



Conserved Prosegment Residues Stabilize a Late-Stage Folding Transition State of Pepsin Independently of Ground States

Derek R. Dee^{1*}, Yasumi Horimoto², Rickey Y. Yada^{1,2*}

1 Biophysics Interdepartmental Group, University of Guelph, Guelph, Ontario, Canada, **2** Department of Food Science, University of Guelph, Guelph, Ontario, Canada

Abstract

The native folding of certain zymogen-derived enzymes is completely dependent upon a prosegment domain to stabilize the folding transition state, thereby catalyzing the folding reaction. Generally little is known about how the prosegment accomplishes this task. It was previously shown that the prosegment catalyzes a late-stage folding transition between a stable misfolded state and the native state of pepsin. In this study, the contributions of specific prosegment residues to catalyzing pepsin folding were investigated by introducing individual Ala substitutions and measuring the effects on the bimolecular folding reaction between the prosegment peptide and pepsin. The effects of mutations on the free energies of the individual misfolded and native ground states and the transition state were compared using measurements of prosegment-pepsin binding and folding kinetics. Five out of the seven prosegment residues examined yielded relatively large kinetic effects and minimal ground state perturbations upon mutation, findings which indicate that these residues form strengthened and/or non-native contacts in the transition state. These five residues are semi- to strictly conserved, while only a non-conserved residue had no kinetic effect. One conserved residue was shown to form native structure in the transition state. These results indicated that the prosegment, which is only 44 residues long, has evolved a high density of contacts that preferentially stabilize the folding transition state over the ground states. It is postulated that the prosegment forms extensive non-native contacts during the process of catalyzing correct inter- and intra-domain contacts during the final stages of folding. These results have implications for understanding the folding of multi-domain proteins and for the evolution of prosegment-catalyzed folding.

Citation: Dee DR, Horimoto Y, Yada RY (2014) Conserved Prosegment Residues Stabilize a Late-Stage Folding Transition State of Pepsin Independently of Ground States. PLoS ONE 9(7): e101339. doi:10.1371/journal.pone.0101339

Editor: Jose M. Sanchez-Ruiz, Universidad de Granada, Spain

Received: August 8, 2013; **Accepted:** June 5, 2014; **Published:** July 1, 2014

Copyright: © 2014 Dee et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC) (http://www.nserc-crsng.gc.ca/index_eng.a?sp) and the Canada Research Chairs (CRC) Program (<http://www.chairs-chaires.gc.ca/home-accueil-eng.aspx>). DRD was also supported by an Ontario Graduate Scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: ryada@uoguelph.ca

† Current address: Department of Physics, University of Alberta, Edmonton, Alberta, Canada

Introduction

General thermodynamic and kinetic features of protein folding are known, such as that the natively folded conformation of a protein is thermodynamically stabilized [1] and that there is a relationship between topology and folding rate, which holds for both two-state and multi-state folding proteins [2,3]. However, zymogen-derived proteins which require a prosegment (PS) domain to catalyze folding, such as the serine peptidases α LP [4], SGPB [5], and subtilisin [6], and the aspartic peptidase pepsin [7], deviate from the common thermodynamic and kinetic trends in protein folding to varying degrees. For example, as shown in **Fig 1A**, the native states of α LP, pepsin and SGPB are thermodynamically unstable ($\Delta G_{N-I} = -4$ kcal/mol) [8], metastable ($\Delta G_{N-I} = -0.1$ kcal/mol) [7] and marginally stable ($\Delta G_{N-I} = +1$ kcal/mol) [9], relative to an intermediate state, respectively. Additionally, in the absence of the PS domain, these proteins fold much slower than would be estimated based on their topology (**Fig 1B**).

When the PS is included these deviations are corrected: the PS shifts the folding equilibrium towards the PS-native state complex

and catalyzes folding by stabilizing the folding transition state (TS) (bold, blue line in **Fig 1A**). Once folding is complete, the PS is removed and the folding and unfolding activation barriers increase, leaving behind a kinetically trapped native state (black line in **Fig 1A**). There is an intriguing separation of low- and high-barrier folding landscapes, with and without the PS, respectively, and understanding how a PS domain stabilizes the folding TS should be informative for understanding kinetic folding/unfolding barriers in general, which remain poorly understood [10]. Despite several studies of PS-catalyzed folding [11–14], the mechanism by which a PS stabilizes the TS remains unknown.

Pepsin, which is derived from its zymogen pepsinogen, folds to a thermodynamically stable yet non-native form (termed refolded pepsin, Rp) upon removal from denaturing conditions [15]. Rp is inactive, contains native-like secondary and tertiary structure and has a greater thermal stability ($\Delta T_m = +5^\circ\text{C}$) [15] and reduced picosecond diffusive motions when compared to native pepsin (Np) [16]. These features suggest that Rp is a late-stage folding intermediate, which in turn indicates that the PS operates late in the pepsinogen folding pathway.

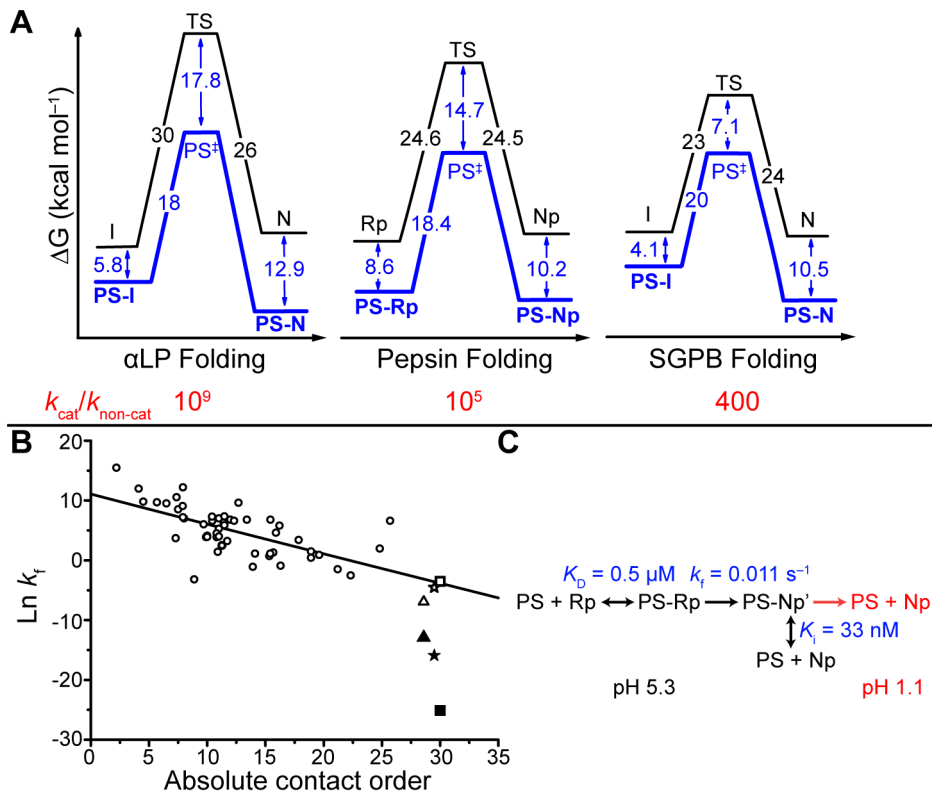


Figure 1. Zymogen-derived proteins deviate from common trends in protein folding. (A) Comparison of the non- and PS-catalyzed folding of α LP [8], SGPB [5] and pepsin [7]. (B) Relation between topology and folding rate for a number of two- and three-state folding proteins (circles, data taken from [3,21,22]). The folding rate of α LP (squares), SGPB (triangles) and pepsin (stars) is accelerated to the value (hollow points) expected based on the topology, only when the PS is included. (C) Reaction scheme of pepsin PS-catalyzed folding. The PS binds Rp and catalyzes its conversion to Np at pH 5.3, where the PS is a strong inhibitor of Np. The PS dissociates from Np at pH < 3. doi:10.1371/journal.pone.0101339.g001

The folding of Rp to Np serves as a useful model for examining late-stage folding transitions between compact-misfolded and native states, and under native conditions (*i.e.*, pH 5.3 and no denaturants), which are often difficult to access experimentally. Indeed, the study of such misfolding events is often only possible at the single-molecule level [17]. Given the increased risk of misfolding in multi-domain proteins, via incorrect domain-domain contacts [18,19], it is prudent to examine the various mechanisms by which proteins have evolved to avoid such issues.

