Understanding the Mechanism of Prosegment-catalyzed Folding by Solution NMR Spectroscopy*

Received for publication, August 10, 2013, and in revised form, November 13, 2013. Published, JBC Papers in Press, November 21, 2013, DOI 10.1074/jbc.M113.505891

Shenlin Wang (王申林)1,2, Yasumi Horimoto3, Derek R. Dee4, and Rickey Y. Yada5,6,†

From the Departments of 1Physics and 2Food Science, and the 3Biophysics Interdepartmental Group, University of Guelph, Guelph, Ontario N1G 2W1, Canada and the 4Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

Background: Native pepsin can be irreversibly denatured into a thermodynamically stable but non-active form. Results: Prosegment converts the misfolded pepsin to a structure similar to the precursor form of pepsin.

Conclusion: Hydrophobic interactions between prosegment and pepsin lower the barrier of pepsin folding toward the native complex. Significance: The mechanism for how the prosegment mediates formation of native pepsin is proposed.

Multidomain protein folding is often more complex than a two-state process, which leads to the spontaneous folding of the native state. Pepsin, a zymogen-derived enzyme, without its prosegment (PS), is irreversibly denatured and folds to a thermodynamically stable, non-native conformation, termed refolded pepsin, which is separated from native pepsin by a large thermodynamically stable, non-native conformation, termed refolded pepsin, which is separated from native pepsin by a large activation barrier. While it is known that PS binds refolded pepsin and catalyzes its conversion to the native form, little structural details are known regarding this conversion. In this study, solution NMR was used to elucidate the PS-catalyzed folding mechanism by examining the key equilibrium states, e.g. native and refolded pepsin, both in the free and PS-bound states, and pepsinogen, the zymogen form of pepsin. Refolded pepsin was found to be partially structured and lacked the correct domain-domain structure and active-site cleft formed in the native state. Analysis of chemical shift data revealed that upon PS binding refolded pepsin folds into a state more similar to that of pepsinogen than to native pepsin. Comparison of pepsin folding by wild-type and mutant PSs, including a double mutant PS, indicated that hydrophobic interactions between residues of prosegment and refolded pepsin lower the folding activation barrier. A mechanism is proposed for the binding of PS to refolded pepsin and how the formation of the native structure is mediated.

Kinetic barriers are important controls for protein folding and stability (1), although the detailed nature of these energy barriers is not well understood (2–6). While proteins folded to the native state are generally thermodynamically stabilized (7), a number of zymogen-derived proteases are kinetically trapped and thermodynamically unstable/metastable. Examples include the aspartic protease (AP)4 pepsin (8) and the serine proteases α-lytic peptidase (αLP) (9) and subtilisin (10). The zymogen precursors of these proteins contain a prosegment (PS) that is removed upon activation of the mature enzyme. PSs serve a number of functions, including intracellular sorting, providing stability, and enzyme inhibition (11). In particular, the PS plays a distinct role in stabilizing the folding transition state and thus catalyzing the folding of numerous proteases (12–14).

Within the AP family, PS-catalyzed folding has been verified only for porcine pepsin (15, 16). This study examines the PS-catalyzed folding of pepsin, which serves as a model for understanding the nature of folding/unfolding barriers and the folding mechanism of related APs and proteases from other families. Using what is known about structure-function relationships could provide further critical information in the development of APs as targets in pharmaceutical intervention in human diseases such as hypertension (renin), breast cancer (cathepsin D), Alzheimer disease (beta-secretase), AIDS (HIV protease), and malaria (plasmspens) (17, 18).

Pepsin is initially formed as an inactive zymogen, pepsinogen, which contains an additional 44 residues of PS on the N terminus (Fig. 1A). The PS is removed autocatalytically upon acidification (pH < 5) of pepsinogen, yielding active native pepsin. X-ray structures of native pepsin (PDB code: 4PEP) (Fig. 1B) and pepsinogen (PDB code: 3PSG) (Fig. 1C) were previously reported (19, 20). They share similar overall secondary structures and topology, consisting predominantly of β-strands, forming two compact and topologically similar N- and C-terminal domains, which are connected via a rigid six-stranded β-sheet plate (19). In pepsinogen, the PS interacts with both N- and C-terminal domains and occupies the substrate binding cleft, competitively blocking access to the two catalytic aspartic residues (D32 and D215) (Fig. 1B). The PS consists of three α-helices and a β-strand (Fig. 1D) that forms part of a six-stranded β-sheet plate (Fig. 1C). Upon removal of the PS, the β-strand formed by PS residues L1p-K9p (where p denotes a prosegment residue) is replaced in an identical conformation by residues

* This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs (CRC) Program.
† To whom correspondence should be addressed: Department of Food Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Tel.: 519-824-4120 (ext. 58915); Fax: 519-824-6631; E-mail: ryada@uoguelph.ca.

