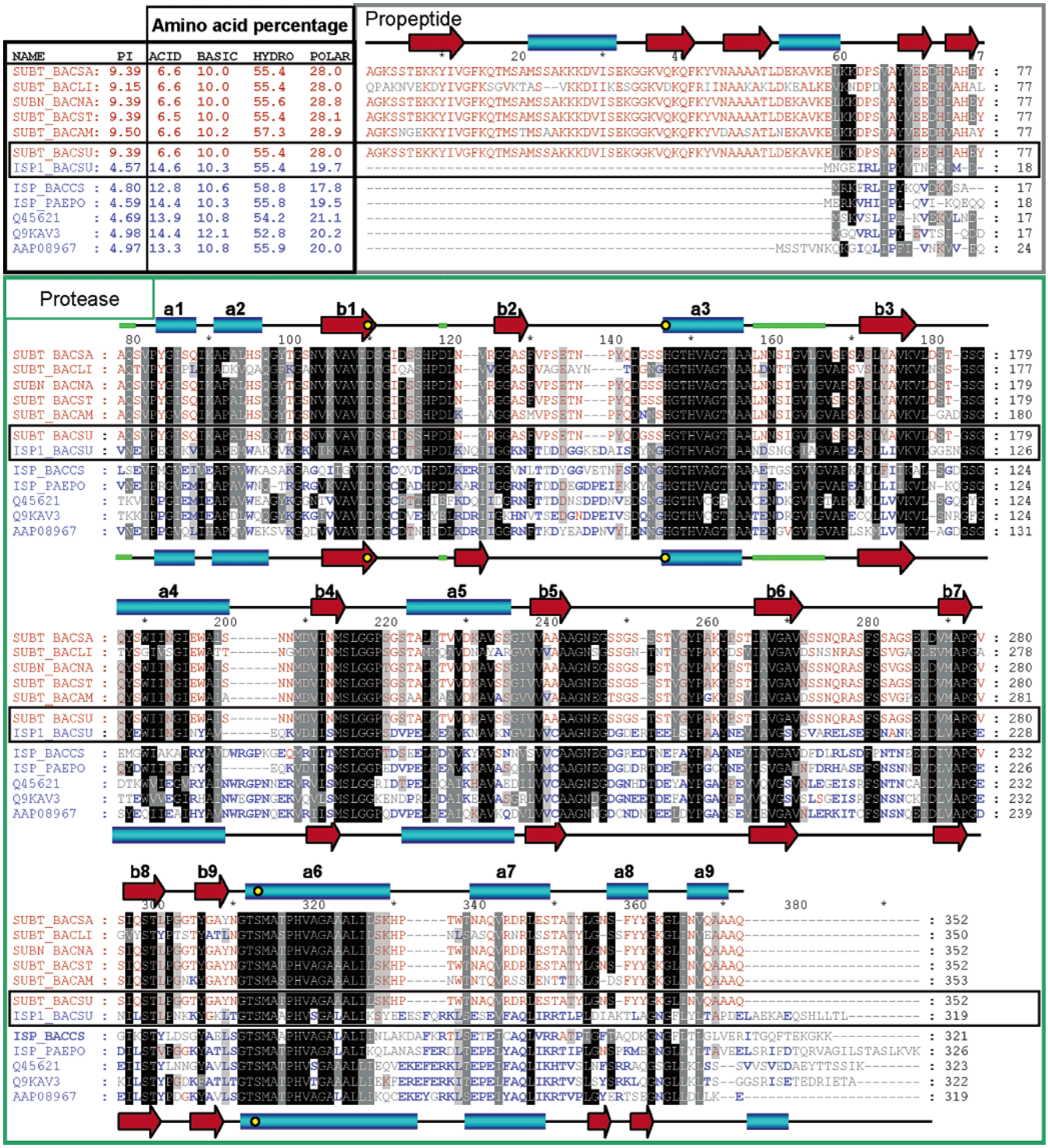


FIGURE 1: Sequence conservation between ISPs (blue) and ESPs (red). Multiple-sequence alignment (MSA) between representative ISPs and ESPs. SUBT_BACSU (SbtE) and ISP1_BACSU (ISP1) are our models whose secondary structures are depicted above and below the MSA, respectively. Active sites (yellow circles), calcium binding domains (green bars), and sequence conservation (black shading, >75% identical; gray shading, >70% similar) are depicted. Specific regions are conserved (>65% identical) only within ESPs (red type) or ISPs (blue type). The amino acid composition and isoelectric points are shown N-terminal to each sequence (ACID, D, E; BASIC, K, R, H; HYDRO, A, V, L, F, G, I, Y, W, P, M, C; POLAR, S, N, Q, T).

selection ($\omega_2$) along the foreground branch. In model A, $\omega_0$ and $\omega_1$ are fixed, while in model B, they are estimated as free parameters (*47*, *48*).

The foreground branch for our analysis is highlighted in blue (Figure 6a). When LRTs suggest positive selection, the posterior probabilities for each site to have a specific $\omega$ class can be calculated using Bayes theorem (*47*−*49*). Our evolutionary analysis of ISPs and ESPs (branch site models A and B) from *Bacillus* suggests that adaptive evolution followed gene duplication leading to ISPs. The Bayes approach identified ∼22% sites for which $\omega > 1$ and hence have positively evolved in the ISP clade. Interestingly, most of these sites have been identified as being critical for folding, activity, and stability of SbtE through independent mutagenesis and directed evolution studies (*50*). Table 3 lists sites that show a strong positive selection in ISPs together with

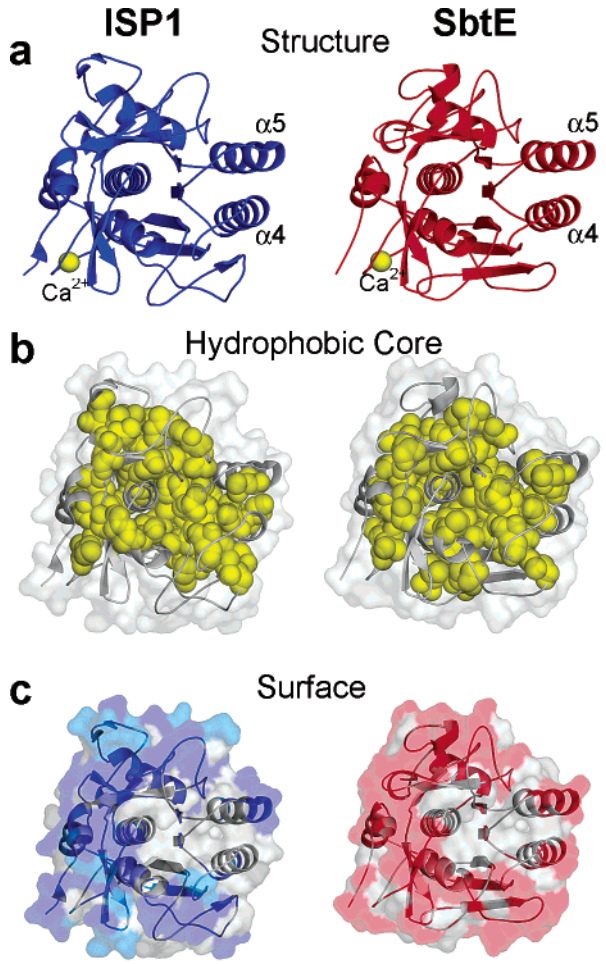


FIGURE 2: Structural conservation between ISPs (blue) and ESPs (red). (a) Ribbon models of SbtE (PDB entry 1SCJ) and ISP1 (homology model, validated as described in Materials and Methods). α4 and α5 are critical for propeptide interaction with SbtE. (b) Conserved hydrophobic core (yellow space-filling diagram) within ISPs and ESPs. (c) Sequence conservation from MSA mapped onto structures. Residues conserved within both ESPs and ISPs are colored gray, while red or blue denotes residues that are conserved within only ESPs or ISPs, respectively. Residues that are conserved within ESPs, but without selection constraints in ISPs, are colored cyan.

the corresponding sites in ESPs and the role of that site in folding, stability, and activity of ESPs as identified through independent studies. Consistent with studies on other protein families, trees generated using different programs (DNAPARS and DNADIST) do not affect the analysis (data not shown).

