Exploring the sequence and structural determinants of the energy landscape from thermodynamically stable and kinetically trapped subtilisins: ISP1 and SbtE

Miriam R. Hood[1], Susan Marqusee[1,2,3,#]

1. Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720
2. Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720
3. California Institute for Quantitative Biosciences (QB3), Berkeley, Berkeley CA 94720

Corresponding author: Susan Marqusee, 576 Stanley Hall, UC Berkeley, Berkeley CA 94720-3320

Abstract

A protein's energy landscape, all the accessible conformations, their populations, and their dynamics of interconversion, is encoded in its primary sequence. While we have a good understanding of how a protein's primary sequence encodes its native state, we have a much weaker understanding of how sequence encodes the kinetic barriers such as unfolding and refolding. Here we have looked at two subtiliase homologs from the *Bacillus subtilis*, Intracellular Subtilisin Protease 1 (ISP1) and Subtilisin E (SbtE) that are expected to have very different dynamics. As an intracellular protein, ISP1 has a small pro-domain thought to act simply as a zymogen, whereas the extracellular SbtE has a large pro-domain required for folding. We examined the global and local energetics of the mature proteases and how each pro-domain impacts their landscapes. We find that ISP1's pro-domain has limited impact on the energy landscape while the mature SbtE is thermodynamically unstable and kinetically trapped. The impact of the pro-domain has opposite effects on the flexibility of the core of the protein. ISP1's core becomes more flexible while SbtE's core becomes more rigid. ISP1 contains a conserved amino-acid insertion not present in extracellular subtilisin proteases, which points to a potential source for these differences. These homologs are an extreme example of how changes in the primary sequence can dramatically alter a proteins energy landscape, both stability and dynamics, and highlight the need for large scale, high throughput studies on the relationship between primary sequence and conformational dynamics.

Introduction:

Proteins are an essential group of macromolecules for life, playing important roles in nearly all biological functions. For most proteins, the active conformation is a compact, folded structure whose topology and dynamics are finely tuned and encoded in the linear sequence of amino acids. Protein folding, the process needed to obtain this structure, must take place on a timescale consistent with its environment and function. Similarly, once folded, both the global and local dynamics must occur on the time- and size-scale needed for function and growth of the organism. This protein folding/unfolding process occurs along an energy landscape, often referred to as a funnel, which contains any partially folded states accessible to the protein as well as the kinetic barriers separating them. This entire landscape is encoded in a protein's amino acid sequence. Despite significant advances in our understanding of how a protein's sequence encodes the native, or folded conformation (Jumper et al., 2021), we have a relatively poor understanding of how the lifetime or barriers between conformations are encoded. These barriers play numerous, important roles in biology – such as preventing inappropriate proteolytic degradation, as timers for complex systems, and in the pathogenesis of misfolding diseases (Sanchez-Ruiz, 2010). It is clear, however, that these barriers are not simply a result of the fold or topology of the protein, as proteins with the same overall native fold can have vastly different landscapes (Spudich, Miller, & Marqusee, 2004; Young, Skordalakes, & Marqusee, 2007).

While the native state of most proteins is thought to be at the free energy minimum on this landscape (Anfinsen's "thermodynamic hypothesis"), this thermodynamic control is not universal. There are now several published examples of proteins for whom the native state is kinetically trapped and the native conformation is maintained by kinetic control (an extremely high barrier to unfolding) (Baker & Agard, 1994; Sanchez-Ruiz, 2010). In many of these cases, a chaperone assists the protein in getting over this kinetic barrier – often a genetically encoded pro-region of the protein that is cleaved once the protein folds into its native state, leaving a kinetically-trapped native protein. Alpha-lytic protease (αLP), perhaps

the most well characterized example of such a protein, was shown to have no observable fluctuations from the native state (a high degree of rigidity) and significant resistance to proteolytic degradation (Jaswal, Sohl, Davis, & Agard, 2002). This lead to the widely accepted hypothesis that kinetically trapped, secreted proteases may have evolved such large barriers to improve survival in harsh, protease rich extracellular environments (P. N. Bryan, 2002; Jaswal et al., 2002; Sanchez-Ruiz, 2010). The mechanism by which αLP's large kinetic barrier arose is unknown. It is also unknown whether this is a general feature for all secreted proteases or why a protein would evolve such an unusually large kinetic barrier.

Kinetic barriers are not simply a feature of the topology or conformation as proteins with similar folds can have dramatically different lifetimes. In fact, proteins with the same topology can show both thermodynamic and kinetic control. A deeper understanding of how these barriers are encoded would expand our understanding of how the energy landscape is modulated and allow for the design of conformations with specified dynamics and lifetimes. Here, we investigate an extreme case where two protein sequences from the same organism with the same overall native conformation show dramatic differences in their unfolding barriers. We characterize and compare the energy landscape of both an intracellular and extracellular subtiliase from *Bacillus subtilis*.

Bacterial subtiliases fall into two subfamilies that have high structural and sequence conservation: intracellular subtilisin proteases (ISPs) and extracellular subtilisin proteases (ESPs). While these proteins have 30-50% sequence identity and similar folds, paralogs found in the same *Bacillus* species have been shown to have very different folding landscapes (Subbian, Yabuta, & Shinde, 2004; Vévodová et al., 2010). ISPs are thermodynamically stable proteins synthesized with a small (approximately 20 amino acids), N-terminal intramolecular inhibitor – the pro-domain, which is not required for folding (Gamble, Künze, Dodson, Wilson, & Jones, 2011; Strongin et al., 1978; Subbian et al., 2004). In contrast, ESPs, like other secreted proteases, are at best marginally stable thermodynamically and are thought to maintain their fold through a high kinetic barrier to unfolding (P. N. Bryan, 2002; Eder & Fersht, 1995; Sohl, Jaswal, &

Agard, 1998; James A Wells & Estell, 1988). ESPs require a 70 to 100 amino acid long intramolecular chaperone, the ESP pro-domain, to reach their native state. Once the protein is folded, the pro-domain is auto-digested, trapping the protein in the native state (P. N. Bryan, 2002; P. Bryan et al., 1995; Yabuta, Takagi, Inouye, & Shinde, 2001). While several ESPs have been characterized biophysically and engineered for industrial purposes, what aspects of their primary sequence convey their impressive kinetic barrier is unclear (Fu, Inouye, & Shinde, 2000; James A Wells & Estell, 1988; You & Arnold, 1996).

Despite the high resource cost of synthesizing a 100 amino-acid chaperone for a single protein-folding event, the ESP pro-domain remains a conserved feature within the subfamily. Sequence analysis shows evidence for positive selection on the pro-domain's size and charge during the evolution of ISPs, perhaps to facilitate protein degradation and turnover and reduce excessive proteolytic activity in the cell (Subbian et al., 2004). However, this analysis does not reveal what aspects of sequence or structure account for the differences in the energy landscape nor what pressures shaped the landscape. There are many potential environmental pressures that may select for increased unfolding barriers for secreted proteases, including a lack of proteostasis machinery outside the cell, the need to function in a harsh environment, and prevention of autolysis or degradation by competing proteases (Jaswal et al., 2002; Sanchez-Ruiz, 2010).

As a secreted protease, it is possible that ESPs' high kinetic barriers, like αLP's, have evolved to prevent degradation in the extracellular environment; however, this has not been established. Subtilisins are produced by *Bacillus* bacteria which are found largely in the soil but span a wide range of environments from temperate to extreme. In order to survive these inhospitable conditions, *Bacilli* form long-lasting, metabolically-inactive endospores (Turnbull, 1996). Germination of these spores can be triggered by L-amino acids (alanine, valine, and asparagine) – which begins the breakdown of the endospore cortex, allowing cell expansion and water uptake to occur and facilitate the return of metabolism (Peter, 2014). ESPs are believed to degrade other extracellular proteins as a source of nutrients for *Bacilli* (Graycar, Bott,

Power, & Estell, 2013), and may face a variety of selective pressures in the extracellular environment beyond avoiding proteolytic degradation by competing proteases.

These two subfamilies (ISP and ESP) present an ideal system for investigating the energy landscapes of two proteins with the same fold, but vastly different kinetic barriers. Here we compare the folding landscapes and dynamics of an ISP and ESP from *Bacillus subtilis*, Intracellular Subtilisin Protease 1 (ISP1) and Subtilisin E (SbtE) (Fig. 1). We use several biophysical approaches to monitor both the global stability and the kinetics of folding/unfolding as well as hydrogen-deuterium exchange monitored by mass spectrometry (HDX-MS) to monitor local stability and dynamics. We find that despite their high sequence similarity and structural homology, ISP1 and SbtE have very different conformational landscapes. In addition to differences in the native state energetics and dynamics, the effects of their pro-domains are also very different. While the SbtE pro-domain both speeds up folding to the native state and provides added stability to the folded state (Kobayashi & Inouye, 1992; Li, Hu, Jordan, & Inouye, 1995; Zhu, Ohta, Jordan, & Inouye, 1989), the ISP1 pro-domain has minimal effects on both folding and stability. Interestingly, the differences in their global energetics are reflected in local changes within the core – particularly the central alpha helix – pointing to a potential basis for these differences. Additionally, while the ESP has a very large barrier to unfolding, in the absence of the pro-domain it is not completely rigid, showing a greater degree of hydrogen exchange in the core than anticipated – contrary to findings for other kinetically-trapped, secreted proteases.

Results:

Intracellular Subtilisin Protease 1 (ISP1) Conformational Landscape: Global stability and re/unfolding kinetics

We first evaluated the global stability and folding/unfolding kinetics of the Intracellular Subtilisin Protease 1 (ISP1), both the mature and pro-protein. To avoid auto-proteolysis, we generated inactive variants by

changing the active-site serine (residue 250) to alanine in both variants (ISP1 S250A and Pro-ISP1 S250A). Global stability was determined by carrying out equilibrium chemically-induced denaturation monitored by following the circular dichroism (CD) signal at 222 nm as a function of guanidinium chloride (GdmCl) for both Pro-ISP1 S250A and ISP1 S250A.