The present study undertook to examine the energy landscape of PS-catalyzed pepsin folding by measuring the contributions of specific PS residues to stabilizing the folding TS and ground states (Rp and Np). Our method took advantage of the bimolecular PS-catalyzed folding reaction of pepsin (**Fig 1C**). Synthetic PS peptide was added exogenously to Rp and Np and the changes in equilibrium stability of PS-Np relative to PS-Rp upon mutation, $\Delta\Delta G_{PS(Np-Rp)}$, were determined from the difference in the changes in binding energies, as shown in **Fig S1**. Changes in activation energy upon mutation, $\Delta\Delta G^\ddagger$, were determined by measuring changes in the rate of PS-catalyzed folding.

A simple comparison of $\Delta\Delta G^\ddagger$ and $\Delta\Delta G_{PS(Np-Rp)}$ yields the so-called Φ -value [20]. Φ -values are calculated as the ratio of the change in activation energy to the change in equilibrium stability ($\Phi = \Delta\Delta G^\ddagger / \Delta\Delta G_{N-D}$) upon introducing a point mutation. Under the classical interpretation, a residue can belong to a region of either unfolded or native structure in the TS, giving rise to the limiting Φ -values of 0 or 1 [23]; however, a range of fractional Φ -values are more commonly observed [24,25]. Φ -value analysis is

often applied to studies of unimolecular folding [26–31], in which all energies are measured relative to one state, generally the unfolded state (and thus mutation effects on this state go unresolved [32]). Compared to Φ -value analysis of unimolecular folding, the bimolecular approach allows a comparison of the effects of point mutants on each individual ground state and the TS. The $\Delta\Delta G$ values obtained for each individual state are more informative than the Φ -values derived from them, yet Φ -values are also reported here as a point of comparison.

Application of Φ -value analysis to the PS-catalyzed folding of pepsin yielded predominantly abnormal Φ -values ($\Phi > 1$, $\Phi < 0$), reflecting the finding that most of the mutations resulted in a greater destabilization of the PS-TS than of either PS-Rp or PS-Np. This greater sensitivity to perturbation of the PS-TS complex likely indicates either the presence of strong non-native interactions or reduced conformational strain in the TS, or a combination of both factors.

Results

Selection of point mutations

PS mutations were chosen on the basis of examining sequence conservation and the available zymogen crystal structures of porcine pepsinogen and human progastricsin (**Fig S2**). As proteins are believed to have evolved stable, mutually supportive native contacts to avoid degeneracy on the folding landscape (*i.e.*, minimally frustrated contacts) [33], highly conserved residues may be particularly important for proper folding. A conserved domain

search within the Pfam database [34] revealed that PS residues 1–29 of pepsinogen correspond to the A1 propeptide conserved motif, which is found in 402 sequences from 107 animal species. Pepsinogen homologues were identified using a sequence similarity search and these were further aligned, focusing on the 44-residue PS domain, revealing that PS residues L6 and R13 are strictly conserved (**Fig S3**). Only the R13 and K36 side-chains of the PS form hydrogen bonds with the mature pepsin domain, with hydrogen bonds formed between R13 and pepsin D11 and between K36 and the catalytic residues of pepsin, D32 and D215. Seven residues were chosen for mutation, shown in **Fig 2A**: the strictly conserved L6 and R13, the semi-conserved V4, S11, F25 and K36, and the non-conserved I17. F25 was chosen primarily to probe the effect of mutating the second α -helical segment, which makes no direct contacts with the pepsin domain, such that the N-terminal β -strand and all three α -helical segments of the PS were probed with mutations.

The residues were replaced with Ala in the corresponding seven synthetic peptides. Ala substitutions were shown to be generally conservative mutations that do not introduce non-native interactions, which would further complicate data analysis, while at the same time provide a measurable destabilization [32].

Effects of PS mutants on binding and folding catalysis

The PS-catalyzed folding and binding data are summarized in **Fig's 2** and **S4** and **Table 1**. All of the mutants markedly slowed the rate of PS-catalyzed folding except for I17A, which had no effect. The L6A mutant gave the slowest folding rate, while the other mutants resulted in similar rates. The narrow range of effects that mutations had on the folding rate suggests that the PS stabilizes the folding TS via contacts made along an extensive portion of the PS and not from localized contacts. For some of the mutants, *e.g.* S11A, a burst-phase in the folding kinetics was noticeable (**Fig S4**), although the basis for this feature is not yet clear. Generally the PS mutants had a small impact on PS affinity for Rp. One exception was R13A, which reduced the binding affinity 7-fold, while the effects of the other mutants were modest (2 to 3-fold reduction in affinity) or had no effect at all (I17A and F25A were similar to PS_{wt}). The mutations resulted in a wider distribution of affinities for Np, measured by PS inhibition of Np. In this case, V4A was similar to PS_{wt} while R13A and I17A resulted in the largest reduction in affinities for Np.

