BIO DESIGN

pISSN 2288-6982 | eISSN 2288-7105

MINI REVIEW P 117-128

Probing Transient Partial Unfolding in Proteins by Native-State Proteolysis

Chiwook Park*

Department of Medicinal Chemistry and Molecular Pharmacology, Bindley Bioscience Center, Purdue University, West Lafayette, IN 47907, USA. *Correspondence: chiwook@purdue.edu

Even under native conditions, proteins can have various non-native conformations, which are in equilibrium with the native functional conformations. Though the populations of the non-native conformation are usually small under native conditions, knowledge of the energy and the structure of the non-native forms is critical in understanding how proteins acquire and lose their structures. Native-state proteolysis is an experimental approach to selectively investigate a dominant partially unfolded form using proteolysis as a structural probe. From analyzing the kinetics of proteolysis of a protein under native conditions, we determine the energetics of unfolding to the partially unfolded form. The effect of urea on the energetics of partial unfolding reports the degree of unfolding in the partially unfolded form. Also, by assessing the effect of a point mutation on the energetics of partial unfolding, we elucidate the structure of the partially unfolded form. Partially unfolded forms probed by native-state proteolysis provide valuable information on the mechanisms of protein folding and protein degradation. In this review, we survey the principle and the applications of native-state proteolysis and also examine the pros and cons of the method in comparison with other experimental approaches.

INTRODUCTION

Well-defined and compactly-folded structures are essential for biochemical functions of most proteins. The folded structure of a protein is achieved through a spontaneous folding process of searching the global energy minimum under native conditions. Because protein folding is a reversible process, the folded form of a protein is in thermodynamic equilibrium with its unfolded forms. The free energy difference between the folded and globally unfolded forms (ΔG_{unf}°) defines the thermodynamic stability of a protein (Figure 1A). The typical thermodynamic stability of proteins is in the range of 4 – 12 kcal/mol. Due to this gap in free energy, the population of proteins in globally unfolded forms is miniscule under native conditions. However, globally unfolded forms are not the only non-native forms that proteins can assume under native conditions. Many partially unfolded forms are also in equilibrium with the folded form (Figure 1A).

The native-state ensemble of protein conformations, therefore, consists of various non-native conformations (partially and globally unfolded forms) as well as the folded form (Englander, 2000; Hilser et al., 2006). The population of protein molecules in each non-native conformation in the ensemble is determined by the Boltzmann distribution according to their free energy relative to the native form. Also, according to the ergodic principle, a trajectory of the conformation of a protein molecule includes transient unfolding to non-native conformations and that the probability of the protein molecule in each non-native conformation is again determined by their free energy relative to the native form. The higher energy levels of non-native conformations keep their populations small under native conditions. Typical biophysical characterizations or biochemical assays report ensemble-averaged values, which are dominated by the molecules in the native forms. When our interest is in these ensemble-averaged values, the presence of the rarely-populated non-native forms is not a concern. However, knowledge on these non-native forms is critical in understanding how proteins acquire their native structures (protein folding) and how proteins lose their native structures (protein inactivation or degradation).

A conformational energy landscape is a map of every possible conformation of a protein according to their energy and structural distance relative to its native form. The current view of conformational energy landscapes of natural proteins is frequently compared to a funnel with native forms at the tip and globally unfolded forms near the outer rim (Figure 1B) (Dill and Chan, 1997; Wolynes et al., 1995). The energy landscapes may contain partially unfolded forms at scattered local minima between the globally unfolded forms and the native forms. In this funnel-like conformational energy landscapes, unfolded protein molecules near the funnel rim find their routes to the tip of the funnel to achieve the native structures. The molecules may stay in local minima along their routes as a partially folded (or partially unfolded) form, i.e. folding intermediates (Kim and Baldwin, 1982, 1990). Because the same conformational energy landscape of a protein defines both its folding process and the distribution of its conformations under native conditions, knowledge of the partially unfolded forms that exist under native conditions provides us with a valuable insight on the route the protein takes during the folding process (Bai, 2006; Chamberlain and Marqusee, 1997;

Englander, 2000).