## RESULTS

*Sequence, Structural, and Functional Analysis of Bacillus ESPs and ISPs.* A database search for SbtE-like bacterial proteases identified two subfamilies (ISPs and ESPs) within subtilases that exhibit significant sequence, structure, and functional conservation (*11*, *32*). High levels of sequence identity (~50%) within these subgroups (Figure 1), and availability of crystallographic structures for several ESPs, including SbtE (*34*), allow development and validation of reliable homology models (*51*) for ISPs (Figure 2a). Subtilases even with a low level of sequence identity have been known to adopt similar three-dimensional structures (*32*). For example, the eukaryotic subtilases Kex2 and furin

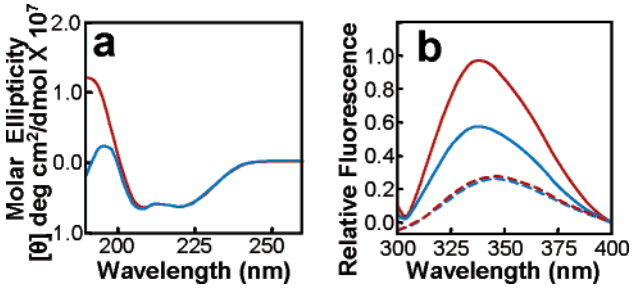


FIGURE 3: Secondary and tertiary structure comparison between ISP1 (blue) and SbtE (red). (a) Secondary structure comparison using CD. Table 1 contains structural contributions. (b) Intrinsic tryptophan fluorescence. Native state (solid lines) SbtE displays fluorescence higher than that of ISP1 due to an additional tryptophan. Unfolded proteins (dashed lines) lose fluorescence intensity.

(prohormone convertases) adopt structural scaffolds (*52*, *53*) that are similar to bacterial subtilisins despite a low level of sequence identity (26% with Kex and 37% with furin). Since ISP1 and SbtE have sequences that are >50% identical, we modeled the structure of ISP1 using subtilase templates. Because of the high levels of sequence identity, the modeled structure of ISP1 can be considered with confidence. When sequence conservation seen in the multiple-sequence alignment (Figure 1) is mapped onto the structure of ESPs and ISPs, conserved sites (black and gray highlight) map to the protein core, while sequence differences between the two families (blue and red type) map predominantly onto the protein surface (Figure 2b,c). Analysis of the amino acid distribution in the two families demonstrates that the percentages of hydrophobic residues within ESPs and ISPs are similar (see amino acid distribution in Figure 1). Further, quantitation (Table 1) of the hydrophobic core establishes similar core strength (see Materials and Methods for details; Figure 2b) within ESPs and ISPs. Calcium binding sites that are critical for ESP stability (*13*, *37*), are also well-conserved within ISPs (green bars in Figure 1). However, although ISPs and ESPs share significant sequence and structural similarities, ISPs lack propeptides that are critical for ESP folding. The propeptides of ESPs are extremely charged (36% charge) and intrinsically unstructured polypeptides (*54*, *55*). Interestingly, ISPs are biased for negatively charged residues which is evident in their acidic isoelectric points (Figure 1 and Table 1), and charges that are localized within propeptides of ESPs appear to be evenly distributed within the protease domain of ISPs.

To obtain insights into structure and folding of ISPs, we cloned, expressed, and purified ISP1 (intracellular serine protease 1) from *B. subtilis* (Materials and Methods) to compare and contrast with SbtE, our ESP model from the same species (*11*, *40*). The secondary structure of SbtE, pro-SbtE, ISP1, and their active site variants were compared using circular dichroism (CD) spectroscopy, and the fractions of secondary structure content were estimated as described previously (*35*, *36*, *40*, *56*). Figure 3a demonstrates that the secondary structure spectra of ISP1 and SbtE are superimposable, and their deconvolution yields similar overall secondary structure content (Table 1), and is consistent with our homology model. Thermostability estimated using changes in CD ellipticity as a function of temperature indicates that the active site variants of ISP1 and SbtE display similar melting temperatures (Table 1). Both ISP1 and SbtE exhibit

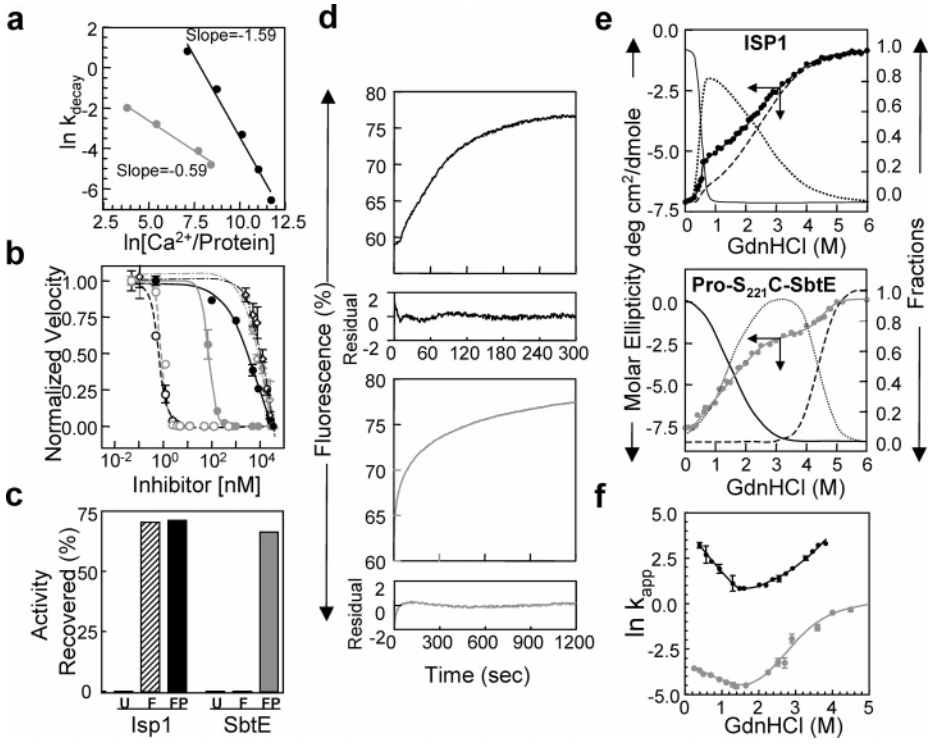


FIGURE 4: Characterization of ISP1 (black) and SbtE (gray). (a) Activity and stability (measured as $k_{decay}$) as a function of $Ca^{2+}$. ISP1 activity decays 3-fold faster as measured from relative slopes. (b) $EC_{50}$ plots for ISP1 and SbtE inhibition by SSI (○), propeptide (●), and BSA (◇). (c) Activity recovered upon refolding. Denatured ISP1 (U) recovers activity upon refolding (F, gray bar) and is unaffected by the propeptide (FP, black bar). Unfolded SbtE (U) when refolded is inactive (F) unless it is folded in the presence of its propeptide (FP, gray bar). (d) Refolding kinetics of $S_{246}$A-ISP1 (top) and pro-$S_{221}$C-SbtE (bottom). Residuals for best-fit curves are below each graph. (e) Folding−unfolding equilibrium of $S_{246}$A-ISP1 and pro-$S_{221}$C-SbtE monitored using CD at 222 nm. Resolved fractions of native (—), unfolded (− − −), and intermediate (···) states are shown. (f) Chevron plots fitted to a three-state model (*8*, *42*) yield the following: $m_{IU} = 3.37$ and $m_{NU} = 4.50$ for $S_{246}$A-ISP1 (black) and $m_{IU} = 1.78$ and $m_{NU} = 3.33$ for pro-$S_{221}$C-SbtE (gray).

a blue shift and an increase in intrinsic tryptophan fluorescence with an emission maximum at 330 nm, when compared with their unfolded states (Figure 3b). Because of the presence of an additional tryptophan residue, native SbtE displays a higher fluorescence than ISP1. The change in fluorescence observed upon denaturation and renaturation can be used to monitor the folding kinetics of ISP1 and SbtE.