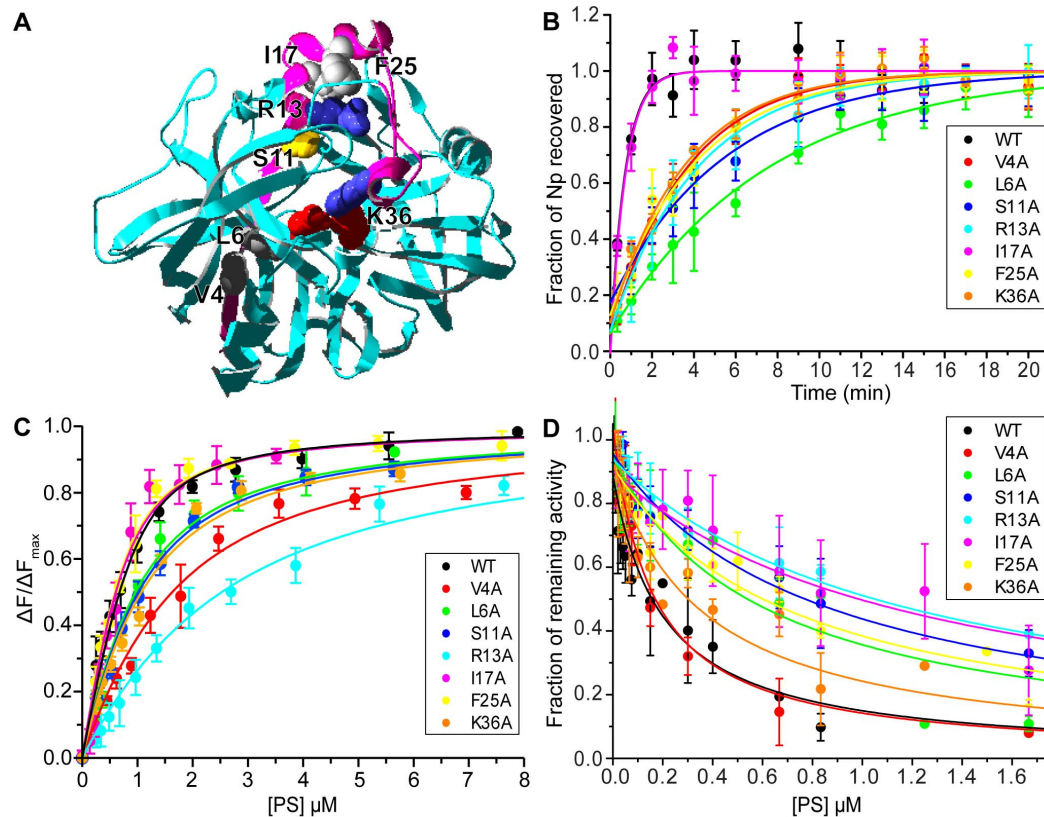


Figure 2. Effects of PS point mutants on binding and catalyzing pepsin folding. (A) Structure of pepsinogen (PDB code: 3PSG) with the PS (pink) located between the N- and C-terminal lobes, forming part of a six-stranded β -sheet, and K36 of the PS interacts with the catalytic residues, D32 and D215 (red). PS residues selected for mutation to Ala are shown in space-filling form and coloured according to type (grey-hydrophobic, orange-polar, blue-basic, red-acidic). (B) Comparison of wild-type and mutant PS-catalyzed folding of pepsin. The rate of PS-catalyzed folding (k_f) was determined by adding PS to Rp, at pH 5.3, 15°C (see **Text S2**: folding rate followed Arrhenius temp-dependence from 0–15°C, shown in **Fig S5**), and measuring the formation of Np based on enzyme activity measured at pH 1.2, 25°C. The data were fit according to a monoexponential function to obtain k_f . (C) Comparison of wild-type and mutant PS affinity for Rp. PS-Rp binding was determined by following the increase in Trp-fluorescence of pepsin as a function of [PS]. The data were fit according to eq 1 to determine the dissociation constant, K_d , at 20°C, pH 5.3. (d) Comparison of wild-type and mutant PS affinity for Np. The reduction in Np activity was measured as a function of [PS]. The data were fit according to a competitive inhibitor model, eq 2, to determine the inhibition (dissociation) constant, K_i , at 20°C, pH 5.3. All data are reported as the average \pm SD of 3–5 measurements for each PS peptide.

doi:10.1371/journal.pone.0101339.g002

Table 1. Changes in binding and folding constants^a and associated free energies^b upon mutation of the PS.

PS-mutant	k_f (min^{-1})	K_d (μM)	K_i (nM)	$\Delta\Delta G_{PS-Rp}$	$\Delta\Delta G^\ddagger$	$\Delta\Delta G_{PS-Np}$	$\Delta\Delta G_{PS(Np-Rp)}$	Φ
WT	1.42 ± 0.34	0.31 ± 0.03	64.8 ± 11.4					
V4A	0.27 ± 0.03	1.19 ± 0.23	50.9 ± 6.9	0.79 ± 0.12	0.94 ± 0.15	-0.14 ± 0.13	-0.93 ± 0.18	-1.0 ± 0.3
L6A	0.13 ± 0.01	0.64 ± 0.09	193.5 ± 38.8	0.43 ± 0.10	1.36 ± 0.15	0.64 ± 0.16	0.20 ± 0.18	6.6 ± 6.0
S11A	0.18 ± 0.02	0.68 ± 0.07	270.1 ± 31.5	0.47 ± 0.08	1.18 ± 0.15	0.83 ± 0.12	0.36 ± 0.15	3.2 ± 1.4
R13A	0.24 ± 0.02	1.96 ± 0.29	382.2 ± 37.2	1.08 ± 0.10	1.02 ± 0.15	1.03 ± 0.12	-0.05 ± 0.15	-21 ± 67
I17A	1.40 ± 0.26	0.21 ± 0.09	393.8 ± 65.1	-0.21 ± 0.24	0.01 ± 0.18	1.05 ± 0.14	1.26 ± 0.28	0.0 ± 0.1
F25A	0.26 ± 0.04	0.25 ± 0.03	230.4 ± 29.1	-0.12 ± 0.08	0.96 ± 0.17	0.74 ± 0.13	0.86 ± 0.15	1.1 ± 0.3
K36A	0.28 ± 0.03	0.72 ± 0.14	113.3 ± 20.3	0.50 ± 0.13	0.94 ± 0.15	0.33 ± 0.15	-0.17 ± 0.19	-5.5 ± 6.4

^aFolding rate constants were measured at 15°C while binding constants were determined at 20°C. Data are given as the mean \pm SD obtained from non-linear curve fitting.

^bFree energy units are in kcal/mol with \pm SD derived by propagation of errors.

doi:10.1371/journal.pone.0101339.t001

Changes in the PS-catalyzed folding energy landscape

The changes in PS-Rp and PS-Np binding energy upon mutation ($\Delta\Delta G_{PS-Rp}$ and $\Delta\Delta G_{PS-Np}$) were obtained from measurements of K_d and K_i . The change in PS-TS binding energy ($\Delta\Delta G_{PS-TS}$) was taken as the change in the folding activation energy, $\Delta\Delta G^\ddagger$, added to $\Delta\Delta G_{PS-Rp}$, as ΔG^\ddagger was determined relative to PS-Rp. The changes in binding and folding activation energies are given in **Table 1**. The effects of each PS mutation on the PS-catalyzed folding landscape are readily compared by plotting the changes in binding energies of the denatured (PS-Rp), transition (PS-TS) and native (PS-Np) states, as shown in **Fig 3A**.

PS-Rp was destabilized by mutations at V4, L6, S11 and R13, which corresponds to the N-terminal β -strand and α -helix-1 of the PS within pepsinogen (**Fig 2A**), indicating that this region may play a dominant role in defining the initial PS-Rp complex. Conversely, both I17A and F25A had a negligible effect on PS-Rp binding, indicating that the PS is likely unstructured in this region within PS-Rp. All of the mutations, except V4A, R13A, and K36A, were more destabilizing to PS-Np than to PS-Rp. V4A had a very small stabilizing effect on PS-Np and a relatively large destabilizing effect on PS-Rp. Conversely, R13A destabilized both the PS-Np and PS-Rp complexes – the similar magnitudes of