1 Both authors contributed equally to this work.
2 Supported by the Canadian Institutes of Health Research (CIHR) Fellowship.
3 To whom correspondence should be addressed: Department of Food Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Tel.: 519-824-4120 (ext. 58915); Fax: 519-824-6631; E-mail: ryada@uoguelph.ca.
4 The abbreviations used are: AP, aspartic protease; PS, prosegment; αLP, α-lytic peptidase; PS-Rp, PS-refolded pepsin complex.
Mechanism of Prosegment-catalyzed Folding

FIGURE 1. X-ray structures of pepsin and pepsinogen. A, schematic of the domain structure. Pepsinogen, an inactive zymogen, contains an N-terminal 44 residue prosegment (PS, colored blue) domain. Native pepsin has a bilobal conformation consisting of N- (residues 1–175, colored orange) and C- (residues 176–326, colored turquoise) domains that are connected via a six-stranded β-sheet plate. B, ribbon representation of native pepsin (PDB ID: 4PEP). Two catalytic aspartic residues, D32 and D215 (red spheres), sit on either side of the active site cleft formed between the N- and C-domains. Native pepsin consists of predominantly β-strand (≈50%) and random coil (38%) with limited α-helix (12%) (19). C, ribbon representation of pepsinogen (PDB ID: 3PSG) (20). D, ribbon representation of PS. p denotes the prosegment, and numbering is relative to the N terminus.

I1-T12 of native pepsin, which requires a large conformational change of the mature N terminus moving over 40 Å. Pepsinogen can be reversibly unfolded by means of urea or alkaline pH (21), while native pepsin is irreversibly denatured above pH 6. Although the disulfide bonds remain intact (22), neutral and alkaline-denatured pepsin is partially unfolded, with biochemical data suggesting an unstructured N-terminal domain and a compact C-terminal domain (8, 21), while NMR and solution x-ray scattering (SAXS) data suggest at least partial unfolding of both domains (8, 23, 24). When alkaline-denatured pepsin is re-acidified (pH≈5.3), it forms an inactive, partially refolded state (refolded pepsin) (8), which is a compact misfolded state with secondary structure intermediate between the alkaline-denatured and native forms, yet with surprisingly greater thermal stability than native pepsin. Most notably, a large activation barrier (≈25 kcal/mol) separates refolded pepsin from the native state (15); addition of the PS to refolded pepsin reduces this folding barrier by 6 kcal/mol, enhancing the folding rate by a factor of 105.

Further exploration of the PS foldase mechanism may help to understand the nature of kinetic barriers separating compact misfolded and native states. Critical to this effort is to obtain a detailed understanding of the structure of refolded pepsin and the folding dynamics of the conversion of the PS-refolded pepsin complex (termed PS-Rp hereafter) into the native state. However, insight into PS-Rp has so far been limited due to its transient nature; addition of the PS to refolded pepsin results in the rapid and near complete folding to a native-like state, termed PS-Rp+, from which native pepsin is obtained after dissociation of the PS at pH<3 (15, 16). A recent examination of a number of PS mutants revealed two substitutions, I17pA and F25pA, were among the most destabilizing (each by 1 kcal/mol) to the PS-native pepsin complex while having no substantial effect on the PS-refolded pepsin complex (25). Therefore, we hypothesize that a double mutant, I17pA/F25pA, could shift the equilibrium closer to PS-refolded pepsin, thus enabling further structural studies for understanding the PS-catalyzed folding mechanism.

NMR spectroscopy is widely utilized in studying protein folding (26–32). In the present study, NMR is used to reveal important details regarding the PS-catalyzed pepsin folding mechanism by examining the key equilibrium states, i.e. refolded pepsin and native pepsin, both free and in complex with the PS, and pepsinogen. Upon binding the PS, it was found that refolded pepsin folds to a conformation that more closely resembles pepsinogen than the native pepsin. Analysis of the conformations of refolded pepsin bound with different PS mutants by solution NMR provides insight into the PS-catalyzed folding pathway.