We next compared enzymatic properties of ISP1 and SbtE. ISP1 is completely inhibited by ethylenediaminetetraacetic acid and phenylmethanesulfonyl fluoride, like SbtE (data not shown), and exhibits a 15-fold higher specific activity toward a chromogenic substrate (Table 1). Since subtilases are known to be dependent upon $Ca^{2+}$ ions for their stability and/or catalytic activity, the relative affinities of ISP1 and SbtE for $Ca^{2+}$ were estimated as described in Materials and Methods. Both proteases display strong $Ca^{2+}$ dependence for their stability and activity, with ISP1 having an only 3-fold lower $Ca^{2+}$-dependent proteolytic stability (Figure 4a).

To gain deeper insight into the similarities and differences between the enzymatic properties of ISP1 and SbtE, their binding affinities for known subtilase inhibitors were examined. *Streptomyces* subtilisin inhibitor (SSI), one of the strongest inhibitors of bacterial subtilases (*57*), can interact with SbtE ($K_I = 0.17$ nM) and ISP1 ($K_I = 0.26$ nM) with similar affinity (Table 1 and Figure 4b). The propeptide inhibits SbtE with a $K_I$ of 18.5 nM. While this affinity is approximately 10-fold weaker than that observed for SSI, it is similar to the affinity of other ESPs such as subtilisin BPN′ and Carlsberg (*57*), for the propeptide of SbtE. On the other hand, the propeptide inhibits ISP1 with an affinity ($K_I =$ 1410 nM) that is ∼100-fold weaker than that for SbtE. Interestingly, this affinity is similar to that of a nonspecific protease substrate, such as bovine serum albumin (BSA), for both ISP1 and SbtE (Table 1 and Figure 4b). The diminished affinity of the propeptide for ISP1 is a likely consequence of changes in specific surface residues that constitute the propeptide−protease interface. Hence, together, the experimental data given above suggest that ISP1 and SbtE display catalytic properties of classical serine proteases but differ in their interaction with propeptides.

*A Comparison of the Folding Pathways of ISP1 and SbtE.* To examine reversibility of folding, denatured ISP1 was rapidly diluted into renaturation buffer (*40*), and recovered activity was used as a direct indicator of folding. Figure 4c establishes that ISP1 folding is reversible and its efficiency is ∼70%, probably due to autoproteolysis. Under identical conditions, the protease domain of SbtE folds exceedingly slowly, but folding is rapidly catalyzed by the propeptide added *in trans* (Figure 4c). Such a reversible folding−unfolding equilibrium provides sufficient evidence for a thermodynamically stable native state (*21*, *22*, *58*). Hence, unlike ESPs whose folding is under kinetic control (*11*, *13*), ISPs adopt thermodynamically stable folds independent of propeptides.

To investigate the thermodynamic versus kinetic folding of ISPs and ESPs, subsequent studies were performed using active site variants that lack complications of proteolysis (*38*). $S_{246}$A-ISP1 adopts a secondary structure that can be superimposed with that of ISP1 but lacks proteolytic activity (data not shown). Folding kinetics of $S_{246}$A-ISP1 and pro-$S_{221}$C-

SbtE (*34*, *38*) were monitored using stopped-flow fluorescence (Figure 4d). Both best fit a double-exponential rate equation, with rate constants of 20.23 and 0.012 $s^{-1}$ for $S_{246}$A-ISP1 and 0.021 and 0.0016 $s^{-1}$ for pro-$S_{221}$C-SbtE, respectively. Comparison of the two phases observed during folding suggests that the rapid phase of $S_{246}$A-ISP1 is $\sim 10^3$-fold faster than that of pro-$S_{221}$C-SbtE. Observed slow phases may be consequences of proline isomerization (*33*). Unassisted folding of SbtE is extremely slow (*33*) ($t_{1/2} \sim 1500$ years), and its kinetics cannot be accurately measured using fluorescence. Fersht and co-workers (*33*) have estimated that the propeptide enhances the folding rate of the protease domain by $\sim 10^6$-fold. Hence, while they both have conserved structures, folding of ISP1 appears to be $\sim 10^9$ times faster than propeptide-independent folding of SbtE.

Rigorous analysis of folding kinetics reveals a burst phase within the dead time of the stopped-flow instrument. The occurrence of burst phase kinetics in several folding models suggests the presence of transient intermediates which are established counterparts of equilibrium intermediates (*43*). Since analysis of intermediates can provide insights into folding pathways, we examined equilibrium folding−unfolding transitions of $S_{246}$A-ISP1 and pro-$S_{221}$C-SbtE (Figure 4e). The folding−unfolding equilibrium of $S_{246}$A-ISP1 is completely reversible with two distinct transitions, and is represented by at least three states (*43*), the native (N), intermediate (I), and unfolded states (U) (see Materials and Methods). Parameters that best represent folding and unfolding of $S_{246}$A-ISP1 are listed in Table 1. Pro-$S_{221}$C-SbtE undergoes autoprocessing to give a thermodynamically stable stoichiometric complex, the structure of which has been determined (*34*). This complex also shows a three-state equilibrium transition (Figure 4e). Studies suggest that the I state observed in $S_{246}$A-ISP1 is on-pathway (data not shown) as shown in the case of prosubtilisin (*33*). This I state is more structured (70% of the native ellipticity at 222 nm) than that observed in pro-$S_{221}$C-SbtE ($\sim$33% of the native ellipticity), and deconvolution (Figure 4e) establishes that the I state of $S_{246}$A-ISP1 is maximally populated in 0.8 M GdnHCl when compared with the I state of pro-$S_{221}$C-SbtE (2.5 M GdnHCl). At high denaturant concentrations, deviations from linearity (rollovers) are observed in the Chevron plots (*59*) [ln($k_{obs}$) vs denaturant] of both $S_{246}$A-ISP1 and pro-$S_{221}$C-SbtE (Figure 4f). This is typically ascribed to folding intermediates (*8*) and is consistent with our equilibrium studies. The extent of solvent exposure between the states involved in a transition is defined by *m* values, and the $m_{IU}/m_{NU}$ ratio obtained from Chevron plots (*8*, *42*) (Figure 4f) suggests a more compact intermediate state for ISP1 (0.7716) than for Pro-SbtE (0.536), probably due to presence of the propeptide. Nonetheless, ISP1 folds to its thermodynamically stable native state ($\Delta G_{NU} = 5.3$ kcal/mol) through a compact transient intermediate, while pro-SbtE folds and autoprocesses (*34*) into a stable stoichiometric complex ($\Delta G_{NU} = 9.9$ kcal/mol) via an expanded intermediate (Table 1). Degradation of the inhibitory propeptide from this complex results in a kinetically trapped protease domain (*33*, *40*) whose unfolding does not show a stable intermediate (Table 1). Although propeptide degradation lowers the thermodynamic stability of SbtE (Table 1; $\Delta G_{release} = 10.4$ kcal/mol), it ensures proteolytic stability (*13*, *21*, *40*) by increasing the unfolding energy barrier (Figure 5). It is important to note that the observed differences in the folding pathways of ISP1 and SbtE are not mere consequences of their cellular location, because the intracellular or extracellular expression of SbtE does not yield active protease in the absence of its cognate propeptide (*60*). Hence, ISP1 and SbtE have similar sequences, structures, and catalytic activities but fold through significantly different pathways. This suggests that the inability to spontaneously fold may not be a property of that particular scaffold, but may be a consequence of fine sequence details.