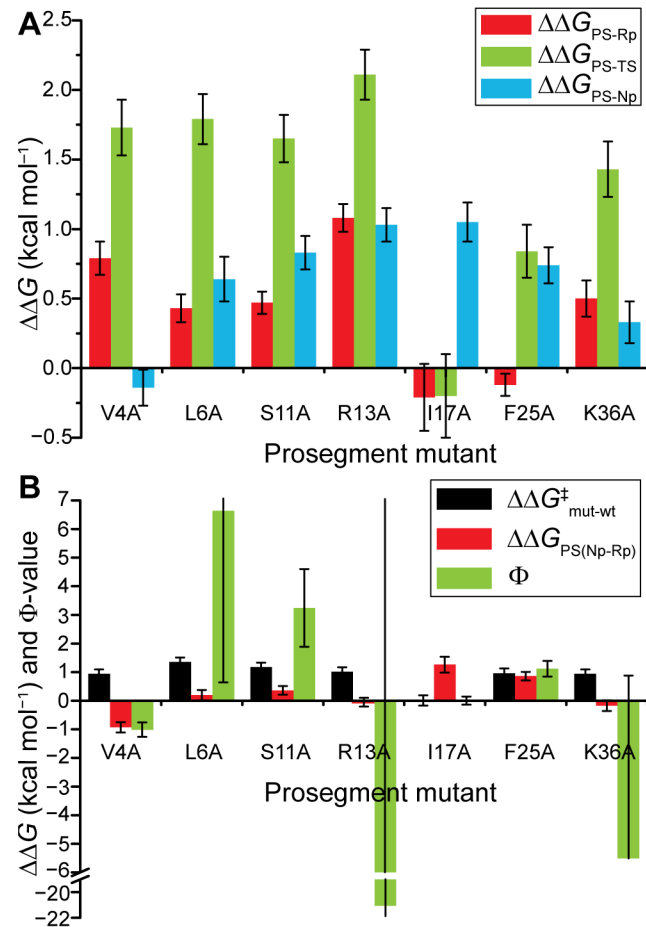


Figure 3. Changes in the PS-catalyzed folding energy landscape upon mutation of the PS peptide. (A) The changes in energy of each conformation were determined as changes in binding energies. (B) Φ -values calculated from the ratio of the changes in activation energy ($\Delta\Delta G^\ddagger$) and free energy difference between PS-Np and PS-Rp ($\Delta\Delta G_{PS(Np-Rp)}$). Error bars show \pm SD derived by propagation of errors. doi:10.1371/journal.pone.0101339.g003

$\Delta\Delta G_{PS-Rp}$ and $\Delta\Delta G_{PS-Np}$ suggest that R13 had similar contacts in both PS-Rp and PS-Np. K36A was also nearly equally destabilizing to PS-Rp and PS-Np, suggesting a similar structure in both complexes, although overall K36A had less of an impact on binding to Rp and Np than R13A.

With the exception of I17A, the mutations were most destabilizing to the PS-TS complex, particularly the mutation of the strictly conserved R13. I17A had a negligible effect on PS-Rp and PS-TS stability, yet was one of the most destabilizing mutations to the PS-Np complex, indicating that this non-conserved residue makes no contribution to catalyzing folding but contributes to driving the equilibrium towards PS-Np.

Φ -values were obtained, using the data in **Fig 3A**, by subtracting $\Delta\Delta G_{PS-Rp}$ from both $\Delta\Delta G_{PS-TS}$ (to obtain $\Delta\Delta G^\ddagger$) and $\Delta\Delta G_{PS-Np}$ (to obtain $\Delta\Delta G_{PS(Np-Rp)}$), and then dividing the first result by the second (**Fig 3B**). A Φ -value close to either 0 or 1 would indicate that a PS residue adopts a conformation identical to that in either PS-Rp or PS-Np in the PS-TS complex, respectively. As seen in **Table 1** and **Fig 3B**, most of the mutants gave rise to large positive or negative Φ -values due to the relatively small $\Delta\Delta G_{PS(Np-Rp)}$ and large $\Delta\Delta G^\ddagger$ values. L6A, R13A and K36A, in particular, yielded exceptionally large Φ -values (as are the associated errors) owing to $\Delta\Delta G_{PS(Np-Rp)}$ close to zero while $\Delta\Delta G^\ddagger$ is ~ 1 kcal/mol. As discussed below, these values likely reflect the formation of non-native interactions and/or reduced conformational strain in the TS.

Discussion

PS stabilizes TS independently of ground states

For small, two-state folding proteins, a predominance of low fractional Φ -values is interpreted as a diffuse TS with weakened native-like structure, while Φ -values clustering towards low and high fractional values is interpreted as a polarized TS structure, with some regions forming native-like contacts and others being unfolded [35]. The Φ -values presented for PS-pepsin show a very different trend: instead of fractional values, five of the seven residues were characterized by large positive or negative Φ -values indicating a highly structured TS state with strengthened interactions. A useful means by which to compare the kinetic effect of various mutations is to use a Brønsted plot [35], as shown in **Fig 4**. It can be seen that the increase in folding activation energy occurs independently of the ground state perturbation, indicating that all of the PS residues examined, except for I17, play a common role in defining the folding barrier. The R13A and K36A mutations gave particularly large kinetic effects, with $\Delta\Delta G_{PS(Np-Rp)}$ close to 0 and $\Delta\Delta G^\ddagger$ of ~ 1 kcal/mol. As strong kinetic effects were seldom observed out of a comparison of hundreds of mutations from various small, single domain proteins [24,35], this data supports the idea that the PS plays a unique role in stabilizing the folding TS.

Physical basis for large kinetic effects and abnormal Φ -values

The perturbations introduced by each mutation are characterized in most detail by the individual binding energy changes, $\Delta\Delta G_{PS-Rp}$, $\Delta\Delta G_{PS-Np}$ and $\Delta\Delta G_{PS-TS}$ (**Fig 3A**). Ala-scanning allowed for the identification of key residues that provide extra PS-TS stabilization, yet the nature of these interactions remains open to speculation. A consequence of the larger and opposite effects on the TS compared to the ground states (**Fig 4**) is that this gives rise to Φ -values that fall outside the typical range of 0 to 1. When interpreting the large kinetic effects observed in PS-catalyzed

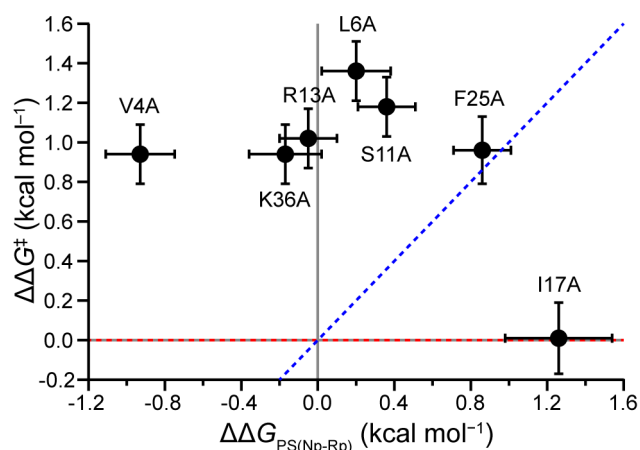


Figure 4. Brønsted plot. A comparison of the mutation effects on the folding activation energy as a function of the change in equilibrium stability. Dashed lines indicate the trend lines for $\Delta\Delta G$ values that would give rise to Φ -values of 0 or 1 and error bars show \pm SD. doi:10.1371/journal.pone.0101339.g004

pepsin folding, it is worth considering previous reports of abnormal Φ -values determined for unimolecular folding.