The denaturation profile of ISP1 S250A showed two cooperative transitions, one with a midpoint near 0.93 ± 0.02 M GdmCl and the other with a midpoint near 1.92 ± 0.02 M GdmCl (Fig. 2A). These data were fit to a three-state model (U ⇋ I ⇋ N) with linear extrapolation (Barrick & Baldwin, 1993) resulting in a ΔG_{NI} = 5.88 ± 0.36 kcal•mol^{-1} and ΔG_{IU} of 1.86 ± 0.50 kcal•mol^{-1}, summing to a ΔG_{NU} of 7.74 ± 0.86 kcal•mol^{-1} (Table 1). The denaturation profile of Pro-ISP1 S250A also revealed two transitions, one with a midpoint near 1.1 ± 0.08 M GdmCl and the other with a midpoint near 1.7 ± 0.33 M GdmCl (Fig. 2A). The data were again fit to a three-state model (ΔG_{NI} = 7.34 ± 0.29 kcal•mol^{-1}, ΔG_{IU} of 1.23 ± 0.21 kcal•mol^{-1}, summing to a ΔG_{NU} of 8.57 ± 0.21 kcal•mol^{-1}) (Table 1). Thus, the presence of the pro-domain has minimal effect on the global energetics of ISP1.

To assess the kinetic barrier in the absence and presence of the ISP1 pro-domain we monitored the time course of refolding and unfolding for the same variants initiated by either diluting into (unfolding) or out of (refolding) high concentrations of guanidinium (see *Materials and Methods)*. For both ISP1 S250A and Pro-ISP1 S250A, the entire refolding process takes place within the deadtime of the instrument (5.6 milliseconds for stopped-flow fluorescence and ~15 seconds for manual mixing and CD). Thus, the rate of refolding must be greater than 180 s^{-1} ($\tau_{unfolded}$ of less than 5.6 ms). In contrast, unfolding of ISP1 S250A and Pro-ISP1 S250A shows observable kinetics when monitored by fluorescence that fit to a single exponential accounting for the entire signal change (Fig. S1). Linear extrapolation of lnk_u as a function of GdmCl concentration (Fig. 2B) results in an unfolding rate of 0.143 ± 0.0521 s^{-1} for ISP1 S250A and 0.0885 ± 0.00898 s^{-1} for Pro-ISP1 S250A (Table 1). Thus, the presence of the pro-domain does not appear to impact the unfolding rate for ISP1.

Subtilisin E (SbtE) Conformational Landscape

Next, we evaluated both the global stability and folding/unfolding kinetics of the extracellular Subtilisin E (SbtE). As expected, the refolding and unfolding of SbtE is too slow to allow determination of global stability using standard equilibrium experiments. Instead, we turned to kinetic measurements to monitor the unfolding and folding rates and used these to obtain an estimate for the global stability. To generate mature SbtE, we used wild-type (active) SbtE, which when expressed with the pro-domain, auto processes to yield mature SbtE (see *Materials and Methods*).

Unfolding of mature SbtE was monitored by CD, with addition of an inhibitor to prevent autodigestion during the unfolding process (Keller, Seufer-Wasserthal, & Jones, 1991; Lindquist & Terry, 1974), at pH 5 or 8 (See *Materials and Methods*). The unfolding rate was measured at pH 5 for comparison with refolding experiments described below and at pH 8 for comparison with ISP1 S250A and Pro-ISP1 S250A unfolding experiments. SbtE unfolding traces were fit to a single exponential (Fig. S2 and S3) and linear extrapolation of $\ln k_u$ as a function of guanidinium concentration (Fig. 2B) results in a $\ln k_u$ in the absence of denaturant of $1.92\times10^{-7} \pm 1.27\times10^{-7}$ s^{-1} at pH 8 (Table 1). Linear extrapolation of $\ln k_u$ as a function of guanidium concentration at pH 5 results in a $\ln k$u of $2.8\times10^{-5} \pm 1.9\times10^{-5}$ s^{-1} (Table 1). Thus, unlike ISP1, the presence of the pro-domain increases the unfolding rate of SbtE by over 6 orders of magnitude compared to the unfolding of the mature protein (Subbian et al., 2004).

The refolding of mature SbtE is too slow to be monitored by traditional spectroscopy. However, since activity is a much more sensitive assay for folded, native protein, the concentration of active protease can be used as a proxy for the fraction folded protease, as seen for αLP (Baker & Agard, 1994; Baker, Sohl, & Agard, 1992; Sohl et al., 1998; Takagi, Maeda, Ohtsu, Tsai, & Nakamori, 1996). Therefore, the unfolded protein was diluted into folding conditions and assayed for active protease as a function of incubation time (see *Materials and Methods*). Again, to prevent autodigestion during refolding, a KOAc buffer at pH5

with phenylboronic acid was used. In the absence of the pro-domain, no activity was detected even after six days. Using a standard curve with the active protease, we can estimate our assay has a lower detection limit of 2×10^{-10} M folded protein. Since even a small amount of folded SbtE can rapidly digest the remaining unfolded SbtE, a second sample was taken at each time point, supplemented with pro-domain *in trans*, and assayed for activity to determine the total intact protein remaining at each time point (Fig. 3, see *Materials and Methods*). Based on the lower detection limit of the activity assay, the maximum amount of folded protein was normalized by the amount of intact protein at each time point (to calculate a maximum fraction folded). These data were fit to the model from Sohl et al. 1998 to generate a maximum value for k_f of 4.92×10^{8} s^{-1} at 4°C and 5.07×10^{8} s^{-1} at 25°C (Table 1). Comparing the rates measured here to previously published rates for Pro-SbtE, we find that the pro-domain increases the refolding rate by at least 9 orders of magnitude (Subbian et al., 2004).

Local Stability by HDX-MS:

To investigate the effects on the energy landscape at the amide level, we turned to hydrogen-deuterium exchange monitored by mass spectrometry (HDX-MS), first on the mature proteins (ISP1 S250A and SbtE) and then on the pro-proteins (Pro-ISP1 S250A and Pro-SbtE S221A). We monitored the exchange process in a continuous exchange experiment (25°C, pH 8, see *Materials and Methods*), following the uptake of deuterium as a function of time. HDX was monitored using mass spectrometry at the level of peptides from time points as short as 15 s to as long as 1 year (see *Materials and Methods*). Both proteins digest well with >95% coverage of the protein sequence and > 500 peptides from ISP1 S250A and > 150 peptides from SbtE (Fig. S4).

The data were analyzed using the standard Linderstrøm-Lang model (Hvidt & Nielsen, 1966), where hydrogen exchange occurs under one of two kinetic regimes, so-called EX1 and EX2, that can be easily differentiated by mass spectrometry. Each report on a different aspect of the fluctuations that allow

exchange. In EX1, the observed rate of exchange reports on the kinetics of this fluctuation or opening event and in EX2 the observed rate of exchange reports on the fraction of the population in the exchange-competent or open state (thermodynamics). When monitored by mass spectrometry, EX2 exchange results in a single peak that shifts from a lighter to heavier average m/z over time, while EX1 results in two peaks (bimodal behavior) where the intensity of the lighter peak decreases over time as the intensity of heavier peak increases.

Despite similar overall folds, the two mature proteins show very different hydrogen exchange behavior (Fig 4A). The HDX behavior for ISP1 S250A is as expected for a thermodynamically stable protein. The majority of the peptides show EX2 exchange, and the exchange process is complete within two weeks, consistent with the expected global stability of 7 kcal/mol (where exchange occurs from the rare fraction of unfolded proteins). SbtE, however, showed much more complicated HDX behavior. While the majority of peptides examined did show EX2 behavior – we observed only minimal exchange over the time course of the experiment. Thus, the observed rate of exchange was very slow (not reaching full exchange even after one year), too slow to arise from the unfolded state based on the stability estimates in Table 1 and suggesting that these exchange events arise from very high energy (rare) fluctuations on the native side of the barrier.

Analysis of the individual peptides allowed us to look at these fluctuations at a more local level, where there are interesting differences in behavior, most notably in the core alpha helix of the proteins (Fig. 1A). For ISP1 S250A, peptides in this helix exchange rapidly (in less than 15 s), despite arising from the central core of this thermodynamically stable protein (Fig. 4B). Thus, this core region is surprisingly flexible, and exchange is not limited by global unfolding of the protein. In contrast, in SbtE, peptides from this region show no observable exchange, even out to a year (Fig. 4B, Fig. S6), consistent with a model where exchange from this region is likely in EX1 and requires crossing the high-energy kinetic barrier to global unfolding. The core beta sheets and calcium-binding loops of SbtE, however, show very slow EX2 exchange

(Fig. 4B), consistent with a rigid core of the protein that still undergoes some fluctuations to exchangeable conformations. In sum, while ISP1 S250A's exchange pattern suggests a high degree of flexibility in the core-region of the protein, SbtE's core exchange pattern suggests a high degree of rigidity.

Next, we carried out HDX on the proteins containing the pro-domain using the inactive variants Pro-ISP1 S250A and Pro-SbtE S221A. We monitored continuous HDX at 25°C, pH 8 from time points as short as 15s to as long as 2 weeks. Again, both proteins are amenable to HDX-MS with >95% coverage of the protein sequence and >700 peptides from Pro-ISP1 S250A and >400 peptides from Pro-SbtE S221A (Fig. S4). Both pro-proteins behaved like typical thermodynamically stable proteins, exhibiting EX2 exchange throughout the entire structure with exchange rates consistent with thermodynamically stable proteins. For ISP1, compared to the mature protein, the addition of the pro-domain slows hydrogen exchange, consistent with the overall increase in global stability (Table 1). There are also notable differences at the local level. In particular, Pro-ISP1 S250A's core alpha helix (Fig. 1A) shows slowed exchange consistent with exchange from the globally unfolded state. In sum, these data suggest that compared to the mature ISP1 S250A, the pro-protein behaves more like a typical folded protein with a rigid core.

In contrast, for Pro-SbtE S221A the presence of the pro-domain increases the observed hydrogen exchange compared to SbtE. The one exception to this observation is at the active-site loops that contact the pro-domain (Fig. 4B, Fig. S5), as would be expected from the increased burial of these residues. The core alpha helix, which in the absence of the pro-domain appears to exchange from the globally unfolded state, shows increased exchange and therefore increased flexibility when the pro-domain is present (Fig. 4B). Thus, unlike the intracellular protein, the presence of the pro-domain increases the fluctuations from this core alpha helix . In sum, removal of the pro-domain results in a decrease in the overall flexibility of SbtE, which is consistent with conversion of a thermodynamically stable protein to a kinetically trapped protein.