EXPERIMENTAL PROCEDURES

Peptide Preparation—Synthetic peptides were obtained from Canpeptide Inc. (Pointe-Clair, 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 wildtype (PSwt), single mutant forms (PSI17pA and PSF25pA), and double mutant form (PSI17pA/F25pA).

Cloning, Expression, and Purification of Pepsinogen and Pepsin—Cloning, expression, and purification of the various protein constructs were as previously described (33). Briefly, soluble expression of pepsinogen was facilitated via fusion with thioredoxin using the expression vector, pET-32b(+) (Novagen, Mississauga, ON Canada) and included a His-Tag for purification.

Preparation of Refolded Pepsin—Purified recombinant pepsin was denatured with 30 mM NaOH and equilibrated for 15 min, followed by buffer-exchange into 20 mM sodium acetate buffer (pH 5.3) to generate refolded pepsin (16).

Fluorescence Titration of Refolded Pepsin with PSI17pA/F25pA—The change in intrinsic tryptophan fluorescence of pepsin was used to measure the binding of PS to refolded pepsin and to determine the dissociation constant, KD. The measurement was conducted according to a previous method (16). Refolded pepsin solutions were diluted to 1.5 μM in 20 mM sodium acetate buffer, pH 5.3 and mixed with the PS, ranging from 0 to 10 μM. After the samples were incubated at 20 °C for 20 min, they were measured by a Shimadzu RF5301 spectrofluorophotometer (Shimadzu Corp., Kyoto, Japan), with excitation at 295 nm and emission measured at 340 nm. The change in fluorescence, ΔF, at each PS concentration was normalized relative to the maximum change, ΔFmax, and fit according to Equation 1,

$$\frac{\Delta F}{\Delta F_{\text{max}}} = \frac{[P][S]}{[K_D] + [PS]}$$  
(Eq. 1)

where [Rp] and [PS] are the concentrations of refolded pepsin and PS, respectively, and K_D is the dissociation constant. PS-binding measurements were repeated five times.
**NMR Spectroscopy**—All NMR data were collected on a Bruker Avance II spectrometer operating at a proton frequency of 600.130 MHz and equipped with a cryoprobe (NMR center, University of Guelph). All experiments were performed at 22 °C. Samples for NMR experiments contained 0.2 mM protein (BMRB code: 18245 and 18246) were used in the analysis of the native pepsin spectra (33). Secondary structure elements were predicted by chemical shift index (CSI) (37, 38) using Cα and Cβ chemical shifts after corrections for deuteration isotope effects (39).

**NMR Titration of Refolded Pepsin with PS Peptides**—The synthetic peptides corresponding to the wild type PS domain and its mutants (I17pA-PS, F25pA-PS, and I17pA/F25pA-PS) were used to study interactions with refolded pepsin. The lyophilized peptide powders were dissolved in the same buffer as that used for pepsin samples. For titration experiments, 15N-labeled peptide powders were dissolved in the same buffer and native pepsin contained predominantly -helical structure in solution, which matched closely the reported crystal structures. CSI analysis showed that both pepsinogen and native pepsin were in close agreement with the reported chemical shifts of native pepsin (40), indicating that all Cys residues of both pepsinogen and native pepsin were oxidized, as identified by the Cβ chemical shifts (40), indicating that all Cys residues were disulfide bonded.

Further verification of the native pepsin and pepsinogen structures in solution was provided by comparing the N-terminal segment of native pepsin with the corresponding region in pepsinogen, which is expected to switch from -helical to -strand structure and undergo a ~40 Å shift upon removal of the PS from pepsinogen (Fig. 2D). CSI analysis indicated that residues G2-G19 of pepsinogen contained -helical structure (Fig. 2B), while residues G2-Y9 and E13-G19 in native pepsin formed -strand structures (Fig. 2A). Accordingly, substantial variations in amide NH chemical shifts (Δδ(NH)>0.5ppm) between pepsinogen and native pepsin were observed for most of those residues (Fig. 2C). Therefore, our findings are consistent with those made by comparing the x-ray structures of native pepitin and pepsinogen (Fig. 2D), which both indicated that removal of the PS allowed for major conformational changes within the first ~20 residues of mature pepsin.