Although Φ -values outside the range 0 to 1 account for as much as 10 to 20% of those reported [23], it was argued that many of these unusual Φ -values are not reliable as they are associated with small $\Delta\Delta G_{N-D}$ and $\Delta\Delta G^\ddagger$ values [35]. In practice, what is considered the lower limit of $\Delta\Delta G_{N-D}$ from which reliable Φ -values may be calculated differs among reports, ranging from 1.7 [35] to 0.6 [36] to 0.2 kcal/mol [30]. Fortunately, analysis of pepsin PS-catalyzed folding did not rely on direct measurements of the generally small values of $\Delta\Delta G_{PS(Np-Rp)}$, as the individual binding energies provide specific detail on the perturbations introduced upon mutation to each of the PS-Rp and PS-Np complexes (**Fig 3A**). Even in the cases where $\Delta\Delta G_{PS(Np-Rp)}$ is close to 0 (e.g., L6A, R13A and K36A), the $\Delta\Delta G^\ddagger$ values are substantial (~ 1 kcal/mol) such that the associated Φ -values can be reliably classified as ‘abnormal’ (outside the range 0 to 1), even if they cannot be measured quantitatively owing to the correspondingly large errors. For example, R13A yields a Φ -value of -21 ± 67 that is unreliable statistically, yet the underlying comparison that the Φ -value represents, that $\Delta\Delta G^\ddagger = 1.02 \pm 0.15 > \Delta\Delta G_{PS(Np-Rp)} = -0.05 \pm 0.15$ (all units in kcal/mol), is reliable.

It was previously observed that abnormal Φ -values occur more frequently for mutations that could lead to changes in stability or configurational dynamics of the denatured state, such as mutation of charged and polar residues or of Ala and Gly [37]. As the pepsin PS is unstructured on its own, as verified by CD (**Fig S6**), mutations would have minimal effects on the structure and stability of the PS, and the measured changes in binding energies can be ascribed entirely to changes in PS-pepsin interactions.

An analysis of 806 mutants from 24 proteins indicated that Φ -values are strongly influenced by packing density and local interactions [24]. Residues at the surface tend to make fewer and more localized contacts than internal residues and thus can adopt a native-like structure when only a few local contacts are formed. Furthermore, the $\Delta\Delta G_{N-D}$ values for mutations at locations with few contacts are generally smaller than those for buried residues with many native contacts. Thus, mutations of surface residues result in both smaller $\Delta\Delta G_{N-D}$ and larger Φ -values, while mutations at core residues tend to give larger $\Delta\Delta G_{N-D}$ and smaller Φ -values [24,37]. The PS residues examined in the present study

are mostly buried from the solvent yet are not core residues, as the PS sits at the surface (**Fig 3A**). Fewer native contacts made by PS residues could explain the relatively small $\Delta\Delta G_{PS(Np-Rp)}$ values, yet it does not explain the larger effects on $\Delta\Delta G^\ddagger$.

Φ -values outside the classical range of 0 to 1 are the result of opposite or larger energetic effects in the TS than in the native state, although the microscopic basis for this is not certain. One hypothesis is that these unusual Φ -values arise from alternative flow channels down the folding funnel [38], with alternative folding paths (and thus a different TS to be crossed) becoming more predominant upon introduction of a point mutation. While the notion of alternative flow channels is consistent with the view of a funneled folding landscape, the only supporting evidence to date has come from native-centric lattice Gō models [38]. An alternative interpretation is that unusual Φ -values can arise when side-chains form non-native contacts in the TS [39,40]. Both experiments and simulations have shown that non-native interactions can accelerate or decelerate folding [39–43]. Suspected non-native interactions were found to involve both hydrophobic and electrostatic interactions, and these can stabilize or destabilize the TS [26,40,44]. In addition to non-native interactions, unusual Φ -values may arise upon mutation of a group that experiences different conformational strain in the TS and native state. Mutations that change the size of hydrophobic side-chains can stabilize the TS (by optimizing side-group packing) while destabilizing the native state, due to the different compactness of the TS, thereby producing negative Φ -values [30]. Similarly, mutations that stabilize the native state yet destabilize the TS due to differences in conformational strain also produce negative Φ -values [45]. Conformational strain present in the TS but not in the native state can also give rise to Φ -values >1 [46]. In PS-pepsin the V4A mutation was stabilizing to the native state and destabilizing to the TS, suggestive of a slightly frustrated or overly-packed native state and a more optimally packed TS. In fact, all the anomalous Φ -values for PS-pepsin (both negative and positive) resulted from $\Delta\Delta G^\ddagger > \Delta\Delta G_{PS(Np-Rp)}$.

The latter result is consistent with the finding that the PS has picomolar affinity for the TS compared to nano- and micromolar affinity for Np and Rp, respectively [7]. It seems likely that the PS has achieved this higher affinity for the TS via a concerted optimization of side-chain packing, hydrogen bonding and electrostatics, as evidenced by polar, hydrophobic and charged groups selectively stabilizing the TS (**Fig 3A**). Furthermore, this concerted optimization likely involves both non-native and strengthened native contacts, given the range of interactions involved. In the native fold [47], PS residues R8, R13 and K36 form ion pairs with pepsin residues E13, D11 and D32, D215, respectively, and these may be optimized in PS-TS. S11 is flanked on either side by R8 and R13 and thus may influence the strength of these interactions. V4 and L6 are likely optimally packed in the TS compared to in PS-Rp and PS-Np, in line with previous evidence that non-native packing accelerates folding [39–43]. Further insight into the nature of these contacts may be gleaned from future studies involving double-mutant cycles (*e.g.*, to study the influence of ion-pairs) and the systematic reduction of side-chain size (*e.g.*, Val \rightarrow Ala \rightarrow Gly), and both approaches have been used previously [20,26,27,40].

Coarse-grain simulations have been used to understand the nature of non-native interactions in small, single-domain proteins lacking stable folding intermediates [40,48,49]. To our knowledge, such approaches have not yet been applied to larger proteins, although the folding of DehI, a 311-residue protein with a knotted-fold, was simulated using a native-centric Gō model that did not include non-native contacts [50]. It would be challenging, but not

impossible, to use similar coarse-grain approaches to model the folding of pepsinogen (370 residues). Using simulation to gain insight into the nature of the non-native contacts formed within PS-TS would be greatly facilitated by knowing the high-resolution structure of PS-Rp. This would require characterizing the PS-Rp complex before it folded to PS-Np. The data presented here (**Fig 3A**) indicate that PS-Rp could be ‘trapped’ for further structural analysis by using a PS with a double mutation (such as PS_{I17A/F25A}), which would be expected to shift the folding equilibrium from PS_{I17A/F25A}-Np to PS_{I17A/F25A}-Rp. This hypothesis was confirmed recently using ¹H-¹⁵N TROSY NMR to show that PS_{I17A/F25A}-Rp is structurally very similar to Rp alone [51].

Only I17A ($\Phi = 0$) and F25A ($\Phi = 1.1$) gave typical Φ -values, indicating that I17 adopts native-like structure after formation of the TS, while F25 adopts native-like structure during formation of the TS. F25 is located on the second α -helix of the PS, which runs across the top of the active site cleft in pepsinogen (**Fig’s 2A and S2**), the C-terminus of which (residue 29) marks the end of the conserved A1 propeptide motif [34]. Given that F25A yielded a Φ -value close to 1 indicates that this α -helical segment may be structured in the TS, suggesting its importance to PS-catalyzed folding.

PS stabilizes a late-stage transition between compact misfolded and native states

For pepsin [7,15], α LP4 [8], and SGPB [5], the PS catalyzes folding from a compact, well-structured denatured state, indicating that the PS acts at a late stage in the folding process. In the case of pepsin, Rp was characterized by a ΔG_{unf} of 5.8 kcal/mol, a 10% increase in unordered secondary structure and identical tertiary structure to Np (both yield $R_g \sim 20$ Å) [15]. Similarly, the stable denatured state of α LP was found to have secondary and tertiary structures intermediate between the native and unfolded forms, with a 9% increase in unordered secondary structure and a 40% increased hydrodynamic radius [4]. The intermediate α LP gave a ΔG_{unf} of 1 kcal/mol, which was 4 kcal/mol more stable than the native state [4]. The SGPB intermediate was characterized by a ΔG_{unf} of 0.5 kcal/mol but was less stable than the native state by 0.8 kcal/mol [5].