Increased local stability and proteolytic resistance:

As mentioned above, core beta sheets in SbtE show slow exchange – this exchange is in the EX2 kinetic regime as shown by a shifting centroid (Fig. S7) and under thermodynamic control. On the other hand, the core alpha helix may be in the EX1 kinetic regime and under kinetic control. The presence of equilibrium fluctuations in SbtE differs from αLP, which was shown to only be under kinetic control (Jaswal et al., 2002).

To determine if the decreased fluctuations seen in the core of the mature protein, SbtE, correlates with increased proteolytic resistance, as with αLP (Jaswal et al., 2002), we carried out proteolytic competition assays using thermolysin. Proteolysis was monitored by SDS-PAGE gel analysis (see *Materials and Methods*). Thermolysin and subtilisins have similar substrate preferences, both showing a preference for bulky, hydrophobic amino acids (Heinrikson, 1977; Keil, 1992; Takagi et al., 1996; J A Wells, Cunningham, Graycar, & Estell, 1987). Figure 5 shows that digestion of thermolysin is nearly completely within 15 s while SbtE remains undigested out to 24 hrs, >10,000 times longer. Thus, in this assay, the observed decreased local stability does appear to correlate with increased proteolytic resistance.

Discussion:

Ever since the classic work of Anfinsen and Levinthal (Anfinsen, 1973; Levinthal, 1968), how proteins manage to navigate their vast conformational landscape to fold in a biologically relevant timescale has been an important and heavily investigated area. It is clear, however, that once folded, the lifetime of the native conformation (both its fluctuations and global unfolding) is also under selective pressure. The mechanisms that control these partial and global unfolding events and the lifetime of the folded state are not well understood. Here, we have examined two subtiliases from the same organism, with the same basic fold, that function in very different environments (ISP1 and SbtE from *Bacillus subtilis*). From a comparison of their global energetics and dynamics as well as their local dynamics, we find that, despite

sharing the same fold and a high level of sequence identity, ISP1 and SbtE have very different energy landscapes at all levels: their global stability (thermodynamics), unfolding/refolding kinetics, and fluctuations at the subglobal level are all different. While each has a pro-domain, the small pro-domain of the intracellular protein (ISP1) has very little effect on the energetics and folding of the protein. For the extracellular protein, the pro-domain serves as an apparent intramolecular chaperone, speeding up folding, and generating a thermodynamically unstable, kinetically trapped mature protein with a high barrier for unfolding. The mature SbtE shows rare, but measurable, fluctuations from the native state, unlike the well-characterized kinetically trapped protein Alpha-lytic protease (αLP) (Jaswal et al., 2002).

As mentioned before, other secreted proteases, like αLP, are known to be thermodynamically unstable and kinetically trapped (so called 'kinetic stability') and dependent on their pro-domain for folding. The biophysical studies of αLP laid the foundation for our understanding of what the energy landscape of a kinetically trapped protein may look like and why they may have evolved (Baker & Agard, 1994; Baker et al., 1992; Jaswal et al., 2002; Sohl et al., 1998). In these studies, the sequence αLP was compared with other chymotrypsin-like proteases, but as αLP has no known direct homologs, it was not subjected to the comparative analysis carried out here. Subsequent studies examining the relationship between pro-domain size and kinetic stability for extant, but inactive, αLP relatives and their ancestral states found that pro-domain size correlated with the height of the kinetic unfolding barrier and that the proteins cluster phylogenetically based on pro-domain size, providing additional insight into structural features that impact the energetics of different αLP family members (Nixon et al., 2021). The work presented here on two subtiliases is , to our knowledge, the first complete and direct comparison of two homologs, from the same species, and the difference in their global and local energetics.

Differences in global energetics and kinetics

The intracellular protease ISP1 behaves, as expected, as a typical thermodynamically stable, three-state folder, both in the presence and absence of its pro-domain. Like most intracellular subtiliases, ISP1 has a small (17 aa) pro-domain. Despite the evolutionary conservation of the pro-domain, we find that the pro domain has little impact on the energy landscape of ISP1. The presence of the pro domain increases the global stability of the protein by less than a kcal/mol (7.74 ± 0.86 to 8.56 ± 0.34). Similarly, the kinetic barrier for unfolding is quite low, both in the presence and absence of the pro domain (k_u = 1.43×10^{-1} ± 5.21×10^{-2} s^{-1} and 8.85×10^{-2} ± 7.33×10^{-3} s^{-1}, respectively). Thus, for the intracellular protease, the pro-domain (only 17 amino acids) does not appear to function as an intramolecular chaperone, but more likely as a regulator for the proteolytic activity of the enzyme (zymogen).

The pro-domain plays a very different role in the extracellular protease. For SbtE, the pro-domain has a large impact on the energy landscape of the protease. The presence of the 78 amino acid pro-domain increases the stability of the protein. Although we cannot precisely determine the global stability of mature SbtE (because we are unable to detect any folded protein (by activity) within our experimental limits), we can say that the protein is not thermodynamically stable because the unfolding rate is faster than the folding rate. Thus, like αLP, SbtE is kinetically trapped. We estimate the pro-domain increases the stability of the protein by at least 13 kcal/mol (from an estimated maximum of 3.73 kcal/mol to almost −10 kcal/mol).

Although kinetically trapped, SbtE's unfolding barrier is not as dramatically high as seen for αLP – with an unfolding rate not that different from many thermodynamically stable proteins (a folded state lifetime of 100 days vs 1.2 years for αLP). However, the lifetimes of both proteins are likely well beyond the biologically relevant timescale – questioning the evolutionary pressures on such unusually high barriers.

Differences in local fluctuations:

In addition to differences in their global energetics, the mature SbtE and ISP1 also show very different fluctuations at the local level. We found that ISP1 behaves like a typical thermodynamically stable protein with fluctuations from the native state observable by hydrogen exchange. These fluctuations are all in the EX2 regime – showing a protein with fluctuations that span in energy from near the native state to the globally unfolded protein. The pro-domain was found to mildly increase the stability of the protein, and to rigidify the protein, as monitored by HDX, as might be expected by the binding of a small domain or region. Since the pro-domain is widely believed to act solely as an inhibitor to ISP1's proteolytic activity, a reduction in fluctuations (which may facilitate activity) due to the pro-domain is reasonable.

The pro-domain has very different effects on the local dynamics of the extracellular protein, SbtE. When monitored by HDX, mature SbtE shows minimal fluctuations, whereas the Pro-SbtE behaves like a typical folded protein, with fluctuations that span in energy from near the native state to the globally unfolded protein. While the increased fluctuations in the presence of SbtE's pro-domain was expected (the pro-domain lowers the barrier between states making excursions from the native state more accessible), we were surprised by the extremely slow hydrogen exchange we do observe in SbtE. Based on the k_{int} for the peptides in this region the ΔG of HDX for the amides in these peptides would be approximately 6.5 kcal/mol, which is greater than the global stability of the protein (ΔG_{NU}). Given the high kinetic barrier to unfolding, these equilibrium fluctuations must arise from the native side of the unfolding barrier.

The pro-domain affects the flexibility in the core alpha helix of each protein in opposite ways. As mentioned above, in ISP1, the presence of the pro-domain rigidifies this core helix as seen by the increase in fluctuations measured by hydrogen exchange. Thus, the increase in stability provided by the pro-region is transmitted to the core alpha helix and appears to rigidify this region. We see the opposite trend in the extracellular protein SbtE, where the presence of the pro-domain increases the flexibility of the core helix. In fact, we see no measurable hydrogen exchange for this helix in the mature protein. Interestingly, while

the relatively large pro-domain dramatically increases the global stability of the protein, it generates increased flexibility within the core. How these energetic effects are uncoupled is unclear.

Sequence basis for the observed differences in the energy landscape

While ISP1 and SbtE have similar folds, they have only 45% sequence identity – making it difficult to identify the sequence basis for the differences in their energy landscape. The observed differences in the flexibility of the core alpha-helix suggest a potential region of the protein to focus on. In fact, ISP1 has an insertion at the C-terminus of the alpha helix that is conserved amongst many ISPs (Fig. 6). This insertion may be one source of the difference in folding/unfolding mechanisms between SbtE and ISP1. Furthermore, the active site serine resides at the N-terminus of this helix and in the case of ISP1, this whole region becomes more flexible in the absence of the pro-domain. Increased dynamics may benefit the enzymatic ability of ISP1. Since ISP1 does not reside in harsh extracellular conditions there is no reason to maintain a rigid core, and therefore native state. How SbtE maintains activity with such a rigid core is unclear.

Comparison between SbtE and alpha-lytic protease

A widely accepted hypothesis for the evolution of kinetically stable secreted proteases is that, once secreted into the extracellular space, they need to avoid degradation by the many other proteases present in the environment. For most proteins, proteolysis requires fluctuation from the native conformation (Park & Marqusee, 2004). Indeed, αLP was shown to have no observable fluctuations from the native state over the course of six months and outlast trypsin and chymotrypsin in an *in vitro* competition (Jaswal et al., 2002). Surprisingly, our results suggest that such complete suppression of fluctuations is not needed for increased proteolytic resistance, since SbtE does show equilibrium fluctuations, albeit rare, and nevertheless seems to outcompete thermolysin in a competition assay. Thus, while SbtE may have a greater degree of flexibility than αLP, it is still quite resistant to proteolysis. Perhaps there exists different

selective pressures in different environments requiring different levels of rigidity – or the increased resistance to proteolysis of αLP is not the only evolutionary pressure on the rigidity of this protein.

In sum, we have examined two homologous proteins from the subtiliase family with similar folds and sequences. We have found that despite these similarities the two proteases have very different energetics at the global and local levels. Differences in their local energetics within the core alpha helix have revealed that an insertion into this core helix may be an evolutionarily important feature of their landscape. How these features and kinetic barriers are encoded in protein sequence remains an outstanding question in the field. Our study suggests the need for high throughput experiments to explore the sequence dependence of a protein's kinetic barriers for a specific fold.