Analysis of 1H-15N TROSY spectra revealed additional insight into the conformational and dynamic changes that accompany the conversion of pepsinogen into native pepitin. The NH resonances of residues in two continuous regions (I73-L84 and S104-Y114, corresponding to -5, -6, -7, and -3 in the N-domain of native pepitin) were not detected in the 1H-15N TROSY spectra of pepsinogen, indicating that these regions experience multiple conformational states, exchanging with each other on an intermediate time scale (μs-ms), which leads to severe resonance broadening. In contrast, most of the NH signals of these residues were detectable in the native pepsin spectra suggesting an increased local rigidity in the N-domain upon PS removal. As found in the pepsinogen crystal structure, residues I73-L84 and S104-Y114 comprise regions wrapping the helix formed by residues E4-Y9, which shifted about 40Å upon PS removal. Therefore, this result indicated that the additional local backbone motions of the I73-L84 and S104-Y114 in pepsinogen are dependent upon the structural rearrangement of E4-Y9.

The above data, therefore, support the notion that native pepitin is more compact and rigid than pepsinogen, even though both proteins share an overall similar topology as shown in crystal structures. These results also provide the first new insights into restrictions on backbone motions and increase in structural compactness of pepitin accompanying the structural rearrangements caused by PS removal.

**NMR Characterization of Refolded Pepsin and PS-Rp*—**1H-15N TROSY spectra from both pepsinogen (Fig. 3A) and native pepsin (Fig. 3B) showed well dispersed resonances, indicating essentially well-folded proteins. In contrast, the dispersion of signals in 1H-15N TROSY spectra from refolded pepitin was much poorer, indicating that refolded pepitin is less structured, more disordered, and flexible than pepsinogen and native pepitin (Fig. 3C). However, about sixty well-resolved NH signals have amide proton chemical shifts higher than 8.5 ppm or lower.

**Mechanism of Prosegment-catalyzed Folding**

**RESULTS**

**NMR Analysis of Native Pepsin and Pepsinogen and Comparison with Crystal Structures**—The structures of native pepsin and pepsinogen were analyzed using CARA program (34). Three cross peaks appearing at 10.1/128.6 ppm, 8.2/108.9 ppm and 7.3/121.1 ppm corresponding to residues of PS-free refolded pepitin were selected to follow the completeness of the titrations.
than 7.5 ppm (Fig. 3C), suggesting that instead of a fully unstructured protein, refolded pepsin is a partially folded state and may contain regions having secondary structures. Notably, the formation of $\beta$-strands, or a similar H-bonding network involving backbone atoms, was suggested by the presence of a number of cross peaks that have $^1$H and $^{15}$N chemical shifts higher than 8.5 ppm and 125 ppm, respectively, which are indicative of $\beta$-strand structure (40). This observation is consistent with CD and SAXS data, which showed that refolded pepsin is partly folded having secondary and tertiary structures intermediate between native pepsin and the more extensively unfolded alkaline denatured state (8).

To assess the structural rearrangements that occur during PS-catalyzed folding of pepsin, $^{15}$N-labeled refolded pepsin was titrated with an unlabeled PS peptide and the process was fol-
lowed using the $^1$H-$^{15}$N TROSY spectra. A stepwise titration of the PS into $^{15}$N-refolded pepsin led progressively to the appearance of new peaks in the TROSY spectra and, concomitantly, the disappearance of some NH signals corresponding to free refolded pepsin (Fig. 4, A–C).

The newly appeared signals were similarly dispersed as those from native pepsin and pepsinogen, indicating that the PS converted refolded pepsin into a more compact conformation, increasing rigidity to an extent close to pepsinogen (Fig. 4, D and E). The process was completed when the molar ratio between PS and refolded pepsin reached 1:1 (Fig. 4, D–E), indicating that refolded pepsin binds the PS with a stoichiometry of 1:1, and with a dissociation constant ($K_D$) estimated to be $\sim 10^{-10}$ M or lower. This $K_D$ value is consistent with previous measurements performed using tryptophan fluorescence spectroscopy (15).