Table 1: Comparison of ISP1, SbtE, and Pro-SbtE from *B. subtilis* as Models for ISPs and ESPs

| | ISP1 | SbtE | pro-SbtE |
|---|---|---|---|
| **Sequence** | | | |
| identity with ISP1 (%) | 100 | 50 | 50 |
| total length (no. of residues) | 319 | 275 | 352 |
| propeptide (no. of residues) | 17 | 0 | 77 |
| isoelectric point (pI) | 4.7 | 6.3 | 8.5 |
| charged residues (%) | 23.10 | 9.40 | 14.77 |
| contact order (CO) | 27.36 | 28.41 | NA |
| hydrophobic core (HCI) | 5.36 | 6.04 | NA |
| **Activity** | | | |
| optimum pH | 8.5 | 8.5 | NA[a] |
| $K_I$ (nM) for SSI | 0.17 | 0.26 | NA[a] |
| $K_I$ (nM) for ProS[b] | 1410 | 18.50 | NA[a] |
| $K_I$ (nM) for BSA | 4700 | 2140 | NA[a] |
| specific activity$_{N-succ-AAPFpNA}$ (units/ng) | 2698 | 210 | NA[a] |
| **Structure** | | | |
| helix (%) | 21.0 | 25.5 | 22.4 |
| sheet (%) | 26.4 | 22.8 | 21.7 |
| $\beta$-turn (%) | 17.3 | 17.4 | 17.4 |
| random coil (%) | 35.3 | 34.3 | 38.8 |
| $T_m$ (°C) | 58.7 | 58.3 | 54.5 |
| **Equilibrium Unfolding** | | | |
| $\Delta G_{NI}$ (kcal/mol) | 3.56 | NA | 1.62 |
| $m_{NI}$ (kcal mol$^{-1}$ M$^{-1}$) | 7.14 | NA | 1.11 |
| $\Delta G_{NU}$ (kcal/mol) | 5.3 | 2.433[c] | 9.98 |
| $m_{NU}$ (kcal mol$^{-1}$ M$^{-1}$) | 7.85 | 0.73[c] | 3.02 |

[a] The cleaved complex (pro-SbtE) cannot bind inhibitors. [b] $\Delta G_{release}$ for pro from SbtE $= -RT \ln K_I$ (10.4 kcal/mol). [c] Values are illustrative (SbtE unfolding is irreversible).

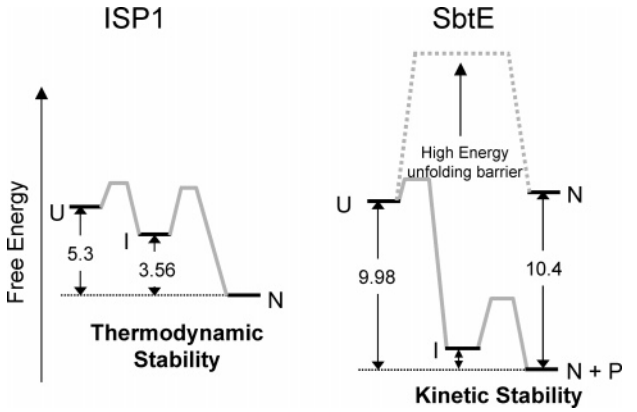


FIGURE 5: Representative folding energy landscape of ISPs and ESPs. Representative free energy landscape of kinetically and thermodynamically driven protein folding. Denatured ISP1 (U) folds through an intermediate (I) to its thermodynamically stable native state (N). Pro-SbtE folds into a thermodynamically stable complex (N + P). Propeptide degradation from this complex kinetically traps the native state (N) by increasing the unfolding energy barrier and by uncoupling folding (solid gray line) and unfolding (dashed gray line) pathways.

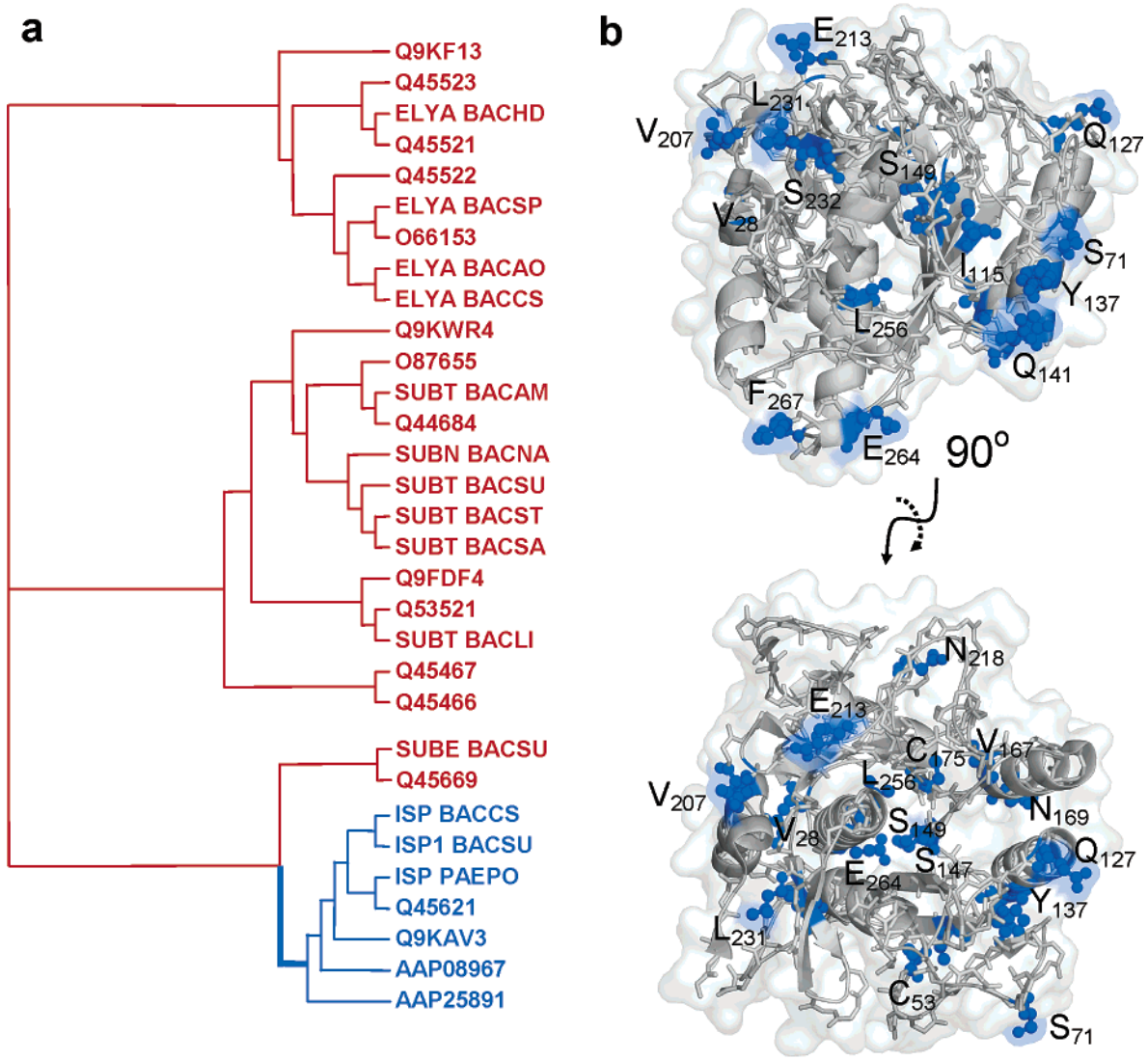


FIGURE 6: Adaptive evolution of ISPs. (a) Phylogeny of ISPs (blue) and ESPs (red) from *Bacillus*. Branch lengths are not to scale. The lineage leading to ISPs (thick blue line) was tested for positive selection. (b) Positively selected residues (posterior probability ≥ 0.95) are mapped onto the ISP1 structure.