Given the unique properties of Rp, a thermodynamically stable [7], rigid [16], native-like yet inactive form, it is reasonable to suppose that Rp is a late-stage misfolded state that lacks the correct domain-domain interactions found in Np. The active site cleft of Np is formed between the N- and C-terminal lobes (**Fig 2A**), and it is possible that the PS catalyzes the correct formation of inter-domain contacts. This scenario is consistent with the relatively small changes in secondary and tertiary structure that accompany the Rp to Np transition [15]. This picture is also consistent with the concept of independent foldon units (in this case the N- and C-terminal domains), which fold independently, followed by a rate-limiting docking step [52]. In the case of pepsin, and perhaps other zymogen-derived proteins, the rate-limiting docking of domains is slow enough that a PS is required to act as a foldase.

With the discovery of PS-assisted folding in a few evolutionarily unrelated serine peptidases, it was suggested that PS-catalyzed folding has developed through convergent evolution [11,13], a notion supported by the similarities that exist in the folding mechanisms of pepsin and α LP. Interestingly, the PS of pepsinogen is 44 residues in length whereas the PS domains for α LP and SGPB are much longer, at 166 and 76 residues, respectively. The PS-catalyzed folding of pepsin yields a folding rate enhancement ($k_{\text{cat}}/k_{\text{non-cat}}$) far greater than that of SGPB, yet is less than that of α LP (**Fig 1AB**). Thus, PS length is not

necessarily correlated with the power of the PS as a folding catalyst. This comparison also highlights that the pepsin PS is a highly efficient folding catalyst: at a mere 44 residues in length it provides substantial TS stabilization per residue (**Fig 3A**), likely via a high density of strengthened native and/or non-native contacts.

Mutations to the PS of α LP were also found to have a greater impact on the PS-catalyzed folding rate than on binding the native or intermediate states, determined by measuring K_i and K_M values, respectively [53]. PS mutations Y26F, E30A, and Y26F/E30A resulted in essentially no change in K_i or K_M while k_{cat} was reduced by a factor of 10, 2, and 60, respectively [53]. These findings reflect the fact that the PS binds most tightly to the folding TS for α LP, and such interactions may involve non-native contacts, similarly as for pepsin.

The PS may 'buffer' the folding landscape via non-native interactions

The physiological basis for the existence of PS-catalyzed folding as a folding mechanism is not clear. For α LP it was shown that the PS aids in the formation of a kinetically trapped native state that is highly rigid, thus conferring enhanced resistance to proteolysis for this extracellular serine peptidase [5,54]. In contrast, pepsin was shown to have a relatively flexible native conformation [16], and it is reasonable to conclude that pepsin has evolved a different mechanism of resistance to proteolysis: Np is kinetically stable at acidic pH where most exogenous proteases would be inactivated, allowing for digestion by pepsin. Pepsin and α LP are both kinetically trapped and are thermodynamically metastable or unstable, respectively, yet these features may not be related, given that many thermodynamically stable proteins are at least as, if not more, kinetically stable than pepsin or α LP [5,55,56]. PS-catalyzed folding is not required to generate a kinetically stable fold.

Given the above considerations, there must be a more universal role for PS-catalyzed folding. We hypothesize that PS-catalyzed folding allows for more destabilizing contacts to accumulate in the native fold, thereby allowing for a greater search of evolutionary space that would otherwise be restricted by the loss of stability (*e.g.*, this could result in novel functions/substrate specificity). This is akin to the 'buffering capacity' that the chaperonins GroEL/ES were shown to provide in enhancing protein evolvability [57,58]. Indeed, α LP [8] and pepsin [7] are thermodynamically unstable/metastable native states that would not exist without PS-catalyzed folding. In such a scenario, the PS acts to buff or smooth [59] the folding landscape via non-native interactions, catalyzing the folding to a thermodynamically stable PS-native state complex. Upon removal of the PS, these stabilizing contacts are no longer available in the unfolding TS; thus, the unfolding barrier is increased yielding a kinetically trapped native state.

Materials and Methods

Materials

Synthetic peptides were obtained from CanPeptide Inc. (Pointe-Claire, QC, Canada), and were more than 95% pure as judged by LC-MS. Peptides corresponding to the 44-residue PS domain of pepsinogen were obtained in wild-type and the following single mutant forms, in which the wild-type residue was replaced with alanine: V4A, L6A, S11A, R13A, I17A, F25A, and K36A. Porcine pepsin A (EC 3.4.23.1) was purchased from Sigma (St. Louis, MO, USA) and used without further purification. Protein solutions were prepared by mass (wt/vol) and the concentrations were determined from the absorbance at 280 nm, using extinction coefficients of $1490 \text{ M}^{-1}\text{cm}^{-1}$ for the PS peptides and $52,830 \text{ M}^{-1}\text{cm}^{-1}$ for pepsin, estimated using the ProtParam tool

[60]. Rp samples were prepared by first denaturing pepsin by making a 20 mg/ml solution in 30 mM NaOH, with a final pH of 8, yielding alkaline denatured pepsin. Rp was then obtained by diluting an aliquot of the alkaline denatured protein to 0.35 mg/ml in 20 mM acetic acid/NaOH buffer at pH 5.3.

Sequence alignment

Sequences similar to that of porcine pepsinogen were identified using an NCBI-BLASTp search [61], and the results were limited to the 100 top scoring, non-redundant sequences. Multiple sequence alignment of this group was then performed using CINEMA 5 [62], in which a breakpoint was added to isolate the alignment of the PS domain from that of the mature domain.

PS-catalyzed folding of Rp to Np

PS-catalyzed folding was carried out by combining 1 μM Rp and 30 μM PS, in a volume of 100 μl , at pH 5.3 (20 mM acetic acid/NaOH with 100 mM NaCl) and 15°C. Aliquots were taken at several time intervals, diluted 20-fold in 50 mM phosphoric acid buffer, pH 1.2, incubated for 5 min at 25°C and assayed for Np activity using the KPAEFF(NO₂)AL substrate. The recovery of Np activity with time, t , was fit with a monoexponential function ($y = a - b \exp(-k_f t)$) to obtain the PS-catalyzed folding rate constant, k_f . Under these conditions, PS binding to Rp reached equilibrium within the dead-time of mixing (<8 sec), as judged by Trp-fluorescence.

PS binding to Rp

The change in intrinsic tryptophan fluorescence of pepsin was used to measure the binding of PS to Rp and to determine the dissociation constant, K_d . Rp solutions were diluted to between 0.6 and 1.2 μM in 20 mM acetic acid/NaOH buffer, pH 5.3, and mixed with various amounts of PS. After incubating the samples at 20°C for ≥ 10 minutes, the intrinsic tryptophan fluorescence was measured using a PTI spectrofluorophotometer (Photon Technology International, Inc., Birmingham NJ, USA), with excitation at 295 nm and emission measured at 315 nm. The change in fluorescence, ΔF_i , at each PS concentration, $[PS]$, was normalized relative to the maximum change, ΔF_{max} , and fit according to

$$\Delta F_i / \Delta F_{max} = \frac{\left(([Rp] + [PS] + K_d) - \left(\sqrt{([Rp] + [PS] + K_d)^2 - 4[Rp][PS]} \right)^{1/2} \right)}{2} / [Rp] \quad (1)$$

where $[Rp]$ is the total concentration of pepsin.