Transient excursion of a folded protein molecule to a partially unfolded form may result in irreversible inactivation of proteins (Chiti and Dobson, 2006; Dobson, 2003; Frokjaer and Otzen, 2005; Kelly, 1998; Sanchez-Ruiz, 2010). Upon partial unfolding, part of the protein molecule becomes disordered, and hydrophobic surfaces are exposed. The disordered region is susceptible to modifications that may not occur in the native form of the protein. Also, the exposed hydrophobic surfaces may cause aggregation. Irreversible inactivation may occur through globally unfolded forms as well. Under native conditions, however, partially unfolded forms are much more prevalent than globally unfolded forms (Figure 1A). The resistance of a protein against irreversible inactivation, therefore, is not determined by the free energy for global unfolding ($\Delta G_{unf}°$) but rather by the free energy for partial unfolding ($\Delta G_{op}°$). Irreversible inactivation by partial unfolding has a critical implication in protein engineering. When we want to increase the robustness of a protein, it is necessary to suppress partial unfolding by increasing the energy gap between the native forms and the partially unfolded forms responsible for irreversible inactivation. Increasing the global stability without suppressing partial unfolding may not result in more robust proteins.

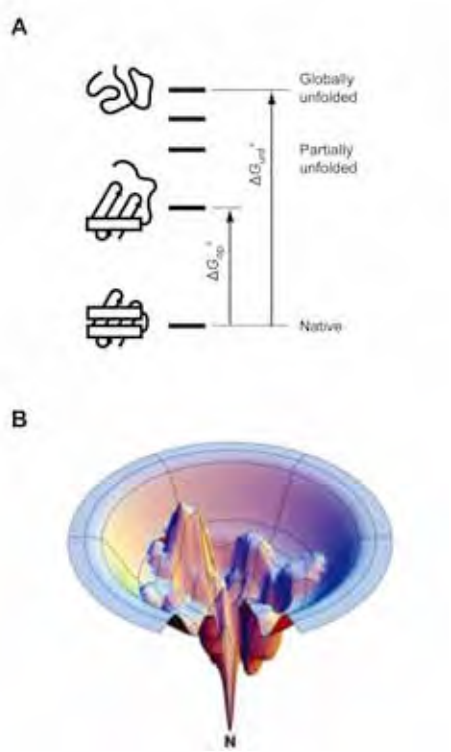

FIGURE 1 | Conformational ensemble of proteins under native conditions. (A) Representative energy diagram of a protein. Multiple partially unfolded forms may be in equilibrium with the native form, and the free energy for partial unfolding ($\Delta G_{op}°$) can be a more relevant parameter to describe the stability of a protein rather than the free energy for global unfolding ($\Delta G_{unf}°$) (B) Schematic representation of the conformational energy landscape of a protein. The native form (N) is located at the tip of the funnel-like energy landscape, and partially unfolded forms may populate at the bumpy region. (The image is originally created by Dill and Chan (Dill and Chan, 1997) and reused under Creative Commons Attribution 4.0 International License.)

Investigating partially unfolded forms under native conditions is challenging because the population of partially unfolded forms is miniscule compared with that of native forms. Typical biophysical methods to monitor protein conformations, such as circular dichroism, fluorescence, and NMR, do not allow us to detect these rare partially unfolded forms with the overwhelming majority of native forms in the background. Therefore, the detection of partially unfolded forms needs to rely on chemical reactions that occur only in partially unfolded forms, such as native-state hydrogen/deuterium exchange (HX) (Englander et al., 1997; Hvidt and Linderstrøm-Lang, 1954). When exposed to the solvent, amide protons are readily exchanged with protons in the solvent. However, amide protons forming hydrogen bonds in proteins are protected against the exchange and only exchanged upon the breakage of the hydrogen bonds. By monitoring how fast each amide proton disappears in a protein dissolved in D_2O, one can determine how frequently each amide loses its folded structure (Bai et al., 1995a; Chamberlain et al., 1996). Native-state HX is typically monitored by NMR (Bai et al., 1995a) or mass spectrometry (Englander et al., 2003; Mandell et al., 1998). Cysteine labeling is another approach to monitor partial unfolding (Feng et al., 2001; Silverman and Harbury, 2002). Analogous to native-state HX, chemical modification of a buried cysteine residue occurs only when the residue is exposed to the solvent upon unfolding.