*Evolutionary Analysis of Bacillus ESPs and ISPs.* To understand how two conserved proteins have evolved to fold differently, and to establish the constraints driving such evolution, we performed a phylogeny-based statistical analysis on *Bacillus* ESPs and ISPs, to test for positive selection along the ISP subfamily (Figure 6a). The ratio of the rate of nonsynonymous (d*n*) to synonymous substitution (d*s*), $\omega$, is an unequivocal indicator of evolution indicating neutral ($\omega = 1$), purifying ($\omega < 1$), or positive diversifying ($\omega > 1$) selection (*46*, *49*). Our data set was analyzed using various codon-based site and branch site models in PAML (*46*) that estimate $\omega$ using the maximum likelihood (see Materials and Methods). The major advantage with this analysis is that it allows for variable selective pressures along a gene, by assuming different classes of sites with varying $\omega$ ratios (*46*, *49*). Site specific models allow $\omega$ to vary among sites, but not among branches. Branch site models are more powerful in detecting positive selection as they allow variations in $\omega$ between both various sites along the sequence and different lineages along the tree (*48*). Briefly, the site models were M0 (one $\omega$ for all sites), M1 (allows $\omega = 1$ and $\omega = 0$), M2 ($\omega = 1$, $\omega = 0$, and $\omega > 1$), M3 (discrete model, with three site classes for $\omega$ estimated from data), M7 ($\beta$ distribution for $\omega$ from 0 to 1), and M8 ($\beta$ distribution from 0 to 1, and independent $\omega$ estimated from data). The branch site models were model A and model B (four different classes of $\omega$; see Materials and Methods). A likelihood ratio test (LRT) is used to search for positive selection, that is, for presence of sites for which $\omega > 1$. This is achieved by comparing a null model that does not account for $\omega > 1$ with a more general model that allows for $\omega > 1$. For example, M0, which allows one $\omega$ for all sites, was compared with M3, which is a discrete model with three site classes for $\omega$ estimated from data. Positive selection is indicated when $\omega > 1$ and the LRT is significant (see Materials and Methods). When LRTs suggest positive selection, the posterior probabilities for each site to have a specific $\omega$ class can be calculated using Bayes theorem. Our data set was analyzed as described in Materials and Methods, and the estimated parameters are listed in Table 2.

Log likelihoods for both branch site models show significantly better fits ($p < 0.05$) than their null models (model A with M1, model B with M3) and show positive selection along the ISP clade (Figure 6a and Table 2). Both models show that although most sites are neutral or conserved during evolution ($\omega \leq 1$), ~22% of sites exhibit strong positive

Table 2: Phylogeny-Based Statistical Analysis of *Bacillus* ESPs and ISPs Using Maximum Likelihood Models

| model | likelihood | $\omega$(d*n*/ds) | positive selection and sites selected[a] |
|---|---|---|---|
| site models | | | |
| M0 | −16311.68 | 0.038 | not allowed |
| M1 | −17475.09 | $P_0 = 0.112$, $p_1 = 0.888$, $\omega_0 = 0.000$, $\omega_1 = 1.000$ | not allowed |
| M2 | −17469.63 | $P_0 = 0.113$, $p_1 = 0.843$, $p_2 = 0.046$, $\omega_0 = 0.000$, $\omega_1 = 1.000$,$\omega_2 = 2.172$ | yes, 169N, **231L**[b] |
| M3 | −15873.44 | $P_0 = 0.543$, $p_1 = 0.457$, $p_2 = 0.000$, $\omega_0 = 0.013$, $\omega_1 = 0.080$, $\omega_2 = 6.845$ | no |
| M7 | −15815.78 | $P = 0.737$, $q = 14.409$ | not allowed |
| M8 ($n_{cat} = 5$) | −15817.58 | $P_0 = 1.000$, $p = 0.709$, $q = 13.607$ ($p_1 = 0.000$), $\omega_1 = 1.383$ | no |
| branch site models | | | |
| A | −17422.28 | $P_0 = 0.093$, $p_1 = 0.664$, $p_2 = 0.030$, $p_3 = 0.214$, $\omega_0 = 0.000$, $\omega_1 = 1.000$, $\omega_2 = 999.00$ | yes, 14 sites (**28V**, 53C, **127Q**,[b] **137Y**,[b] 147S,[b] **149S**,[b] 167V, 175C, **207V**,[b] **213E**,[b] **218N**, **232S**, **264E**,[b] **267F**) |
| B | −15856.14 | $p_0 = 0.347$, $p_1 = 0.418$, $p_2 = 0.107$, $p_3 = 0.129$, $\omega_0 = 0.082$, $\omega_1 = 0.0128$, $\omega_2 = 24.447$ | yes, 12 sites (53C, **71S**,[b] 115I,[b] **137Y**,[b] **141Q**, **149S**,[b] 175C, **218N**, **232S**, 256L, **264E**,[b] **267F**) |

[a] Only those sites with a posterior probability of >0.95 for positive selection are shown. Residues in boldface type are those on protein surface and constitute 75% of positively selected residues. There is a >55% overlap between residues identified from branch site model A and model B. [b] Mutations at these sites have been shown to affect the folding, stability, and activity of subtilisins through independent directed evolutionary studies (*50*, *61*, *79*).

Table 3: Positively Selected Sites in ISPs, Identified through Phylogeny Analysis, Corresponding Sites in ESPs and Their Functional Roles from Independent Mutagenesis Studies in ESPs (*50*)

| ISP1 | SbtE | implicated role in subtilisins |
|---|---|---|
| 28V | 10Q | not determined |
| 53C | 35I | not determined |
| 71S | 51V | stability |
| 115I | 92A | activity, stability |
| 127Q | 103Q | activity, affects flexibility |
| 137Y | 113W | folding, stability |
| 141Q | 117N | not determined |
| 147S | 123N | activity, affects flexibility |
| 149S | 125S | activity |
| 167V | 143V | not determined |
| 175C | 151A | not determined |
| 207V | 182S | activity, stability |
| 213E | 188S | folding, stability, thermostability |
| 218N | 193G | not determined |
| 231L | 206Q | stability |
| 232S | 207S | not determined |
| 256L | 231A | not determined |
| 264E | 239P | activity, stability |
| 267F | 242T | not determined |

selection ($\omega > 1$) within ISPs. Hence, while subtilases are well-conserved, it appears that there has been significant adaptive evolution preceding ISPs, to allow propeptide-independent protein folding. We next examined posterior probabilities using an empirical Bayes approach (*46*, *49*) to identify positively selected sites (Table 2). It is important to note that 75% of the sites with a probability of >0.95 are confined to the surface of ISP1 (Figure 6b and Table 2). Because of their industrial importance, ESPs have been extensively analyzed using mutagenesis and directed evolution, and sites that directly affect folding, activity, and stability have been identified (*50*). Interestingly, several positively selected sites identified in our evolutionary analysis overlap with sites determined through such studies (Tables 2 and 3) and are further consistent with earlier findings from our laboratory (*61*). For example, residue 188 in ESPs (corresponding to site 213 in ISP1) has been established to interact with residue 17 (earlier termed position −60) in the propeptide domain during folding (*61*). This site appears to have positively evolved within propeptide-independent ISPs. Hence, while kinetics (Figure 4) establishes that the structurally conserved proteases SbtE and ISP1 (Figures 1−3) can fold through different pathways, our evolutionary analysis (Figure 6 and Table 2) suggests that this selection of pathways may be mediated through positive selection of specific residues.

## DISCUSSION

Subtilases constitute a ubiquitous superfamily of serine proteases, with most members being synthesized with N-terminal propeptide extensions (*11*, *32*). X-ray crystallography and homology modeling concur that catalytic domains of subtilases adopt very similar three-dimensional structures (*32*). The contemporary view is that structurally conserved subtilases should fold into their native states through similar transition states and folding pathways (*3*−*5*, *62*−*65*). This is consistent with extensive work on prokaryotic subtilisins and eukaryotic subtilisin/kexin-like prohormone convertases that emphasize the importance and necessity of propeptide domains in facilitating correct protein folding (*11*, *13*). Much of our mechanistic understanding of propeptide-mediated protein folding has emerged from analysis of the bacterial ESPs (*13*, *32*, *35*−*37*, *39*, *40*, *61*, *66*−*72*), which are extracellular proteases. Detailed *in vitro* and *in vivo* studies have demonstrated that propeptides can direct folding of ESPs into kinetically stable states and are not mere facilitators of correct localization (*11*, *60*, *73*). Propeptides remain tightly associated with their cognate ESPs and must be proteolytically degraded to release enzymatic activity (*57*). Cleavage and degradation of the propeptide at the completion of maturation enable the decoupling of folding and unfolding pathways (*40*) and stabilize the native state by increasing the energy of the transition state for the unfolding reaction (Figure 5). Furthermore, proteolytic cleavage also facilitates formation of high-affinity calcium binding site A, which contributes significantly to the increased energy of the unfolding transition state in ESPs (*32*).