Materials and Methods:

Plasmid Cloning:

The DNA encoding intracellular and extracellular subtilisin were purchased as gBlock gene fragments from IDT. Enzyme digestion and site-directed mutagenesis were used to clone: (1) Pro-Subtilisin E S221A and Pro-Subtilisin E WT into a pSV272 expression vector without the secretion signal and an N-terminal His6-MBP-TEV-(GS)25-PreScission tag and (2) Pro-ISP1 S250A and ISP1 S250A into pET-27b(+) (Novagen) expression vector. Sequences were confirmed with Sanger sequencing.

Extracellular Subtilisin Protein Expression:

Bl21 (DE3)pLysS competent cells were transformed with pSV272 vectors containing either Pro-SbtE S221A or Pro-SbtE WT DNA and plated on LB agar plates with 50 µg/mL kanamycin. Starter cultures were inoculated with single colonies for overnight growth to saturation. For Pro-SbtE S221A, 15 mL of LB media overnight culture was used to inoculate 1 L of LB media with 50 µg/mL kanamycin. Cells were grown at 37°C and 250 rpm to an optical density at 600 nm of 0.6-0.8 and induced with 1 mM IPTG. Cells were

harvested after 3 hrs by centrifugation at 4000 g for 30 min at 4°C. For Pro-SbtE WT, 100 µL of 2xYT media overnight culture was used to inoculate 5 mL of 2xYT media with 50 µg/mL kanamycin and 1% glucose and grown at 37°C and 250 rpm to an optical density at 600 nm of 0.6-0.8. Cells were cooled to 20°C and induced with 1 mM IPTG. Cells were harvested after 18 hrs by centrifugation at 4000 g for 10 min at 4°C. Cell pellets were stored at −80°C until purification.

Intracellular Subtilisin Protein Expression:

Bl21 (DE3)pLysS competent cells were transformed with pET-27b(+) vectors containing either Pro-ISP1 S250A or ISP1 S250A DNA and plated on LB agar plates with 50 µg/mL kanamycin. Starter cultures were inoculated with single colonies for overnight growth to saturation. For Pro-ISP1 S250A and ISP1 S250A, 15 mL of LB media overnight culture was used to inoculate 1 L of LB media with 50 µg/mL kanamycin. Cells were grown at 37°C and 250 rpm to an optical density at 600 nm of 0.6-0.8 and induced with 1 mM IPTG. Cells were harvested after 3 hrs by centrifugation at 4000 g for 30 min at 4°C.

Pro-Subtilisin E S221A Protein Purification:

Pellets were thawed on ice, resuspended in 30 mL buffer A1 (50 mM Tris pH 8, 150 mM NaCl, 1 mM $CaCl_2$, 0.5 mM TCEP) with 1 µL benzonase per 1 L pellet, and lysed by sonication at 30% for 15 minutes on ice. The lysate was centrifuged at 13000 rpm for 30 min at 4°C. The resulting pellet was washed once with 30 mL of 1% Triton in buffer A1 and once with 30 mL of buffer A1 alone. The washed inclusion bodies were resuspended in 10 mL of buffer A with 8 M urea, stirred at RT for 15 minutes, and sonicated for 10 seconds 5x. The denatured inclusion bodies were centrifuged at 14000g for 15 minutes. Protein was refolded by dropwise 10-fold dilution into buffer A1, stirred at RT for 15 minutes, and centrifuged at 13000 rpm for 30 minutes.

The supernatant was incubated with amylose beads for 2 hrs rotating at 4°C. The amylose beads were loaded into a gravity column. Flow through was collected and beads were washed with 50 mL of buffer A.

His6-MBP-TEV-(GS)25-PreScission-Pro-SbtE S221A was eluted from the amylose beads in three 10 mL fractions with 10 mM maltose in buffer A1. The protein concentration of each fraction was measured by ultraviolet-visible (UV-vis) spectroscopy, fractions with protein were pooled, and PreScission protease was added, and the cleavage reaction proceeded overnight at 4°C. The N-terminal His6-MBP-TEV-(GS)25-PreScission tag was removed with subtractive a NiNTA step. The cleavage reaction was diluted to 50 mL in buffer A1 to 2.5 mM imidazole and incubated with NiNTA beads for 2 hrs rotating at 4°C. The NiNTA beads were loaded into a gravity column. Flow through was collected and beads were washed with 50 mL of buffer A1 with 10 mM imidazole. Flow through and wash fractions were both concentrated to <1 mL using Amicon Ultra-15 Centrifugal Filter with a 10 kDa MWCO. The protein concentration was measured by UV-vis spectroscopy and fractions with protein were pooled. The pooled fractions were 0.22 µm filtered and further purified by size-exclusion chromatography on a HiLoad S75 16/600 or HiLoad S200 16/60 (GE) in buffer B (50 mM Tris pH 8, 500 mM $(NH_4)_2SO_4$, 1 mM $CaCl_2$, 0.5 mM TCEP). The peak corresponding to Pro-SbtE S221A was collected, concentration was measured by UV-vis spectroscopy and stored overnight at 4°C for immediate use.

Wildtype Subtilisin E Protein Purification:

Pellets were thawed on ice, resuspended in 0.5 mL BugBuster (Millipore) with 0.5 µL benzonase per 5 mL pellet, and rotated at RT for 2 hrs Lysis released soluble, active SbtE, which cleared the majority of the lysate. Lysate was centrifuged at 14000g for 10 minutes, concentrated to <2 mL, and 0.22 µm filtered. Filtered supernatant was run on a HiLoad S75 16/600 or HiLoad S200 16/60 (GE) in buffer B. The peak corresponding to SbtE was collected and concentrated to <2 mL. SbtE was diluted 10-fold in buffer A2 (50 mM Tris pH 8, 25 mM NaCl, 1 mM $CaCl_2$, 0.5 mM TCEP) and further purified on a HiTrap Q HP (GE) and SbtE was collected from the flow-through. The peak corresponding to Pro-SbtE was collected, concentration was measured by UV-vis spectroscopy, and stored at 4°C.

Pro-ISP1 S250A Protein Purification:

Pellets were thawed on ice, resuspended in buffer A2 with 1 µL benzonase per 1 L pellet, and lysed by sonication at 30% for 15 minutes on ice. The lysate was centrifuged at 13000 rpm for 30 min at 4°C. Supernatant was 0.22 µm filtered. Pro-ISP1 S250A was isolated using anion-exchange chromatography on a HiPrep 16/10 Q XL (GE) and eluted with a gradient from 25 mM to 1 M NaCl. Pro-ISP1 S250A typically elutes around 300 mM NaCl. The peak corresponding to Pro-ISP1 S250A was collected, concentrated to <2 mL, and 0.22 µm filtered. Pro-ISP1 was further purified by size-exclusion chromatography on a HiLoad S75 16/600 or HiLoad S200 16/60 (GE) in buffer B. The peak corresponding to Pro-ISP1 S250A was collected, concentration was measured by UV-vis spectroscopy and stored overnight at −80°C for later use.

ISP1 S250A Protein Purification:

Pellets were thawed on ice, resuspended in buffer A2 with 1 µL benzonase per 1 L pellet, and lysed by sonication at 30% for 15 minutes on ice. The lysate was centrifuged at 13000 rpm for 30 min at 4°C. The resulting pellet was washed once with 30 mL of 1% Triton in buffer A2 and once with 30 mL of buffer A2 alone. The washed inclusion bodies were resuspended in 10 mL of buffer A2 with 8 M urea, stirred at RT for 15 minutes, and sonicated for 10 seconds 5x. The denatured inclusion bodies were centrifuged at 14000g for 15 minutes. Protein was refolded by dropwise 10-fold dilution into buffer A2, stirred at RT for 15 minutes, and centrifuged at 13000 rpm for 30 minutes. Supernatant was 0.22 µm filtered. ISP1 S250A was isolated using anion-exchange chromatography on a HiPrep 16/10 Q XL (GE) and eluted with a gradient from 25 mM to 1 M NaCl. ISP1 S250A typically elutes around 250 mM NaCl. The peak corresponding to ISP1 S250A was collected, concentrated to <2 mL, and 0.22 µm filtered. Pro-ISP1 was further purified by size-exclusion chromatography on a HiLoad S75 16/600 or HiLoad S200 16/60 (GE) in

buffer B. The peak corresponding to Pro-ISP1 S250A was collected, concentration was measured by UV-vis spectroscopy and stored overnight at 4°C for immediate use.

Equilibrium Denaturation by Circular Dichroism Spectroscopy:

Protein samples were diluted to a concentration of 0.04 mg/mL in 0 M or 7 M GdmCl in buffer B. High and low denaturant protein samples were either mixed manually to create a range of GdmCl concentrations or high denaturant protein was titrated into low denaturant protein with the ATS-350 titrator (Jasco). Protein was allowed to equilibrate at each GdmCl concentration for at least 5 minutes at 25°C before measurement. CD signal was measured at 222 nm for 30 seconds in a 1 cm quartz cuvette (Starna) with a stirring magnetic stir bar. For manually mixed samples, each sample was recovered and the GdmCl concentration was measured with a refractometer. Denaturation curves were plotted and fit to either a standard six parameter two state fit (Eq. 2) or a nine parameter three state fit (Eq. 3) (Barrick & Baldwin, 1993; Street, Courtemanche, & Barrick, 2008).

$$\theta = \frac{y_N + y_U \cdot e^{-(m\cdot([D]-C_m)/RT}}{1 + e^{-(m\cdot([D]-C_m)/RT}} \ (Equation\ 2)$$

$$\theta = \frac{y_N + y_I \cdot e^{-(m_{NI}\cdot([D]-C_{m,NI})/RT} + y_U \cdot e^{-(m_{NI}\cdot([D]-C_{m,NI})/RT} \cdot e^{-(m_{IU}\cdot([D]-C_{m,IU})/RT}}{1 + e^{-(m_{NI}\cdot([D]-C_{m,NI})/RT} + e^{-(m_{NI}\cdot([D]-C_{m,NI})/RT} \cdot e^{-(m_{IU}\cdot([D]-C_{m,IU})/RT}} \ (Equation\ 3)$$

Where Θ is the CD signal at 222 nm, m is the m-value for each transition, and:

$$y_i = m \cdot x + b \ (Equation\ 4)$$

Unfolding Kinetics by Circular Dichroism Spectroscopy:

Protein samples were diluted to a concentration of 0.4 mg/mL in 0 M or 7 M GdmCl in buffer B. All samples and buffers were allowed to equilibrate overnight at 25°C. For each experiment 1.95 mL of unfolding buffer was placed into a 1 cm quartz cuvette. To initiate folding, 0.05 mL of protein was added to the same

cuvette, rapidly mixed with p1000 micropipetter, capped, and placed in the CD with a stirring magnetic stir bar. An estimate of the deadtime for mixing and starting the measurement was also recorded. CD signal was measured at 222 nm, integrated over 8 seconds, every 10 seconds for 3 to 8 hours. Each sample was recovered and the final GdmCl concentration was determined with a refractometer. Individual unfolding traces were plotted and fit to a single exponential (Eq. 5). Calculated unfolding rates were plotted against GdmCl concentration and fit to a line (Eq. 4) in Python.

$$\theta = \theta^{\circ} + A \cdot e^{-(x-x_0)/\tau} \ (Equation\ 5)$$

Where Θ is the CD signal and τ is the lifetime of the starting state.