The PS-Rp* complex was characterized as having a $^1$H-$^{15}$N TROSY pattern more similar to that of pepsinogen (Fig. 4E) than that of native pepsin mixed with equal equivalent PS (Fig. 4D) or of free native pepsin. Additionally, this complex was characterized by backbone chemical shifts similar to those of pepsinogen. Particularly, the Cα chemical shifts obtained for residues G2-G19 were identical to those of pepsinogen supporting that the first 20 residues of pepsin within PS-Rp* adopt an $\alpha$-helical conformation, a feature characteristic of pepsinogen but not of native pepsin. Furthermore, backbone NH resonances of residues I73–G84 and S104–Y114 were not detected in the PS-Rp* spectra, suggesting that PS-Rp* also behaves similarly as pepsinogen with a higher degree of backbone fluctuations within the N-terminal domain. These results suggest that upon binding the PS, refolded pepsin experiences a transition from a partially folded state to a well-folded conformation in a process of coupled binding and folding, and forms a complex adopting a native-like secondary structure similar to pepsinogen.

Double Mutant of PS Traps the Folding Intermediate State (PS-Rp)—To understand the PS-catalyzed folding mechanism, we need first to understand structures of not only the reactants (PS and refolded pepsin) and the product (PS-Rp*), but also of the folding intermediate, PS-Rp, which is a transient complex and remains a partially folded state. A structural analysis of PS-Rp requires decoupling the binding and folding processes such that PS-Rp can be trapped for further study. On the basis of a previous mutational study (25), we hypothesized that a PS double mutation (PS$_{I17pA/F25pA}$) could shift the folding equilibrium from PS-Rp* to favor PS-Rp. This
hypothesis was tested by, first, measuring the dissociation constant of the PS$_{17pA/F25pA}$-Rp complex using intrinsic tryptophan fluorescence (Figs. 5 and 6A), obtaining a $K_d$ value of $3.02 \pm 0.52 \, \mu M$, which is comparable to that obtained using the wild-type PS (PS$_{wt}$) of $1.42 \pm 0.12 \, \mu M$ (16). Subsequently, the effect of PS$_{17pA/F25pA}$ on the structure of refolded pepsin was tested by titrating refolded pepsin with PS$_{17pA/F25pA}$ and following the reaction using $^1$H-$^1$H TROSY spectra, similar to the reaction between PS$_{wt}$ and refolded pepsin described above. The TROSY spectrum of PS$_{17pA/F25pA}$-Rp (Fig. 5D) differed greatly from...
from that of PSwt-Rp* (Fig. 5A), and was similar to that of free refolded pepsin (Fig. 3C). Therefore, it was concluded that the PS double mutation I17pA/F25pA was effective in trapping the PSI17pA/F25pA –Rp complex, which can be further studied as a PSwt-Rp model. Upon combining equimolar amounts of PSI17pA/F25pA and 15N-refolded pepsin several new resonances appeared corresponding to residues G93, D138, Q143, G162, and G288 of PSwt-Rp* (Fig. 6B). Thus, our initial analysis showed that PS binds through both domains of the refolded pepsin.

It was notable that the double mutation to the PS, I17pA/F25pA, was able to shift the equilibrium strongly in favor of PS-Rp and prevent folding to PS-Rp*. As PSI17pA/F25pA bound refolded pepsin with an affinity similar to that of PSwt, the shift in the folding equilibrium to favor PS-Rp was likely due to the destabilization of PS-Rp*. To investigate how the I17pA/F25pA mutations are destabilizing within the PSI17pA/F25pA–Rp complex, the effect of individual PS point mutants (PSI17pA and PSF25pA) on refolded pepsin binding and folding were also examined.

New resonances were detected in the 1H-15N TROSY spectra of 15N-refolded pepsin upon titrating with either unlabeled PSI17pA or PSF25pA (Fig. 5, B and C). The majority of these new signals had amide chemical shifts identical to those in the TROSY spectrum of PSwt-Rp*, demonstrating that both single mutant PSs catalyzed the refolded pepsin folding to a pepsinogen-like state. The processes reached completion when mutant PS and refolded pepsin were combined in an equimolar ratio, indicating that both mutant PSs bound refolded pepsin and formed cooperatively folded complexes with dissociation constants in the μM range, comparable to that obtained for PSwt-Rp*.

Comparison of the 1H-15N TROSY spectra of PSI17pA-Rp* and PSF25pA-Rp* showed that the resonances of a number of residues were broadened or missing, i.e. residues consisting of a loop turn spatially close to I17p (E239-E244), residues located at the PS-Rp* interface (G168, G217-T218, L276, and G302-F305), or residues involved in the first residues (N8, Y14, G19, and G21) (Fig. 6C). NH resonances were also not detected for residues, T28, S42, C50, F56, D60, E65, G102, and A130, which are in proximity to the flexible regions composed by residues I73-G84 and S104-Y114. The inability to detect those residues in the PSI17pA-Rp* spectra showed that substitution of Ile to Ala in the PS induces local structural disorder and/or enhanced backbone dynamics within the PS-Rp* complex and suggested that I17p has a role in increasing the rigidity of those regions during conversion to PS-Rp*.