A comprehensive database search for SbtE-like bacterial proteases helped identify two distinct subfamilies (ISPs and ESPs) within the family of subtilases that exhibit significant sequence, structure, and functional conservation (*11*, *32*). While ESPs and ISPs display a high level of sequence identity (Figure 1), the latter represents an intracellular subfamily that lacks the classical subtilisin propeptide signature (*54*), which is essential for correct folding of ESPs. The lack of the propeptide signature suggests that ISPs might fold in a manner independent of propeptide domains. This work represents the first description of the propeptide-independent folding pathway of a $Ca^{2+}$-dependent subtilase and a detailed analysis of evolutionary constraints that affect the choice between thermodynamic and kinetic folding and stability.

Figure 1 establishes that the sequences of ESPs and ISPs are approximately 50% identical. A striking feature of this alignment is that specific residues that are conserved within one family are changed and display significantly different amino acid properties from the other. For example, the loop connecting $\beta$5 with $\beta$6 has $D_{182}S$, $E_{189}T$, and $E_{198}S$ sites that change from polar (ESPs) to acidic (ISPs). Similarly, $R_{136}E$, $Y_{137}W$, $E_{158}A$, and $K_{164}D$ are substitutions that are localized within helices $\alpha$4 and $\alpha$5, which contribute interactions to the propeptide−subtilisin interface in ESPs. Specifically, mutations at site 112 in SbtE (corresponding to site 136 in ISPs; $R_{136}E$ substitution) have been shown to significantly lower the binding affinity of SbtE for the propeptide (*36*). Thus, while the two subfamilies have conserved structures, there are specific sites that appear to be selected during their evolution. Interestingly, such sites map mostly to the solvent accessible surface of the protein (Figures 2c and 6b). Furthermore, the level of sequence identity and/or conservation within the hydrophobic core is significant, and quantitation suggests that the average strength of the hydrophobic core, which is the major driving force in protein folding reactions (*74*−*76*), is similar in ESPs and ISPs (Table 1 and Figure 2b).

In this work, ISP1 was extensively studied and compared with our ESP folding model, SbtE. The CD spectroscopic analysis demonstrates that ISP1 and SbtE adopt very similar secondary structures (Figure 3a) and exhibit similar thermostability (Table 1). The inability to spontaneously fold has been attributed to the high stability of ESPs bestowed by the presence of a high-affinity calcium binding site (*32*). A calcium deletion variant of ESPs can fold in a manner independent of its propeptide (*13*, *39*), albeit very slowly (*67*−*69*), and its folding is catalyzed more than $10^6$-fold by the addition of the propeptide (*67*). The sequence alignment of ESPs and ISPs indicates significant conservation among residues that constitute the $Ca^{2+}$ binding site. Consistent with this conservation (Figure 1), ISP1 displays a strong calcium-dependent proteolytic stability. However, the affinity of ISP1 for calcium is approximately 3-fold lower than that of SbtE, which may be due to an insertion prior to $\beta$2 in ISP1 (Figure 1). It is important to note that the presence of this $Ca^{2+}$ binding site does not preclude ISP1 from spontaneous folding and unfolding (Figure 4c,e) independent of any propeptide domain.

The propeptides of subtilases are potent inhibitors of protease activity and exhibit significant cross reactivity. Therefore, the propeptide of SbtE can bind tightly with other ESPs such as subtilisin Carlsberg, whose sequence is ~60% identical. However, the SbtE propeptide is not a good inhibitor of ISP1. On the other hand, SSI, a potent subtilisin inhibitor, binds to both SbtE and ISP1 with similar affinities (Figure 4b and Table 1). The weak affinity of ISP1 for the propeptide, which is similar to the affinity of a nonspecific substrate BSA (Figure 4b), can be attributed to distinct differences in the mode of interaction (*11*, *34*). The propeptide binds with the protease domain in a "side-on" orientation through an interface mediated by helices $\alpha$4 and $\alpha$5 (*34*). On the other hand, SSI binds to the active site of subtilisin in a "top-on" orientation (*77*) through interactions that are completely different from that of the propeptide. Interestingly, the sequence alignment and the modeled structure establish that these helices ($\alpha$4 and $\alpha$5) in ISPs exhibit differences in key residues (Figure 2a,c) that are essential for high-affinity binding to the propeptide (*35*). However, the sequences of these helices are >90% identical within ESPs (Figure 1), and as a consequence, the propeptide can strongly inhibit other ESPs.

Comparative studies demonstrate that folding of ISP1 is propeptide-independent and thermodynamically reversible with a $\Delta G_{UF}$ of −5.3 kcal/mol, while folding or unfolding of the protease domain of SbtE is thermodynamically reversible only in the presence of its cognate propeptide with a $\Delta G_{UF}$ of approximately −10.0 kcal/mol. Interestingly, equilibrium folding−unfolding reactions of ISP1 and pro-SbtE indicate the presence of distinct protein folding intermediates that are maximally populated in 0.8 and 2.5 M GdnHCl, respectively (Figure 4e). The dependence of the apparent folding rates on denaturant (Figure 4f) shows a rollover in the unfolding limbs for both proteins and supports the presence of stable intermediates in the folding pathway. Furthermore, the extent of solvent exposure obtained from the chevron plots (Figure 4f) suggests that the folding intermediate of ISP1 is more compact, in contrast with the diffused state in SbtE. The presence of partially structured intermediates is in contrast with the hypothesis that broad specificity proteases such as subtilisin would select against folding via stable intermediates, because they would be excellent substrates for autolysis (*13*). While such partially structured states are commonly found in several protein models and are important intermediates in the overall folding pathway, their precise role in folding of ISP1 and pro-SbtE remains unknown.

Our studies establish that, SbtE, an ESP that is secreted in harsh, extracellular, protease-rich environments, folds to its native state with the help of a propeptide domain (*11*, *12*, *35*−*37*, *40*, *61*, *66*, *78*). The presence of the propeptide domain enables it to reach a kinetically trapped native state that has a high unfolding energy barrier and thus offers greater proteolytic stability. On the other hand, the intracellular homologue, ISP1, with similar sequence, topology, hydrophobic core, and catalytic activity, can fold to its native state even without the help of a propeptide domain. Further, it reaches a thermodynamically stable native state through a more compact folding intermediate. While the contemporary view is that homologous proteins fold through similar pathways and folding transition states (*3*−*5*, *62*−*65*), our results demonstrate that ISP1 (propeptide-independent) and SbtE (propeptide-dependent) reach similar native states through different folding pathways. Evolutionary analysis of

the constraints underlying the evolution of such closely related subfamilies shows that the ISPs have indeed positively diverged from the ESPs.

It is important to note that the observed differences in folding pathways of ISP1 and SbtE are not mere consequences of their cellular location, because the intracellular or extracellular expression of SbtE does not yield an active protease in the absence of its cognate propeptide (*11*, *73*). While the kinetic barriers apparently enhance the stability of ESPs in harsh extracellular environments (*13*, *21*), similar barriers leading to high stability in ISPs may impede intracellular protein turnover. This is consistent with our finding that intracellular expression of ESPs (aqualysin 1) is detrimental to cell growth due to extensive proteolysis (unpublished data). This suggests that biological requirements, and not just specific conformations, dictate selection of folding pathways. A crucial question is how two homologous subtilase subfamilies have evolved to select different folding pathways. Several studies demonstrate the importance of topology, contact order, hydrophobicity, and stability in driving protein folding pathways and kinetics (*2*−*5*). Interestingly, the properties listed above appear to be conserved within ISPs and ESPs, and differences within their sequences map specifically onto the protein surface. Furthermore, evolutionary analysis shows that adaptive evolution, which precedes ISPs, is predominantly through positive selection of specific surface residues. These differences may allow ISPs to fold in a propeptide-independent, thermodynamically driven pathway.