Unfolding and Refolding Kinetics by Stopped-Flow Fluorescence:

Protein sample was diluted to a concentration of 150 µM in 0 M or 7 M GdmCl in buffer B. All samples and buffers were allowed to equilibrate for at least 5 minutes at 25°C. Protein sample was loaded into 2 mL slip tip syringe and attached to position 4 on the SFM 400 (BioLogic) and buffer was loaded into a 10 mL slip tip syringe and attached to position 3. Positions 1 and 2 were loaded with MilliQ water and not used during the experiment. For each shot of a refolding experiment 225 µL of buffer and 25 µL of protein were pushed into the FC-15 cuvette (BioLogic) at a flow of 6.5 mL/s. Five shots were used to clear the cuvette and the following five shots were recorded. Sample was excited at 280 nm and emission at 320 nm was recorded for 50-500 ms sampling every 0.01-1 ms. The dead time of the stopped-flow was 5.6 ms. Individual unfolding traces were plotted and fit with Python scripts with a single exponential equation (Eq. 5). Calculated unfolding rates were plotted against GdmCl concentration and fit to a line (Eq. 4) in Python.

Unfolding and Refolding Kinetics by Stopped-Flow Circular Dichroism Spectroscopy:

Protein sample was diluted to a concentration of 7 mg/mL in 0 M or 7 M GdmCl in buffer B. All samples and buffers were allowed to equilibrate for at least 5 minutes at 25°C. Protein sample was loaded into

reservoir 1 of the Aviv 430 stopped flow and buffer was loaded into reservoir 3. For each shot of a refolding experiment 80 µL of buffer and 8 µL of protein were pushed into 0.1-cm path length cuvette using a delta mixer (Aviv). Five shots were used to clear the cuvette, and the following five shots were recorded. Signal at 222 nm was recorded for 50-500 ms sampling every 0.01-1 ms. The stopped-flow dead time was 70 ms. Individual unfolding traces were plotted and fit with Python scripts with a single exponential equation (Eq. 5). Calculated unfolding rates were plotted against GdmCl concentration and fit to a line (Eq. 4) in Python.

Refolding Kinetics Measured by Subtilisin E Activity:

This protocol was adapted from (Sohl et al., 1998). Protein was dialyzed into 7 M GdmCl in buffer C (50 mM Potassium Acetate pH 5, 500 mM $(NH_4)_2SO_4$, 1 mM $CaCl_2$, 0.5 mM TCEP) with 10 mM phenylboronic acid, a low-affinity (K_i = 8 mM or 67 µM – see refs) subtilisin inhibitor, overnight at 4°C. This procedure limits autodigestion and removes any protein fragments from residual autodigestion during unfolding. Refolding was initiated by rapidly diluting 5 µL of 400 µM protein with 995 µL of buffer C. Refolding was allowed to proceed at 4°C or 25°C. N,N-dimethylformamide (DMF) stock of Subtilisin E substrate, N-succinyl-Ala-Ala-Pro-Phe-P-nitroanilide (s-AAPF-pNA), was diluted in buffer B and concentration was measured absorption at 315 nm. Time points were taken at 24, 48, 72, and 168 hours. At each time point protease activity was measured. The enzymatic reaction was initiated by adding 6 µL of refolding reaction to 54 µL s-AAPF-pNA in buffer B for a final concentration of 0.2 µM protein and 1 µM substrate in a 1-cm path length masked quartz cuvette (Starna). Intact folding intermediate (I) decrease as a function of time due to proteolysis by refolded N. Intact I was measured at each time point, 2 molar excess of Pro-domain was added to an additional 6 µL of the refolding reaction and allowed to equilibrate at 25°C for 2 hours. The sample was then diluted 5-fold with buffer B and treated with 0.004 ug of trypsin and incubated at 25°C for 30 min. The enzymatic reaction was initiated with 1 µM s-AAPF-pNA in 60 µL of the treated sample in a 1-cm path length masked quartz cuvette. Absorption at 410 nm was recorded every 100 ms for 10 minutes. The detection limit is approximately 2 pM in the 60 µL reaction.

Individual activity traces were plotted and fit a line (Eq. 4). A standard curve was used to determine the concentration of active protein. Calculated fraction folded ($[N]/[I]_t$) was plotted against time and fit with $k_{f=}$ $([N]/[I]_0)/t$ (Sohl et al., 1998) to find refolding rates (Sohl et al., 1998).

Subtilisin E Activity Standard Curve:

Protein was serially diluted to a concentration of 2E1 to 2E−5 µM in 80% buffer B, 20% buffer C. DMF stock of s-AAPF-pNA was diluted in buffer B and concentration was measured by absorption at 315 nm. Phenylboronic acid was dissolved in buffer B. The protease reaction was initiated by adding 6 µL of Subtilisin E standard to 54 µL of in buffer B for a final concentration of 1 µM substrate, 7.5 µM phenylboronic acid in a 1 cm masked quartz cuvette (Starna). This was done for all Subtilisin E standards, measuring activity of 2E0 to 2E−6 µM protein at 25°C. Absorption at 410 nm was recorded every 100 ms for 10 minutes. Individual activity traces were plotted and fit with Python scripts to a line). Calculated initial velocities were plotted against protein concentrations and fit with Python scripts to a line.

Protease Survival Assay:

Stock thermolysin was prepared by dissolving ~0.5 mgs of thermolysin (Sigma-Aldrich, P1512) into buffer B and concentration was determined by measuring absorption at 280 nm. Proteolysis was initiated by adding the stock thermolysin to Subtilisin E in buffer B for a final concentration of 10 µM and 10 µM respectively and incubated at 25°C. Samples were removed at at 0.5, 1, 2, 4, 8, 24, 48, 72, 96, 168, and 336 hours and quenched with 100 mM EDTA or 100 mM EDTA and 250 µM PMSF. Each time point was run on a NuPAGE Bis-Tris gel (Invitrogen), stained with Sypro Tangerine, and imaged on a Typhoon imager. Band intensities were quantified in ImageJ. Data was plotted and fit to a single exponential equation.

Continuous Hydrogen-Deuterium Exchange:

For all hydrogen-deuterium exchange (HDX) experiments, buffer B was lyophilized and resuspended in D_2O. Deuterated buffer B and protein samples were pre-equilibrated to 25°C. HDX was initiated by diluting protein samples 10-fold into deuterated buffer B for a final protein concentration of 5 µM. At each time point, samples were quenched by mixing 50 µL of HDX reaction with 50 µL of ice-cold 2x quench buffer (3.5 M GdmCl, 1.5 M Glycine pH 2.4, 500 mM TCEP) and flash freezing in liquid nitrogen. Samples were stored at −80°C. For Pro-SbtE S221A, Pro-ISP1 S250A, and ISP1 S250A, 0.5 µg/mL angiotensin peptide was included as a fiduciary as a back-exchange control, although the data were not back-exchange corrected.

Protease Digestion and LC-MS:

Protein samples were injected into the LC (Thermo Ultimate 3000) and run on the Q Exactive Orbitrap Mass Spectrometer (ThermoFisher) as described previously (Costello et al., 2022). The quenched samples were inline digested with columns made by covalently attaching aspergillopepsin (Sigma-Aldrich, P2143) and the porcine pepsin (Sigma-Aldrich, P6887) to resin using standard aldehyde chemistry. Digestion occurred at a flow rate of 200 µL/min of aqueous phase buffer (0.1% formic acid) at 10C. Peptides were then desalted for 4 minutes on a hand-packed trap column (Thermo Scientific POROS R2 reversed-phase resin 1112906, IDEXC-130B). Peptides were separated on a C18 analytical column (Waters Acquity UPLC BEH C18 Column, 1.7 µm particle size, 1 mm (inner diameter) × 50 mm, 186002344) with a gradient 5-40% organic phase buffer (100% acetonitrile, 0.1% formic acid) at 40 µL/min over 14 minutes. The analytical and trap columns were then washed with 40-90% organic phase buffer sawtooths every 30 seconds. The analytical and trap columns were then equilibrated at 5% organic phase buffer before the next injection. After digestion and desalting, the protease columns were washed with three injections of 100 µL of 1.6 M GdmCl, 0.1% formic acid before the next injection. Peptides were eluted directly into Q Exactive Orbitrap Mass Spectrometer operating in positive mode (resolution 70000, AGC target 3 × 106, maximum IT 50 ms, scan range 300–1,500 m/z). For each protein variant, a tandem mass spectrometry experiment was performed (full MS settings were the same as above, and data-dependent MS2 settings included 17,500,

automatic gain control target of 2 × 105, maximum injection time of 100 ms, loop count of 10, isolation window of 2.0 m/z, normalized collision energy of 28, charge state of 1 and ≥7 excluded and dynamic exclusion of 15 s) on undeuterated samples. LC and MS methods were run using Xcalibur 4.1 (Thermo Scientific).

Peptide Identification:

Peptides were identified from tandem MS data using Byonic (Protein Metrics) as described previously (Costello et al., 2022). The sequence of the expressed construct with the angiotensin peptide sequence added to the C-terminal of the sequence, if present, was used as the search reference. The search parameters were set as follows: specificity to nonspecific, precursor mass tolerance to 6 ppm, and fragment mass tolerance to 20 ppm.

HDExaminer 3 Analysis:

Peptide isotope distributions at each time point were fit in HDExaminer 3 as previously described (Costello et al. 2022). Peptide isotope distributions for each time point were fit in HDExaminer 3. Deuteration levels were determined by subtracting the mass centroids of deuterated peptides from undeuterated control peptides. All analyzed peptides showed EX2 behavior or no exchange.