PS binding to Rp was measured in buffer without added 100 mM NaCl in order to obtain more accurate measurements of K_d , owing to a larger change in signal in the absence of salt. Additionally, without added 100 mM NaCl it was possible to isolate the PS-Rp binding step from the catalyzed folding step thereby improving the determination of K_d . Without 100 mM NaCl, no generation of protease activity was observed on a timescale of 0–2 hours (data not shown, and [7]), indicating that the PS binds Rp yet does not catalyze folding to Np. However, ¹H-¹⁵N TROSY NMR experiments [51] indicated that PS-Rp does fold to the native complex in the absence of 100 mM NaCl, but over longer timescales of days (**Fig S7**).

PS binding to Np

PS binding to Np was determined by using the PS as a competitive inhibitor and measuring the inhibition constant, K_i . Hydrolysis of the KPAEFF(NO₂)AL substrate was measured by

following the decrease in absorbance at 300 nm using a Biochrom Ultraspec 3100pro UV-Vis spectrophotometer (Biochrom Ltd., Cambridge, England) in 20 mM acetic acid/NaOH buffer, pH 5.3, containing 100 mM NaCl. Np samples were diluted to 10 nM and incubated with PS for 5 min at 20°C and assayed for activity. The reaction rates were normalized to the activity in the absence of PS and the data fit using the competitive inhibitor form of the Michaelis-Menten equation

$$v_0 = \frac{V_{\max} \times [S]}{[S] + K_M(1 + [I]/K_i)} \quad (2)$$

where v_0 is the initial reaction rate, V_{\max} is the maximum reaction rate, $[S]$ is the substrate concentration (fixed at 0.1 mM), $[I]$ is the inhibitor concentration, and K_M is the Michaelis constant.

CD spectroscopy

CD data were collected using a Jasco J-810 spectropolarimeter (Jasco corp., Tokyo, Japan), over a wavelength range of 250 nm to 190 nm with a 1 nm resolution, 100 nm/min scan rate, 0.25 s response time and four-fold accumulation of scans. PS_{wt} was diluted to 0.1 mg/ml in either pure water or 20 mM acetic acid/NaOH buffer at pH 5.3, with and without 100 mM NaCl, and loaded into a cell with a 0.1 cm path length. Background spectra were subtracted and the sample spectra converted to units of mean residue ellipticity, MRE , using $MRE = MRW \times \theta_\lambda / (10 \times d \times c)$, where MRW is the mean residue weight (molecular weight/number of residues), θ_λ is the measured ellipticity at a particular wavelength (degrees), d is the pathlength (0.1 cm) and c is the protein concentration (g/cm³).

Calculation of Φ -values from PS-catalyzed folding and binding constants

The change in stability of the folding transition state upon mutation was calculated using

$$\Delta\Delta G^\ddagger = -RT \ln \left(\frac{k_f^{mut}}{k_f^{wt}} \right) \quad (3)$$

where $k_{f, wt}$ and $k_{f, mut}$ are the PS-catalyzed folding rate constants of the wild-type and mutant PS peptides. The change in equilibrium stability of PS-Np relative to PS-Rp upon mutation of the PS ($\Delta\Delta G_{PS(Np-Rp)}$) was determined as the difference between the changes in binding energies, using

$$\Delta\Delta G_{bind} = -RT \ln \left(\frac{K_d^{wt}}{K_d^{mut}} \right) \quad (4)$$

and

$$\Delta\Delta G_{PS(Np-Rp)} = \Delta\Delta G_{PS-Np} - \Delta\Delta G_{PS-Rp} \quad (5)$$

Here, $\Delta\Delta G_{bind}$ refers to either $\Delta\Delta G_{PS-Rp}$, determined from K_d values, or $\Delta\Delta G_{PS-Np}$, determined from K_i values. The Φ -value for each mutant corresponds to eq 3 divided by eq 5. Additional details are included in the supporting information section (**Text S1**).

Supporting Information

Figure S1 PS-catalyzed folding approach to Φ -value analysis. The effect of a mutation on each step of the folding

landscape was determined separately by measuring PS-catalyzed folding and binding affinities rather than directly measuring the equilibrium stability of PS-Np relative to PS-Rp, $\Delta\Delta G_{PS(Np-Rp)}$. The relative changes in binding affinities gave $\Delta\Delta G_{PS(Np-Rp)}$, while $\Delta\Delta G^\ddagger$ was obtained directly from the relative folding rates. (TIF)

Figure S2 Structure of the PS domains of pepsinogen (PDB: 3PSG) and progastricsin (PDB: 1HTR). Ribbon diagram showing select residues of pepsinogen (red side chains, black backbone) and progastricsin (blue side chains, grey backbone), starting from the N-terminus, pepsinogen numbering: V4, L6, R8, S11, R13, I17, F25 and K36. The overall fold is very similar with an average RMSD of 1.24 Å, while particularly for the conserved residues L6 and R13 and the semi-conserved V4, S11 and K36, the structures are identical, with RMSD < 1 Å. (TIF)

Figure S3 Sequence alignment of the PS domain of pepsinogen with the nearest 100 sequences. The first sequence from the top is porcine pepsinogen (PEPA_PIG). (TIF)

Figure S4 Determination of k_f , K_d and K_i for PS-catalyzed folding, PS-Rp binding and PS-Np binding, respectively. Data and fit curves are the same as those shown in Fig 2BCD in the main text, and are plotted for each individual wt- and mutant PS for clarity. (TIF)

Figure S5 Temperature dependence of PS_{wt}-catalyzed folding of Rp to Np. Rates were measured at 0, 5, 10 and 15°C, using recovery of Np activity. Data points are the mean \pm SD of at least three determinations. The linear fit is also indicated. (TIF)

Figure S6 CD spectra of prosegment. Far-UV CD spectra of PS_{wt} in ddH₂O and in 20 mM acetic acid/NaOH, at pH 5.3, with and without 100 mM NaCl added. A negative band at 198 nm is characteristic of random coil structure [Sreerama N, Venyaminov SY, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. Anal Biochem 287: 243–251]. (TIF)

Figure S7 PS-Rp folds to an identical native conformation with and without added salt. ¹H-¹⁵N TROSY NMR spectra were collected for samples of PS combined with Rp in buffers (A) containing 100 mM NaCl and (B) without added 100 mM NaCl. (C) Overlay of the two spectra. The buffer was 20 mM NaOAc pH 5.3 with 10% D₂O, at 22°C. The NMR experimental details were published previously [51]. (TIF)

Text S1 Calculation of Φ -values from PS-catalyzed folding and binding constants. (DOCX)

Text S2 Temperature dependence of PS-catalyzed folding. (DOCX)

Author Contributions

Conceived and designed the experiments: DRD RYY. Performed the experiments: DRD YH. Analyzed the data: DRD YH. Wrote the paper: DRD RYY.