In addition to the same resonances (and the corresponding residues) that were not observed in the PSI17pA-Rp* spectra, more residues were broadened or missing from the TROSY spectra of PSF25pA-Rp* (Fig. 6D). Prominent among the missing residues were F137 of the N-terminal domain and W190 of the

FIGURE 5. Comparison of completeness of refolded pepsin folding upon binding wild type or mutant PSs. The 1H-15N TROSY spectra of PSwt-Rp* (A), PSI17pA-Rp* (B), PSF25pA-Rp* (C), and PSI17pAF25pA-Rp (D) are shown with circles highlighting examples of resonances characterized by a broader line width in refolded pepsin bound with mutant PSs than when bound with wild type PS.
C-terminal domain, which are located on opposite sides of the PS and form part of the hydrophobic cluster between the two domains. This result indicates that in contrast to PSI17pA, which disrupted only local structure of the PS-Rp* binding interface, F25pA induced a global effect on the structure and/or dynamics of the PS-Rp* complex, weakening the interactions between the N- and the C- domains. Together, the I17pA and F25pA point mutants had an additive effect on destabilizing PS-Rp* such that the PS double mutant shifted the equilibrium from PS-Rp* to favor PS-Rp, effectively trapping the folding intermediate.

**DISCUSSION**

**Coupled Binding and Folding**—Examination of the structural changes that occur along the PS-catalyzed conversion of refolded pepsin indicated that upon PS binding refolded pepsin is folded to a compact native-like state with a topology similar to that of pepsinogen (Fig. 4E). The PS-refolded pepsin complex is also structurally different from the complex formed by PS and native pepsin (Fig. 4D). Since this complex folds to a pepsinogen-like state, it is expected that the unstructured PS peptide, upon binding with refolded pepsin, may organize into a similar topology as the PS domain in pepsinogen, which contains a helical kink and one β-strand and interacts with both N- and C-terminal domains of pepsin. The less structured components of the PS may fold simultaneously to a final well-structured complex. Similar cooperatively folding and binding processes have also been observed for the nuclear coactivator binding domain of CBP/300 upon binding to a flexible domain of ACTR (41). The portrait of the structure of refolded pepsin and the structural changes that occur along the PS-catalyzed conversion of refolded pepsin to the pepsinogen-like state (Figs. 7 and 8) contribute substantially to our understanding of the PS-catalyzed folding mechanism.

**Molecular Interactions in PS-catalyzed Folding**—In an earlier study, we showed that the irreversibility of native pepsin unfolding is due mainly to the large energy barrier separating native pepsin from refolded pepsin and the unfolded states (16). In the present study, details about the structure appeared upon addition of PS to refolded pepsin provide insight into the structural origins of the kinetic barrier, which separates the partially folded and native states. Many residues involved in hydrophobic interactions appeared upon addition of PS to refolded pepsin. Hydrophobic collapse plays a major role in driving protein folding (42, 43) and likely plays a central role in the conversion of refolded pepsin to native pepsin. A major topological feature of native pepsin and pepsinogen is the large central hydrophobic core (44), which includes a six-stranded β-sheet plate con-
necting two domains and the region located just behind the active site. This region appeared within PS-Rp*, indicating that formation of native hydrophobic interactions is a major feature of the conversion from refolded to native pepsin.

The importance of hydrophobic interactions to PS-catalyzed folding was also examined by using PS mutants in which hydrophobic side chains were truncated. Previous experiments showed that the I17pA and F25pA substitutions were destabilizing to the PS-native pepsin complex (25). Following this observation, we investigated the folding pathway at residue-level resolution using NMR analysis and found that I17pA induced a number of effects on the PS-Rp* complex. In the x-ray structure of pepsinogen, residue I17p points toward the C-terminal domain (Fig. 7, A and B), especially a loop region (residues E239-M244), and these residues were not demonstrated in the interaction between refolded pepsin and PS_{I17pA*}. Furthermore, residues from the longest helix region in the C-terminal domain (residues G223, I231, I235, S238, and E239) and the following loop region (residues S241, G243, E244, S248, and I252) were missing (Fig. 7C). Residue F25p points toward the N-terminal domain (Fig. 7, A and B) and is in close contact with F111, which may play a functional role in forming hydrophobic interactions (Fig. 7D).