Acknowledgements: We would like to thank Sophie Shoemaker for help with HDX-MS, Eva Gerber for technical advice, the entire Marqusee Lab for advice and suggestions, and David Agard and his lab for advice and reagents. This work was funded by a grant from the NIH (GM 149319). SM is a Chan Zuckerberg Biohub Investigator.

References:

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. doi: 10.1038/s41586-024-07487-w

Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, *181*(4096), 223–230. Retrieved from www.jstor.org/stable/1736447

Baker, D., & Agard, D. A. (1994). Kinetics versus thermodynamics in protein folding. *Biochemistry*, *33*(24), 7505–7509.

Baker, D., Sohl, J. L., & Agard, D. A. (1992). A protein-folding reaction under kinetic control. *Nature*, *356*(6366), 263–265. doi: 10.1038/356263a0

Barrick, D., & Baldwin, R. L. (1993). Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry*, *32*(14), 3790–3796. doi: 10.1021/bi00065a035

Bryan, P. N. (2002). Prodomains and Protein Folding Catalysis. *Chemical Reviews*, *102*(12), 4805–4816. doi: 10.1021/cr010190b

Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., … Gallagher, T. (1995). Catalysis of a protein folding reaction: Mechanistic implications of the 2.0 .ANG. structure of the subtilisin-prodomain complex. *Biochemistry*, *34*(32), 10310–10318. doi: 10.1021/bi00032a026

Costello, S. M., Shoemaker, S. R., Hobbs, H. T., Nguyen, A. W., Hsieh, C.-L., Maynard, J. A., … Marqusee, S. (2022). The SARS-CoV-2 spike reversibly samples an open-trimer conformation exposing novel epitopes. *Nature Structural & Molecular Biology*, *29*(3), 229–238. doi: 10.1038/s41594-022-00735-5

Eder, J., & Fersht, A. R. (1995). Pro-sequence-assisted protein folding. *Molecular Microbiology*, *16*(4), 609–

614. doi: 10.1111/j.1365-2958.1995.tb02423.x

Fu, X., Inouye, M., & Shinde, U. (2000). Folding Pathway Mediated by an Intramolecular Chaperone: THE INHIBITORY AND CHAPERONE FUNCTIONS OF THE SUBTILISIN PROPEPTIDE ARE NOT OBLIGATORILY LINKED. *Journal of Biological Chemistry*, *275*(22), 16871–16878. doi: 10.1074/jbc.275.22.16871

Gamble, M., Künze, G., Dodson, E. J., Wilson, K. S., & Jones, D. D. (2011). Regulation of an intracellular subtilisin protease activity by a short propeptide sequence through an original combined dual mechanism. *Proceedings of the National Academy of Sciences*, *108*(9), 3536–3541. doi: 10.1073/pnas.1014229108

Graycar, T. P., Bott, R. R., Power, S. D., & Estell, D. A. (2013). Subtilisins. In *Handbook of Proteolytic Enzymes*. doi: 10.1016/B978-0-12-382219-2.00693-1

Heinrikson, R. L. (1977). [20] Applications of thermolysin in protein structural analysis. In *Methods in Enzymology* (Vol. 47, pp. 175–189). Academic Press. doi: https://doi.org/10.1016/0076-6879(77)47022-8

Hvidt, A., & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Advances in Protein Chemistry*, *21*, 287–386. doi: 10.1016/s0065-3233(08)60129-1

Jaswal, S. S., Sohl, J. L., Davis, J. H., & Agard, D. A. (2002). Energetic landscape of α-lytic protease optimizes longevity through kinetic stability. *Nature*, *415*(6869), 343–346. doi: 10.1038/415343a

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. doi: 10.1038/s41586-021-03819-2

Keil, B. (1992). *Specificity of Proteolysis*. Springer-Verlag. Retrieved from https://books.google.com/books?id=V8HwAAAAMAAJ

Keller, T. H., Seufer-Wasserthal, P., & Jones, J. B. (1991). Probing the specificity of the S1 binding site of subtilisin Carlsberg with boronic acids. *Biochemical and Biophysical Research Communications*, *176*(1), 401–405. doi: https://doi.org/10.1016/0006-291X(91)90938-4

Kobayashi, T., & Inouye, M. (1992). Functional analysis of the intramolecular chaperone: Mutational hot spots in the subtilisin pro-peptide and a secondsite suppressor mutation within the subtilisin molecule. *Journal of Molecular Biology*, *226*(4), 931–933. doi: 10.1016/0022-2836(92)91042-N

Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique*, *65*, 44–45.

Li, Y., Hu, Z., Jordan, F., & 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. *The Journal of Biological Chemistry*, *270*(42), 25127–25132. doi: 10.1074/jbc.270.42.25127 [doi]

Lindquist, R. N., & Terry, C. (1974). Inhibition of subtilisin by boronic acids, potential analogs of tetrahedral reaction intermediates. *Archives of Biochemistry and Biophysics*, *160*(1), 135–144. doi: https://doi.org/10.1016/S0003-9861(74)80018-4

Nixon, C. F., Lim, S. A., Sailer, Z. R., Zheludev, I. N., Gee, C. L., Kelch, B. A., … Marqusee, S. (2021). Exploring the Evolutionary History of Kinetic Stability in the α-Lytic Protease Family. *Biochemistry*, *60*(3), 170–181. doi: 10.1021/acs.biochem.0c00720

Park, C., & Marqusee, S. (2004). Probing the high energy states in proteins by proteolysis. *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2004.08.085

Peter, S. (2014). Germination of Spores of Bacillus Species: What We Know and Do Not Know. *Journal of Bacteriology*, *196*(7), 1297–1305. doi: 10.1128/jb.01455-13

Sanchez-Ruiz, J. M. (2010). Protein kinetic stability. *Biophysical Chemistry*. doi: 10.1016/j.bpc.2010.02.004

Sohl, J. L., Jaswal, S. S., & Agard, D. A. (1998). Unfolded conformations of α-lytic protease are more stable than its native state. *Nature*, *395*(6704), 817–819. doi: 10.1038/27470

Spudich, G. M., Miller, E. J., & Marqusee, S. (2004). Destabilization of the Escherichia coli RNase H Kinetic Intermediate: Switching Between a Two-state and Three-state Folding Mechanism. *Journal of Molecular Biology*, *335*(2), 609–618. doi: https://doi.org/10.1016/j.jmb.2003.10.052

Street, T. O., Courtemanche, N., & Barrick, D. (2008). Protein Folding and Stability Using Denaturants. *Methods in Cell Biology*. doi: 10.1016/S0091-679X(07)84011-8

Strongin, A. Y., Izotova, L. S., Abramov, Z. T., Gorodetsky, D. I., Ermakova, L. M., Baratova, L. A., … Stepanov, V. M. (1978). Intracellular serine protease of Bacillus subtilis: sequence homology with extracellular subtilisins. *Journal of Bacteriology*, *133*(3), 1401. Retrieved from http://jb.asm.org/content/133/3/1401.abstract

Subbian, E., Yabuta, Y., & Shinde, U. (2004). Positive Selection Dictates the Choice between Kinetic and Thermodynamic Protein Folding and Stability in Subtilases. *Biochemistry*, *43*(45), 14348–14360. doi: 10.1021/bi048397x

Takagi, H., Maeda, T., Ohtsu, I., Tsai, Y.-C., & Nakamori, S. (1996). Restriction of substrate specificity of subtilisin E by introduction of a side chain into a conserved glycine residue. *FEBS Letters*, *395*(2–3), 127–132. doi: https://doi.org/10.1016/0014-5793(96)01014-9

Turnbull, P. C. B. (1996). Bacillus. In S. Baron (Ed.), *Medical Microbiology* (4th ed.). Galveston: University of Texas Medical Branch at Galveston.

Vévodová, J., Gamble, M., Künze, G., Ariza, A., Dodson, E., Jones, D. D., & Wilson, K. S. (2010). Crystal Structure of an Intracellular Subtilisin Reveals Novel Structural Features Unique to this Subtilisin Family. *Structure*, *18*(6), 744–755. doi: 10.1016/j.str.2010.03.008

Wells, J A, Cunningham, B. C., Graycar, T. P., & Estell, D. A. (1987). Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering. *Proceedings of the National Academy of Sciences*, *84*(15), 5167–5171. doi: 10.1073/pnas.84.15.5167

Wells, James A, & Estell, D. A. (1988). Subtilisin - an enzyme designed to be engineered. *Trends in Biochemical Sciences*, *13*(8), 291–297. doi: 10.1016/0968-0004(88)90121-1

Yabuta, Y., Takagi, H., Inouye, M., & Shinde, U. (2001). Folding Pathway Mediated by an Intramolecular Chaperone: PROPEPTIDE RELEASE MODULATES ACTIVATION PRECISION OF PRO-SUBTILISIN. *Journal of Biological Chemistry*, *276*(48), 44427–44434. doi: 10.1074/jbc.M107573200

You, L., & Arnold, F. H. (1996). Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Engineering, Design and Selection*, *9*(1), 77–83. doi: 10.1093/protein/9.1.77

Young, T. A., Skordalakes, E., & Marqusee, S. (2007). Comparison of Proteolytic Susceptibility in Phosphoglycerate Kinases from Yeast and E. coli: Modulation of Conformational Ensembles Without Altering Structure or Stability. *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2007.02.077

Zhu, X., Ohta, Y., Jordan, F., & Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature*, *339*(6224), 483–484. doi: 10.1038/339483a0

Tables:

Table 1: Thermodynamic and kinetic properties of ISP1 S250A, Pro-ISP1 S250A, SbtE, and Pro-SbtE