References

- Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181: 223–230.
- Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277: 985–994.
- Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, et al. (2003) Contact order revisited: influence of protein size on the folding rate. *Protein Sci* 12: 2057–2062.
- Baker D, Sohl JL, Agard DA (1992) A protein-folding reaction under kinetic control. *Nature* 356: 263–265.
- Truhlar SME, Cunningham EL, Agard DA (2004) The folding landscape of *Streptomyces griseus* protease B reveals the energetic costs and benefits associated with evolving kinetic stability. *Protein Sci* 13: 381–390.
- Fisher KE, Ruan B, Alexander PA, Wang L, Bryan PN (2007) Mechanism of the kinetically-controlled folding reaction of subtilisin. *Biochemistry* 46: 640–651.
- Dec DR, Yada RY (2010) The prosegment catalyzes pepsin folding to a kinetically trapped native state. *Biochemistry* 49: 365–371.
- Sohl JL, Jaswal SS, Agard DA (1998) Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* 395: 817–819.
- Jaswal SS, Truhlar SM, Dill KA, Agard DA (2005) Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases. *J Mol Biol* 347: 355–66.
- Sanchez-Ruiz JM (2010) Protein kinetic stability. *Biophys Chem* 148, 1–15.
- Baker D, Shiau AK, Agard DA (1993) The role of pro regions in protein folding. *Curr Opin Cell Biol* 5: 966–970.
- Bryan PN (2002) Prodomains and protein folding catalysis. *Chem Rev* 102: 4805–4816.
- Eder J, Fersht AR (1995) Pro-sequence-assisted protein folding. *Mol Microbiol* 16: 609–614.
- Pauwels K, Van Molle I, Tommassen J, Van Gelder P (2007) Chaperoning Anfinsen: the steric foldases. *Mol Microbiol* 64: 917–922.
- Dec D, Pencer J, Nieh M-P, Krueger S, Katsaras J, et al. (2006) Comparison of solution structures and stabilities of native, partially unfolded and partially refolded pepsin. *Biochemistry* 45: 13982–13992.
- Dec DR, Myers B, Yada RY (2011) Dynamics of thermodynamically stable, kinetically trapped, and inhibitor-bound states of pepsin. *Biophys J* 101: 1699–1709.
- Borgia MB, Borgia A, Best RB, Steward A, Nettels D, et al. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474: 662–665.
- Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J (2007) The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* 8: 319–330.
- Itoh K, Sasai M (2008) Cooperativity, connectivity, and folding pathways of multidomain proteins. *Proc Natl Acad Sci USA* 105: 13865–13870.
- Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme I. theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224: 771–782.
- Gromiha MM, Thangakani AM, Selvaraj S (2006) FOLD-RATE: prediction of protein folding rates from amino acid sequence. *Nucleic Acids Res* 34: W70–74.
- Wishart DS, Arndt D, Berjanskii M, Guo GC, Shi Y, et al. (2008) PPT-DB: the protein property prediction and testing database. *Nucleic Acids Res* 36: D222–D229.
- Ozkan SB, Bahar I, Dill KA (2001) Transition states and the meaning of phi-values in protein folding kinetics. *Nat Struct Biol* 8: 765–769.
- Naganathan AN, Muñoz V (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proc Natl Acad Sci USA* 107: 8611–8616.
- Raleigh DP, Plaxco KW (2005) The protein folding transition state: what are phi-values really telling us? *Protein Pept Lett* 12: 117–122.
- Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254: 260–288.
- Serrano L, Matouschek A, Fersht AR (1992) The folding of an enzyme. III. structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 224: 805–818.
- Grantcharova VP, Riddle DS, Santiago JV, Baker D (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* 5: 714–720.
- Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, et al. (1999) Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 6: 1016–1024.
- Northey JG, Di Nardo AA, Davidson AR (2002) Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol* 9: 126–130.
- Zarrine-Afsar A, Dahesh S, Davidson AR (2007) Protein folding kinetics provides a context-independent assessment of beta-strand propensity in the fyn SH3 domain. *J Mol Biol* 373: 764–774.
- Zarrine-Afsar A, Davidson AR (2004) The analysis of protein folding kinetic data produced in protein engineering experiments. *Methods* 34: 41–50.
- Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14: 70–75.
- Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38: D211–222.
- Sánchez IE, Kiefhaber T (2003) Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J Mol Biol* 334: 1077–1085.
- Fersht AR, Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci USA* 101: 7976–7981.
- Oliveberg M, Wolynes PG (2005) The experimental survey of protein-folding energy landscapes. *Quart Rev Biophys* 38: 245–288.
- Ozkan SB, Bahar I, Dill KA (2001) Transition states and the meaning of Phi-values in protein folding kinetics. *Nat Struct Biol* 8: 765–769.
- Li L, Mirny LA, Shakhnovich EI (2000) Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nat Struct Biol* 7: 336–342.
- Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, et al. (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc Natl Acad Sci USA* 105: 9999–10004.
- Plotkin SS (2001) Speeding protein folding beyond the Gō model: how a little frustration sometimes helps. *Proteins* 45: 337–345.
- Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci* 13: 1750–1766.
- Di Nardo AA, Korzhnev DM, Stogios PJ, Zarrine-Afsar A, Kay LE, et al. (2004) Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation. *Proc Natl Acad Sci USA* 101: 7954–7959.
- Morton VL, Friel CT, Allen LR, Paci E, Radford SE (2007) The effect of increasing the stability of non-native interactions on the folding landscape of the bacterial immunity protein Im9. *J Mol Biol* 371: 554–568.
- Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L, et al. (2002) Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat Struct Biol* 9: 485–493.
- Kelch BA, Salimi NL, Agard DA (2012) Functional modulation of a protein folding landscape via side-chain distortion. *Proc Natl Acad Sci USA* 109: 9414–9419.
- Hartsuck JA, Koelsch G, Remington SJ (1992) The high-resolution crystal structure of porcine pepsinogen. *Proteins* 13: 1–25.
- Zarrine-Afsar A, Zhang Z, Schweiker KL, Makhatadze GI, Davidson AR, et al. (2012) Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain. *Proteins* 80: 858–870.
- Chan HS, Zhang Z, Wallin S, Liu Z (2011) Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu Rev Phys Chem* 62: 301–326.
- Bölinger D, Sulkowska JI, Hsu H-P, Mirny LA, Kardar M, et al. (2010) A Stevedore's protein knot. *PLoS Comput Biol* 6: e1000731.
- Wang S, Horimoto Y, Dec DR, Yada RY (2014) Understanding the mechanism of prosegment-catalyzed folding by solution NMR spectroscopy. *J Biol Chem* 289: 697–707.
- Fersht AR (1997) Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* 7: 3–9.
- Cunningham EL, Mau T, Truhlar SME, Agard DA (2002) The pro region N-terminal domain provides specific interactions required for catalysis of alpha-lytic protease folding. *Biochemistry* 41: 8860–8867.
- Jaswal SS, Sohl JL, Davis JH, Agard DA (2002) Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* 415: 343–346.
- Xia K, Manning M, Hesham H, Lin Q, Bystroff C, et al. (2007) Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proc Natl Acad Sci USA* 104: 17329–17334.
- Manning M, Colón W (2004) Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry* 43: 11248–11254.
- Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 19: 596–604.
- Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459: 668–673.
- Plotkin SS, Wolynes PG (2003) Buffered energy landscapes: another solution to the kinetic paradoxes of protein folding. *Proc Natl Acad Sci USA* 100: 4417–4422.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. *The Proteomics Protocols Handbook*. Totowa: Humana Press. pp. 571–607.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
- Parry-Smith DJ, Payne AWR, Michie AD, Attwood TK (1997) CINEMA - a novel colour interactive editor for multiple alignments. *Gene* 211: GC45–56.