Another important question is how changes in surface residues compensate for the loss of propeptides within ISPs. Both ESPs and ISPs are highly charged proteins but with extremely different isoelectric points (Table 1). Moreover, this charge is localized within propeptides of ESPs, while it is distributed over the protease domain of ISPs. As a consequence, protease domains of ISPs are more polar than ESPs. This difference may help to enhance the conformational entropy of water around the folding polypeptide and assist the hydrophobic core in driving spontaneous folding of ISPs. When folding of ESPs is carried out in the absence of the propeptide, the protease domain adopts a partially structured molten-globule intermediate with solvent-exposed hydrophobic surfaces (*38*). The partially formed hydrophobic core within the protease domain of ESPs may be insufficient to drive folding without the solvation assistance bestowed by the surface charge. This leads to stabilization of the molten-globule intermediate that is prone to aggregation. This concept is consistent with our findings that charge per se, and not its polarity, is critical for folding of subtilases (*35*, *78*). This result, to our knowledge, is the first demonstration that surface residues and their charge may influence the selection of folding pathways and kinetics, and fine sequence details may dictate selection between kinetically and thermodynamically driven protein folding.

## ACKNOWLEDGMENT

We thank Martine Bouton-Landais for technical assistance and Gary Thomas, David Farrens, Svetlana Lutsenko, and Mihail Iordanov for discussions and comments on the manuscript.

## REFERENCES

1. Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation, *Semin. Cell Dev. Biol. 15*, 3−16.
2. Gunasekaran, K., Eyles, S. J., Hagler, A. T., and Gierasch, L. M. (2001) Keeping it in the family: folding studies of related proteins, *Curr. Opin. Struct. Biol. 11*, 83−93.
3. Nishimura, C., Prytulla, S., Jane Dyson, H., and Wright, P. E. (2000) Conservation of folding pathways in evolutionarily distant globin sequences, *Nat. Struct. Biol. 7*, 679−686.
4. Shakhnovich, E., Abkevich, V., and Ptitsyn, O. (1996) Conserved residues and the mechanism of protein folding, *Nature 379*, 96−98.
5. Plaxco, K. W., Simons, K. T., Ruczinski, I., and Baker, D. (2000) Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics, *Biochemistry 39*, 11177−11183.
6. Mirny, L., and Shakhnovich, E. (2001) Evolutionary conservation of the folding nucleus, *J. Mol. Biol. 308*, 123−129.
7. Larson, S. M., Ruczinski, I., Davidson, A. R., Baker, D., and Plaxco, K. W. (2002) Residues participating in the protein folding nucleus do not exhibit preferential evolutionary conservation, *J. Mol. Biol. 316*, 225−233.
8. Raschke, T. M., and Marqusee, S. (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions, *Nat. Struct. Biol. 4*, 298−304.
9. Makarov, D. E., and Plaxco, K. W. (2003) The topomer search model: A simple, quantitative theory of two-state protein folding kinetics, *Protein Sci. 12*, 17−26.
10. Tseng, Y. Y., and Liang, J. (2004) Are residues in a protein folding nucleus evolutionarily conserved? *J. Mol. Biol. 335*, 869−880.
11. Shinde, U., and Inouye, M. (2000) Intramolecular chaperones: polypeptide extensions that modulate protein folding, *Semin. Cell Dev. Biol. 11*, 35−44.
12. Shinde, U. P., Liu, J. J., and Inouye, M. (1997) Protein memory through altered folding mediated by intramolecular chaperones, *Nature 389*, 520−522.
13. Bryan, P. N. (2002) Prodomains and protein folding catalysis, *Chem. Rev. 102*, 4805−4816.
14. Sohl, J. L., Jaswal, S. S., and Agard, D. A. (1998) Unfolded conformations of α-lytic protease are more stable than its native state, *Nature 395*, 817−819.
15. Cunningham, E. L., and Agard, D. A. (2003) Interdependent folding of the N- and C-terminal domains defines the cooperative folding of α-lytic protease, *Biochemistry 42*, 13212−13219.
16. Cunningham, E. L., and Agard, D. A. (2004) Disabling the folding catalyst is the last critical step in α-lytic protease folding, *Protein Sci. 13*, 325−331.
17. Huntington, J. A., Read, R. J., and Carrell, R. W. (2000) Structure of a serpin-protease complex shows inhibition by deformation, *Nature 407*, 923−926.
18. Im, H., Seo, E. J., and Yu, M. H. (1999) Metastability in the inhibitory mechanism of human α1-antitrypsin, *J. Biol. Chem. 274*, 11072−11077.
19. Baldwin, T. O., Ziegler, M. M., Chaffotte, A. F., and Goldberg, M. E. (1993) Contribution of folding steps involving the individual subunits of bacterial luciferase to the assembly of the active heterodimeric enzyme, *J. Biol. Chem. 268*, 10766−10772.
20. Carr, C. M., Chaudhry, C., and Kim, P. S. (1997) Influenza hemagglutinin is spring-loaded by a metastable native conformation, *Proc. Natl. Acad. Sci. U.S.A. 94*, 14306−14313.
21. Jaswal, S. S., Sohl, J. L., Davis, J. H., and Agard, D. A. (2002) Energetic landscape of α-lytic protease optimizes longevity through kinetic stability, *Nature 415*, 343−346.
22. Baker, D., and Agard, D. A. (1994) Kinetics versus thermodynamics in protein folding, *Biochemistry 33*, 7505−7509.
23. Berkenpas, M. B., Lawrence, D. A., and Ginsburg, D. (1995) Molecular evolution of plasminogen activator inhibitor-1 functional stability, *EMBO J. 14*, 2969−2977.
24. Lawrence, D. A., Ginsburg, D., Day, D. E., Berkenpas, M. B., Verhamme, I. M., Kvassman, J. O., and Shore, J. D. (1995) Serpin-protease complexes are trapped as stable acyl-enzyme intermediates, *J. Biol. Chem. 270*, 25309−25312.
25. Im, H., Woo, M. S., Hwang, K. Y., and Yu, M. H. (2002) Interactions causing the kinetic trap in serpin protein folding, *J. Biol. Chem. 277*, 46347−46354.
26. Swalley, S. E., Baker, B. M., Calder, L. J., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (2004) Full-length influenza hemagglutinin