	ISP1 S250A	Pro-ISP1 S250A	SbtE	Pro-SbtE[a]
ΔG_{NI} (kcal mol^{-1})	5.88 ± 0.36	7.33 ± 0.05	NA	1.62
m_{NI} (kcal mol^{-1} M^{-1})	6.36 ± 0.51	6.93 ± 0.50	NA	1.11
ΔG_{IU} (kcal mol^{-1})	1.86 ± 0.50	1.23 ± 0.29	NA	8.36
m_{IU} (kcal mol^{-1} M^{-1})	0.93 ± 0.02	0.735 ± 0.027	NA	
ΔG_{NU} (kcal mol^{-1})	7.74 ± 0.86	8.56 ± 0.34	<−3.73**	9.98
m_{NU} (kcal mol^{-1} M^{-1})	4.2 ± 0.5	6.1 ± 0.2		3.02
k_u (s^{-1})	1.43E−1 ± 5.21E−2	8.85E−2 ± 7.33E−3	1.92E−7 ± 1.27E−7$^{pH\ 8}$, 2.8E−5 ± 1.9E−5$^{pH\ 5}$	1E−1
k_f (s^{-1})	>180	>180	<4.92E−8*$^{25°C}$, <5.07E−8*$^{4°C}$	20.23, 1.2E−2

*Estimated base on lower limit of detection for refolding by activity, **Calculated based on the ratio of the kinetic measurements, (a) As previously reported (Subbian et al., 2004)

Figure legends:

Figure 1: Comparison of the structures and sequences of intracellular and extracellular subtiliases. A) Ribbon representation of ISP1, with the pro-domain shown in yellow the core alpha helix shown in orange, and the active serine side chain as a stick, seen from the front and bottom. Residues 301-319 are not shown. The catalytic triad consists of Ser 250, His 91, and Asp 54 (numbering based on PDB entry 2WV7 (Vévodová et al., 2010)). The structure was generated with AlphaFold3 (Abramson et al., 2024). Ribbon representation of SbtE, with the pro-domain shown in light blue, the core alpha helix shown in teal, and the active site serine side chain as a stick, seen from the front and bottom (PDB entry 1SCJ). The catalytic triad consists of Ser 221, His 64, and Asp 32 (numbering based on PDB entry 1SCJ). **B)** Representation of sequence alignment of ISP1 and SbtE – the pre-sequence from SbtE was not included in the alignment. The length of the alignment is represented by the black line from left to right with the N-terminus shown in green and C-terminus shown in red. Breaks in the line represent insertions into either ISP1 (first three breaks) or SbtE (remaining breaks). Beta sheets and alpha helices present in each protein are represented above and below the alignment. Any secondary elements spanning a break are connected by a dashed line. The ISP1 pro-domain is marked by the light orange box at the top and the SbtE pro-domain is marked by the light blue box at the bottom. ISP1 and SbtE have 45.7% sequence identity and conserved residues are denoted by orange boxes along the sequence.

Figure 2: Equilibrium denaturation and unfolding kinetics demonstrate that ISP1 behaves as a typical thermodynamically stable protein and SbtE is slow to unfold A) Representative guanidinium denaturation curves of Pro-ISP1 S250A (filled circles) and ISP1 S250A (open circles) monitored by CD at 222 nm. Three-state fits for Pro-ISP1 S250A (solid line) and ISP1 S250A (dashed line) are shown. **B)** Unfolding as a function of guanidinium concentration for Pro-ISP1 S250A (filled circles, orange), ISP1 S250A (open circles, orange), and SbtE (open squares, blue) as monitored by the CD signal at 222 nm. Linear regressions for Pro-ISP1 S250A (solid line, orange), ISP1 S250A (dashed line, orange), and SbtE

(dashed line, blue) are shown. All data represented here was collected at pH 8 (see *Materials and Methods).*

Figure 3: Enzymatic assay reports on slow refolding rate of SbtE Fraction folded SbtE is plotted as a function of time for refolding at 25°C (left) and 4°C (right). Fits to $k_{f=}$ $([N]/[I]_0)/t$ are shown (Sohl et al., 1998). No activity was detected in the refolding sample over the course of six days. There is a loss of intact SbtE, competent for folding with pro-domain *in trans*, overtime indicating the presence of folded, active SbtE. The lower detections limit of the assay (2E−10 M SbtE) sets an upper limit on the total amount of folded, active protein.

Figure 4: HXD-MS analysis shows local dynamics are modulated by the pro-domain in opposite manners for ISP1 and SbtE A) Difference Woods plots for ISP1 S250A vs Pro-ISP1 S250A (top) and SbtE vs Pro-SbtE S221A (bottom). Each peptide is plotted, and color coded by time point: 60 seconds in red, 30 minutes in orange, 24 hours in cyan, 48 hours in blue, and 96 hours in purple. Positive values indicate more deuteration (less protection) in the pro-proteins and negative values indicate less deuteration (more protection) in the first protein listed. Regions corresponding to uptakes plots in b are boxed. **B)** Uptake plots showing average number of deuterons per protein as a function of time for representative peptides (average of three replicates, error bars represent standard deviation not back exchange corrected). Plots are colored coded by protein: Pro-ISP1 S250A in orange, ISP1 S250A in dark orange, Pro-SbtE S221A in light blue, SbtE in dark blue. Pro-ISP1 S250A's core helix exchanges with a k_{obs} of ~10^{-4} s^{-1}, ISP1 S250A and Pro-SbtE S221A's core helix exchanges with a k_{obs} of >10^{-1} s^{-1}. ISP1 S250A's core helix exchanges with a k_{obs} of >10^{-1} s^{-1} and SbtE's core helix exchanges with a k_{obs} of <10^{-8} s^{-1}. Pro-ISP1 S250A's calcium binding loops and core beta sheets exchange with a k_{obs} of ~10^{-5}-10^{-6} s^{-1}. ISP1 S250A's calcium binding loops and core beta sheets exchange with a k_{obs} of >10^{-1} s^{-1}. Pro-SbtE's calcium binding loops and core beta sheets all exchange with a k_{obs} of >10^{-1} s^{-1}. SbtE's calcium binding loops and core beta sheets exchange with a k_{obs} of ~10^{-5}-10^{-7} s^{-1}. Pro-ISP1 S250A's substrate binding loops exchange with a k_{obs} of ~10^{-1} s^{-1} and ISP1

S250A's exchange with a k_{obs} of ~10^{-5} s^{-1}. Pro-SbtE S221A's substrate binding loops exchange with a k_{obs} of ~10^{-4} s^{-1} and SbtE's exchange with a k_{obs} of ~10^{-1} s^{-1}.

Figure 5: Subtilisin outlasts thermolysin in an *in vitro* competition A) Representative SDS-PAGE gels showing thermolysin is rapidly degraded but SbtE persists over a 24-hour time course. **B)** Quantification of gel bands from part A. The band from each time point was normalized to their respective 0 second time point for each protease and plotted as a function of time.

Figure 6: ISPs have a conserved insertion at the C-terminus of the core alpha helix Left – Section of sequence alignment showing core alpha helix sequence from ISPs (top of alignment) and ESPs (bottom of alignment) from Bacillus species. ISP core alpha helix region is highlighted by orange ribbon cartoon and ESP core alpha helix region is highlighted by blue ribbon cartoon. Right – Slice of structural alignment of ISP1 and SbtE showing elongated core alpha helix in ISP1 (orange, AlphaFold3 model) versus the core alpha helix in SbtE (blue, PDB: 1SCJ).

Figures:

Figure 1

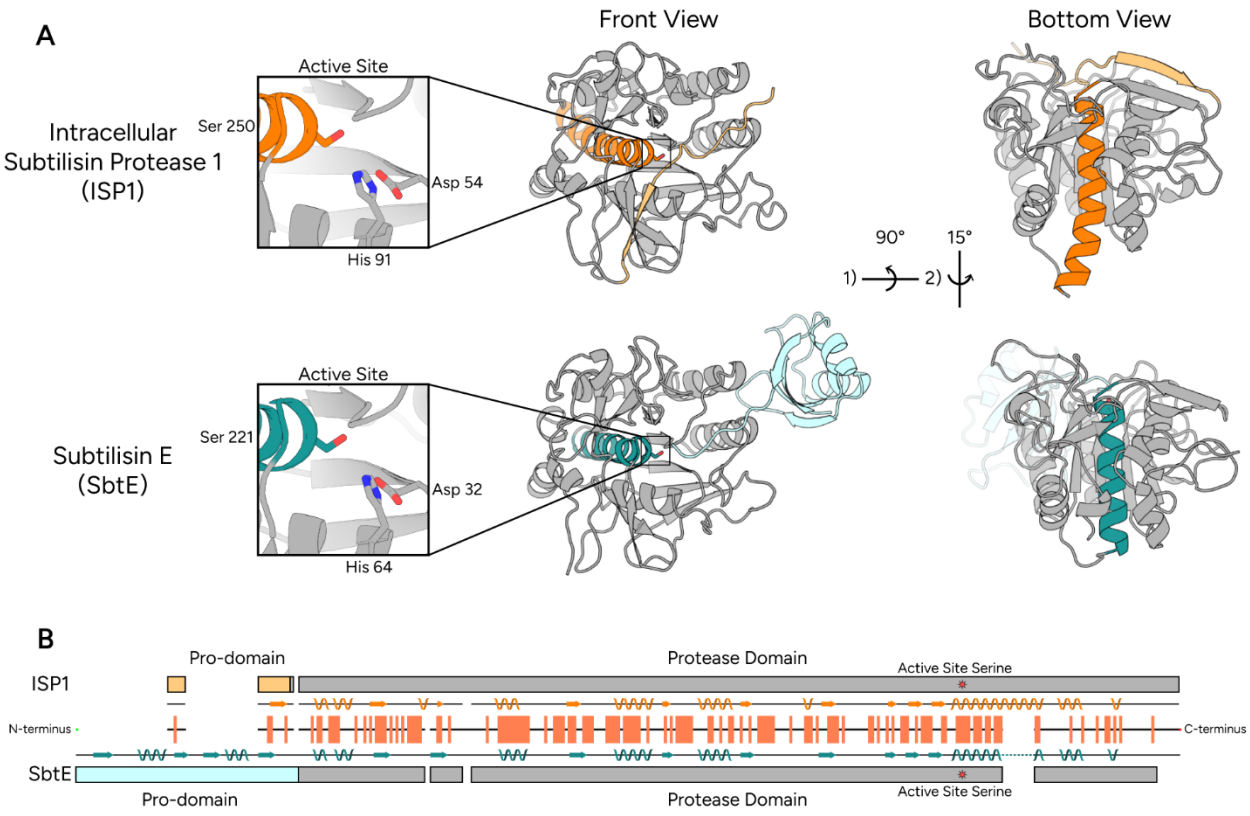

Figure 2

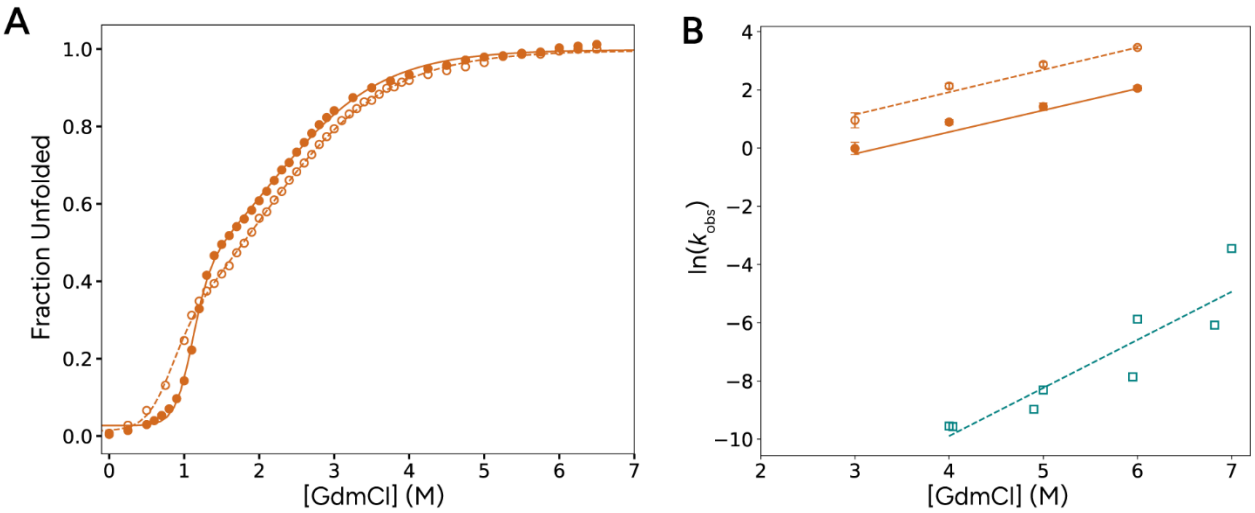

Figure 3

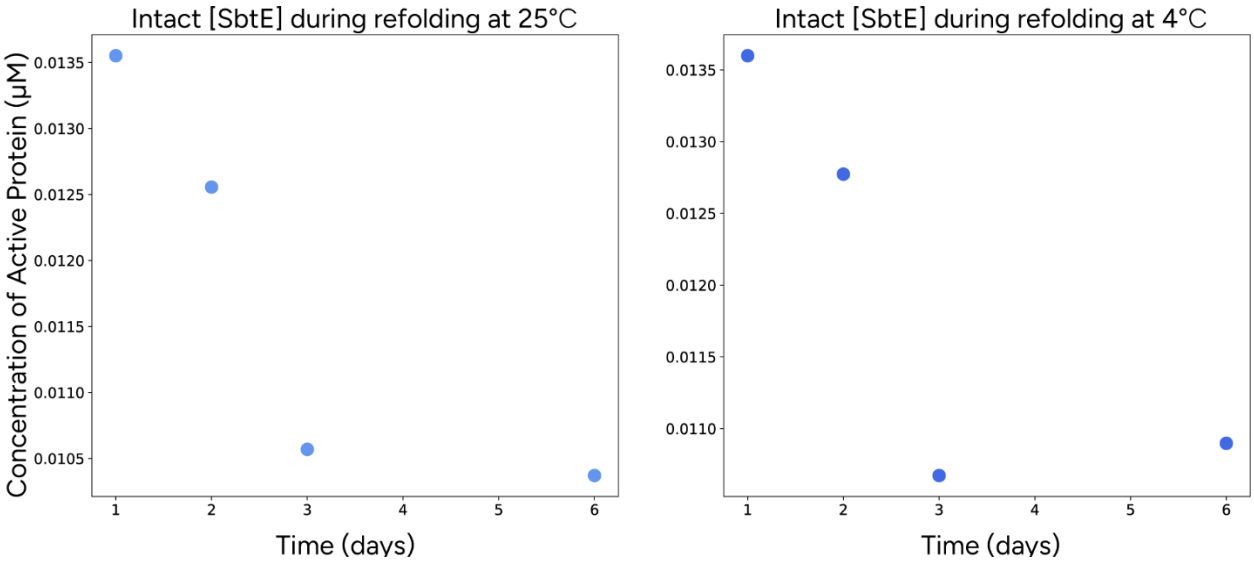

Figure 4

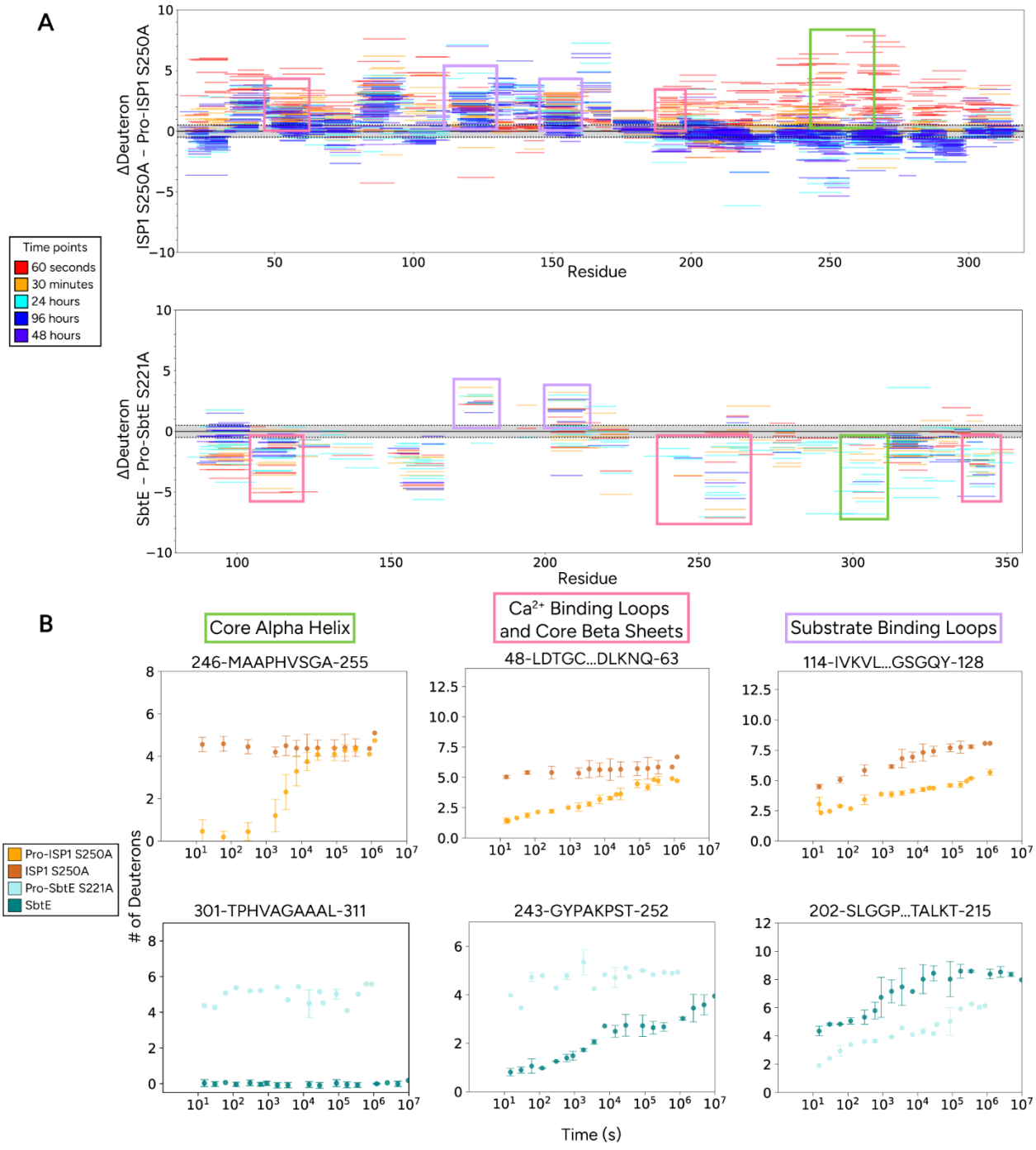

Figure 5

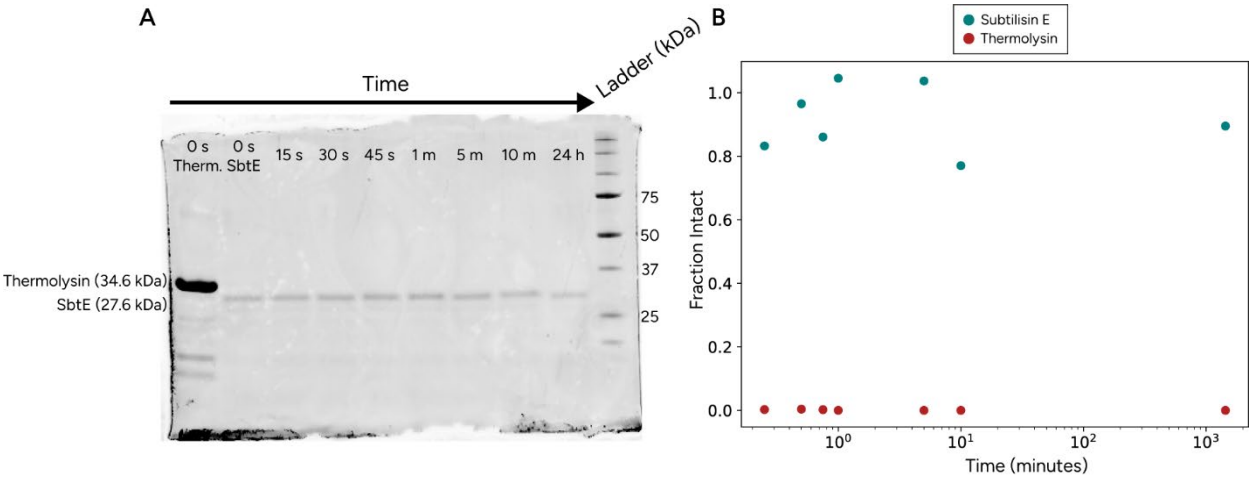

Figure 6