If the helix 2 region of the PS is not anchored early in folding, the binding interface between the N terminus of the PS (1p-16p) and pepsin is likely destabilized. Our NMR data showed that PS_{F25pA-Rp*} is less structured than both the PS_{I17pA-Rp*} and the PS_{wt-Rp*} and, induces a global effect on both the N- and C-domains of pepsin in PS-Rp* complex, not limited to residues around the hydrophobic patch involving F25, as shown in the

FIGURE 7. Hydrophobic interactions that may play a critical role in determining the completeness of PS-catalyzed folding. A, x-ray crystal structure of pepsinogen, with PS (in blue) and residues undetected by NMR (in red). B, top view, and close-up views of (C) residues around I17p, (D) residues around F25p, and (E) the potential interaction between F25p and refolded pepsin (blue), as suggested by examining the x-ray structure of pepsinogen. Undetected residues are shown in red. Undetected glycine residues are shown in yellow.
Mechanism of Prosegment-catalyzed Folding

psigen structure. Thus, this suggested that disruption of helix 2 of the PS, where F25p is located, may be another reason that F25A induces the larger destabilization effects on PS-Rp* complex and, accounts for the reduced folding rate (25).

Except for a few newly detected resonances, no substantial spectra difference was observed between refolded pepes and PS117pA/F25pA-Rp*, suggesting that the double mutant has an additive effect of PS117pA and PSF25pA. Residues I17pA in α-helix 1 and F25pA in α-helix 2 are involved in interactions with the C- and N-terminal domain, respectively. When these two regions are disturbed, PS-Rp cannot be refolded into PS-Rp*, because the burial of a large hydrophobic surface is prevented. Such a scenario is consistent with the reduced change in heat capacity (ΔCv) upon unfolding of refolded pepes (ΔCv = 0.6 kcal/mol/°C) compared with native pepes (ΔCv = 5.2 kcal/mol/°C), determined previously (16), as reduced ΔCv values are associated with increased exposure of hydrophobic groups (45, 46).

Conformational flexibility may also be an important factor in PS-catalyzed folding. Most regions involving glycine, except G76, G78, G82, and G109, were detectable in NMR spectra upon addition of PS to refolded pepes (Fig. 7E). The glycine content of porcine pepes is 10.7%, with 20 Gly in the N-domain and 15 in the C-domain. Fuhrmann et al. suggested that a possible source of large entropy difference in αLp may arise from the high glycine content of the protein (16%) (47). Since glycine lacks a side chain and increases the number of conformations accessible to an unfolded state, its conformational freedom increases the number of unfolded conformations available in refolded pepes. Refolded pepes, therefore, increases entropy and stability, and decreases entropy from the interactions between refolded pepes and PS. Comparison of the number of undetected residues in the NMR spectra of pepes and native pepes showed that the N-terminal domain of pepesogen is more flexible than the C-terminal domain, and that pepesogen is more flexible than native pepes. Our study suggests that the flexible regions of the N-terminal domain may contribute to PS recognition.

Implications for the PS-catalyzed Folding Mechanism—Based on the NMR analysis of the interactions between wildtype and mutant PS peptides with refolded pepes, the mechanism of PS catalyzed pepes folding can be proposed. PSwt cooperatively binds and refolds the partially folded pepes into a well-folded native-like structure, while PS117pA/F25pA binds refolded pepes with a similar affinity yet forms a largely disordered partially folded complex. Single mutant PSs, PSF25pA, and PSI17pA, are also able to bind and refold refolded pepes, but the resulting complexes have structural and dynamic properties intermediate between those of PSwt-Rp* and PSI17pA/F25pA-Rp, which indicated that hydrophobic residues, at least I17p and F25p, have critical roles in refolded pepes refolding but not in refolded pepes binding. Examination of the refolded pepes spectra changes upon addition of PS117pA/F25pA showed that the newly detected resonances corresponded to either Gly or polar residues from both the N- and C-domains, suggesting that the binding between refolded pepes and PS may be driven by electrostatic interactions and involves residues on both domains of refolded pepes. When mapped onto the pepesogen crystal structure, residues G93 and G162 are located spatially close to the β-strands in the N-terminal of PS (L1p-V7p) and G288 is close to the last α-helix in the C-terminal of PS (P33p-K36p), which suggested that the binding may involve residues in multiple regions along the whole peptide sequence of the PS, albeit the complex being flexible. The PS117pA/F25pA-Rp complex characterized in the present study could serve as a model of a partially folded transient complex formed from two less structured proteins, mainly through electrostatic interactions, in the early stage of refolded pepes folding. Subsequently, the hydrophobic interactions between PS and refolded pepes shift the equilibrium toward the well-folded target conformation. This binding before folding process matches with the induced folding mechanism which occurs commonly in the case of coupled binding and folding of intrinsically disordered proteins when encountered with their targeted partners (48).