HA2 refolds into the trimeric low-pH-induced conformation, *Biochemistry 43*, 5902−5911.
27. Shental-Bechor, D., Danieli, T., Henis, Y., and Ben-Tal, N. (2002) Long-range effects on the binding of the influenza HA to receptors are mediated by changes in the stability of a metastable HA conformation, *Biochim. Biophys. Acta 1565*, 81−89.
28. Light, A., and al-Obeidi, A. M. (1991) Further evidence for independent folding of domains in serine proteases, *J. Biol. Chem. 266*, 7694−7698.
29. Cunningham, E. L., Jaswal, S. S., Sohl, J. L., and Agard, D. A. (1999) Kinetic stability as a mechanism for protease longevity, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11008−11014.
30. Cunningham, E. L., Mau, T., Truhlar, S. M., and Agard, D. A. (2002) The pro region N-terminal domain provides specific interactions required for catalysis of α-lytic protease folding, *Biochemistry 41*, 8860−8867.
31. Eder, J., and Fersht, A. R. (1995) Pro-sequence-assisted protein folding, *Mol. Microbiol. 16*, 609−614.
32. Siezen, R. J., and Leunissen, J. A. (1997) Subtilases: the superfamily of subtilisin-like serine proteases, *Protein Sci. 6*, 501−523.
33. Eder, J., Rheinnecker, M., and Fersht, A. R. (1993) Folding of subtilisin BPN′: role of the pro-sequence, *J. Mol. Biol. 233*, 293−304.
34. Jain, S. C., Shinde, U., Li, Y., Inouye, M., and Berman, H. M. (1998) The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 Å resolution, *J. Mol. Biol. 284*, 137−144.
35. Yabuta, Y., Subbian, E., Oiry, C., and Shinde, U. (2003) Folding pathway mediated by an intramolecular chaperone. A functional peptide chaperone designed using sequence databases, *J. Biol. Chem. 278*, 15246−15251.
36. Fu, X., Inouye, M., and Shinde, U. (2000) Folding pathway mediated by an intramolecular chaperone. The inhibitory and chaperone functions of the subtilisin propeptide are not obligatorily linked, *J. Biol. Chem. 275*, 16871−16878.
37. Yabuta, Y., Subbian, E., Takagi, H., Shinde, U., and Inouye, M. (2002) Folding pathway mediated by an intramolecular chaperone: dissecting conformational changes coincident with autoprocessing and the role of $Ca^{2+}$ in subtilisin maturation, *J. Biochem. 131*, 31−37.
38. Shinde, U., and Inouye, M. (1995) Folding pathway mediated by an intramolecular chaperone: characterization of the structural changes in pro-subtilisin E coincident with autoprocessing, *J. Mol. Biol. 252*, 25−30.
39. Bryan, P., Alexander, P., Strausberg, S., Schwarz, F., Lan, W., Gilliland, G., and Gallagher, D. T. (1992) Energetics of folding subtilisin BPN′, *Biochemistry 31*, 4937−4945.
40. Yabuta, Y., Takagi, H., Inouye, M., and Shinde, U. (2001) Folding pathway mediated by an intramolecular chaperone: propeptide release modulates activation precision of pro-subtilisin, *J. Biol. Chem. 276*, 44427−44434.
41. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res. 31*, 3381−3385.
42. Takei, J., Chu, R. A., and Bai, Y. (2000) Absence of stable intermediates on the folding pathway of barnase, *Proc. Natl. Acad. Sci. U.S.A. 97*, 10796−10801.
43. Fujiwara, K., Arai, M., Shimizu, A., Ikeguchi, M., Kuwajima, K., and Sugai, S. (1999) Folding-unfolding equilibrium and kinetics of equine *β*-lactoglobulin: equivalence between the equilibrium molten globule state and a burst-phase folding intermediate, *Biochemistry 38*, 4455−4463.
44. Eder, J., Rheinnecker, M., and Fersht, A. R. (1993) Folding of subtilisin BPN′: characterization of a folding intermediate, *Biochemistry 32*, 18−26.
45. Mizuguchi, M., Arai, M., Ke, Y., Nitta, K., and Kuwajima, K. (1998) Equilibrium and kinetics of the folding of equine lysozyme studied by circular dichroism spectroscopy, *J. Mol. Biol. 283*, 265−277.
46. Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood, *Comput. Appl. Biosci. 13*, 555−556.
47. Yang, Z. (2002) Inference of selection from multiple species alignments, *Curr. Opin. Genet. Dev. 12*, 688−694.
48. Bielawski, J. P., and Yang, Z. (2003) Maximum likelihood methods for detecting adaptive evolution after gene duplication, *J. Struct. Funct. Genomics 3*, 201−212.
49. Swanson, W. J., Yang, Z., Wolfner, M. F., and Aquadro, C. F. (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals, *Proc. Natl. Acad. Sci. U.S.A. 98*, 2509−2514.
50. Bryan, P. N. (2000) Protein engineering of subtilisin, *Biochim. Biophys. Acta 1543*, 203−222.
51. Baker, D., and Sali, A. (2001) Protein structure prediction and structural genomics, *Science 294*, 93−96.
52. Henrich, S., Cameron, A., Bourenkov, G. P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W., and Than, M. E. (2003) The crystal structure of the proprotein processing proteinase furin explains its stringent specificity, *Nat. Struct. Biol. 10*, 520−526.
53. Holyoak, T., Wilson, M. A., Fenn, T. D., Kettner, C. A., Petsko, G. A., Fuller, R. S., and Ringe, D. (2003) 2.4 Å resolution crystal structure of the prototypical hormone-processing protease Kex2 in complex with an Ala-Lys-Arg boronic acid inhibitor, *Biochemistry 42*, 6709−6718.
54. Shinde, U., and Inouye, M. (1993) Intramolecular chaperones and protein folding, *Trends Biochem. Sci. 18*, 442−446.
55. Shinde, U., Li, Y., Chatterjee, S., and Inouye, M. (1993) Folding pathway mediated by an intramolecular chaperone, *Proc. Natl. Acad. Sci. U.S.A. 90*, 6924−6928.
56. Greenfield, N., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry 8*, 4108−4116.
57. Li, Y., Hu, Z., Jordan, F., and Inouye, M. (1995) Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants, *J. Biol. Chem. 270*, 25127−25132.
58. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223−230.
59. Jackson, S. E., and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry 30*, 10428−10435.
60. Zhu, X. L., Ohta, Y., Jordan, F., and Inouye, M. (1989) Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process, *Nature 339*, 483−484.
61. Inouye, M., Fu, X., and Shinde, U. (2001) Substrate-induced activation of a trapped IMC-mediated protein folding intermediate, *Nat. Struct. Biol. 8*, 321−325.
62. Plaxco, K. W., Larson, S., Ruczinski, I., Riddle, D. S., Thayer, E. C., Buchwitz, B., Davidson, A. R., and Baker, D. (2000) Evolutionary conservation in protein folding kinetics, *J. Mol. Biol. 298*, 303−312.
63. Calloni, G., Taddei, N., Plaxco, K. W., Ramponi, G., Stefani, M., and Chiti, F. (2003) Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding, *J. Mol. Biol. 330*, 577−591.
64. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature 416*, 507−511.
65. Friel, C. T., Capaldi, A. P., and Radford, S. E. (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins, *J. Mol. Biol. 326*, 293−305.
66. Shinde, U., Fu, X., and Inouye, M. (1999) A pathway for conformational diversity in proteins mediated by intramolecular chaperones, *J. Biol. Chem. 274*, 15615−15621.
67. Ruan, B., Hoskins, J., and Bryan, P. N. (1999) Rapid folding of calcium-free subtilisin by a stabilized pro-domain mutant, *Biochemistry 38*, 8562−8571.
68. Ruan, B., Hoskins, J., Wang, L., and Bryan, P. N. (1998) Stabilizing the subtilisin BPN′ pro-domain by phage display selection: how restrictive is the amino acid code for maximum protein stability? *Protein Sci. 7*, 2345−2353.
69. Ruvinov, S., Wang, L., Ruan, B., Almog, O., Gilliland, G. L., Eisenstein, E., and Bryan, P. N. (1997) Engineering the independent folding of the subtilisin BPN′ prodomain: analysis of two-state folding versus protein stability, *Biochemistry 36*, 10414−10421.
70. Gallagher, T., Gilliland, G., Wang, L., and Bryan, P. (1995) The prosegment-subtilisin BPN′ complex: crystal structure of a specific 'foldase', *Structure 3*, 907−914.
71. Buevich, A. V., Shinde, U. P., Inouye, M., and Baum, J. (2001) Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies, *J. Biomol. NMR 20*, 233−249.

72. Strausberg, S., Alexander, P., Wang, L., Schwarz, F., and Bryan, P. (1993) Catalysis of a protein folding reaction: thermodynamic and kinetic analysis of subtilisin BPN′ interactions with its propeptide fragment, *Biochemistry 32*, 8112−8119.
73. Ikemura, H., Takagi, H., and Inouye, M. (1987) Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*, *J. Biol. Chem. 262*, 7859−7864.
74. Agashe, V. R., Shastry, M. C., and Udgaonkar, J. B. (1995) Initial hydrophobic collapse in the folding of barstar, *Nature 377*, 754−757.
75. Ptitsyn, O. B. (1995) How the molten globule became, *Trends Biochem. Sci. 20*, 376−379.
76. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol. 4*, 10−19.
77. Takeuchi, Y., Satow, Y., Nakamura, K. T., and Mitsui, Y. (1991) Refined crystal structure of the complex of subtilisin BPN′ and *Streptomyces* subtilisin inhibitor at 1.8 Å resolution, *J. Mol. Biol. 221*, 309−325.
78. Marie-Claire, C., Yabuta, Y., Suefuji, K., Matsuzawa, H., and Shinde, U. (2001) Folding pathway mediated by an intramolecular chaperone: the structural and functional characterization of the aqualysin I propeptide, *J. Mol. Biol. 305*, 151−165.
79. Kobayashi, T., and Inouye, M. (1992) Functional analysis of the intramolecular chaperone. Mutational hot spots in the subtilisin pro-peptide and a second-site suppressor mutation within the subtilisin molecule, *J. Mol. Biol. 226*, 931−933.

BI048397X