Folding and unfolding are crucial steps in regulating biological activity and trafficking target proteins to cellular locations. The destabilizing effects of hydrophobic core substitutions arise from both loss of hydrophobicity and from disruption of the tightly packed arrangement of side chains within the core.

NMR Assignment Data—The NMR assignment data are deposited in the BioMagResBank (BMRB), under accession number 18247 for pepesogen.

Acknowledgments—We thank Dr. Margaret Hundleby for technical writing suggestions and assistance. We also thank Prof. Leonid S. Brown for critical reading and suggestions with the manuscript.

REFERENCES
1. Sanchez-Ruiz, J. M. (2010) Protein kinetic stability. Biophys. Chem. 148, 1–15
2. Eder, J., and Fersht, A. R. (1995) Pro-sequence-assisted protein folding. Mol. Microbiol. 16, 609–614
Mechanism of Prosegment-catalyzed Folding

- Value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15717–15722

- Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447, 1021–1025

- Banci, L., Bertini, I., Cefaro, C., Gioffi-Baffoni, S., Gallo, A., Martinelli, M., Sideris, D. P., Katrakili, N., and Tokatidis, K. (2009) Mia40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat. Struct. Mol. Biol.* 16, 198–206

- Banci, L., Bertini, I., Cefaro, C., Cenacchi, L., Gioffi-Baffoni, S., Felli, I. C., Gallo, A., Gonnelli, L., Luchinat, E., Sideris, D., and Tokatidis, K. (2010) Molecular chaperone function of Mia40 triggers consecutive induced folding steps of the substrate in mitochondrial protein import. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20190–20195

- Meinhold, D. W., and Wright, P. E. (2011) Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9078–9083

- Horimoto, Y., Wang, S., and Yada, R. Y. (2012) 1H, 13C, and 15N backbone resonance assignments of the porcine pepsin and porcine pepsin complex with pepstatin. *Biol. NMR Assign.* 10.1007/s12104-012-9452-4

- Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*. Cambridge University Press

- Salzmann, M., Wider, G., Pervushin, K., Senn, H., and Wuthrich, K. (1999) TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J. Am. Chem. Soc.* 121, 844–848

- Wishart, D. S., and Sykes, B. D. (1994) The C-13 chemical-shift Index - a simple method for the identification of protein secondary structure using C-13 chemical-shift data. *Biol. NMR*. 4, 171–180

- Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J. Am. Chem. Soc.* 123, 2970–2978

- Gardner, K. H., Rosen, M. K., and Kay, L. E. (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* 36, 1389–1401

- Wang, Y., and Jarettzky, O. (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* 11, 852–861

- Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415, 549–553

- Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry* 29, 7133–7155

- Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) The protein folding problem. *Annu. Rev. Biophys.* 37, 289–316

- Sieweck, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., and James, M. N. G. (1990) Molecular and crystal structures of monomeric porcine pepsin refined at 1.8 Å resolution. *J. Mol. Biol.* 214, 143–170

- Gómez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins. *Proteins* 22, 404–412

- Kolaczyk, D. A., Erensen, D. N., and Makkahdaz, G. I. (2001) Heat capacity changes upon burial of polar and nonpolar groups in proteins. *Protein Sci.* 1, 1343–1352

- Fuhrmann, C. N., Kelch, B. A., Ota, N., and Agard, D. A. (2004) The 0.83 Å resolution crystal structure of alpha-lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain. *J. Mol. Biol.* 338, 999–1013

- Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding. *Curr. Opin. Struct. Biol.* 19, 31–38