This review summarizes the recent development of proteolysis as an alternative approach to investigate partial unfolding (Park and Marqusee, 2004). Proteolysis has been used to probe conformational changes in proteins for decades, but most of the applications were qualitative, e.g. identifying the location of transient unfolding from the cleavage patterns in limited proteolysis (Fontana et al., 1997; Heinrikson, 1977). However, a quantitative analysis of proteolytic susceptibility of a protein under native conditions (native-state proteolysis) offers valuable information on the energetics of partial unfolding and the structure of the partially unfolded form responsible for proteolysis (Chang and Park, 2009; Kasper et al., 2014b; Park and Marqusee, 2004). Moreover, native-state proteolysis is relatively simple in interpretation and versatile in the choice of experimental conditions.

MECHANISM OF PROTEOLYSIS

Before we go into the detailed discussion on native-state proteolysis, I will briefly summarize the history of the scientific research on proteolysis to provide the historical perspective to the current view on the mechanism of proteolysis. Here, I will focus our discussion on proteolysis of proteins, not peptides or small substrates. Also, I will not discuss the catalytic

mechanisms of proteases, which is irrelevant when proteolysis is used as a structural probe for unfolding.

The history of the research on proteolysis is as long as the history of enzyme research. Digestion of proteins in food by digestive proteases was actively investigated even in the nineteenth century. In his book "On Digestive Proteolysis" published in 1895, Chittenden has described quite thoroughly the process of proteolysis of food proteins by pepsin and trypsin (Chittenden, 1895). These studies were apparently aimed at understanding the physiology of digestion but also offered an insight on the structure of proteins. Biochemists in the nineteenth century knew that proteins were composed of various amino acids, though not every amino acid was discovered. However, the chemical structure of proteins had been enigmatic to them until Fischer and Hofmeister independently proposed the peptide theory of protein structures in 1902 (Tanford and Reynolds, 2001). Later, as analytical methods had advanced, the release of free amine groups from proteolysis became a strong support for the peptide theory (Van Slyke, 1911). In early twentieth century, biochemists also discovered that denatured proteins are much more susceptible to proteolysis than native proteins (Calvery, 1933), and proteolysis has been utilized as a structural probe to test protein refolding (Bernheim et al., 1942). The greater proteolytic susceptibility of denatured proteins also hinted to the biochemists that proteins may assume more extended conformations upon denaturation. Soon, biochemists suspected that the proteolysis of native proteins actually occurs through denaturation of the proteins. From his investigation on the temperature effect on the proteolysis kinetics, Linderstrøm-Lang proposed in 1938 that proteolysis may occur through the denatured state that is in equilibrium with the native state (Linderstrøm-Lang et al., 1938). It is noteworthy that no protein structures were available in this era. The first atomic structure of a protein was reported much later in 1958, and several more protein structures followed in 1960s (Tanford and Reynolds, 2001). The compact structures of native proteins clearly supported the earlier proposals that proteolysis occurs through denaturation.

However, knowing the structure of native proteins answered only half of the question of how proteolysis of a native protein occurs. The other half of the question requires knowledge of the denatured state (or unfolded state) responsible for proteolysis. It has been observed even in early twentieth century that proteolysis of a protein frequently produces distinct intermediates before complete digestion into short peptides (Tiselius and Eriksson-Quensel, 1939). These partially cleaved proteins, such as plakalbumin (Ottesen and Wollenberger, 1952) and ribonuclease S (Richards, 1958), were once subjects of extensive investigations. In these examples, the cleavage of one specific amide bond out of hundreds of amide bonds in proteins implied that proteolysis does not occur through completely unfolded state (globally unfolded state) of the proteins (Linderstrøm-Lang, 1950; Ottesen, 1967). In 1967, Rupley wrote in his rewiew on proteolytic susceptibility (Rupley, 1967):

> Indeed, a compact conformation conceivably may possess *no* bond which can be cleaved. In such instances, the observed rate of proteolysis may result entirely from attack upon bonds which are exposed in some partially unfolded species, present in low concentration and in equilibrium with the compact form (in effect, attack upon a "denatured" species).

This view of proteolysis of native proteins is basically identical to the current view of proteolysis, though biochemists did not have any direct evidence for the rarely-populated partially unfolded forms in equilibrium with native forms. It was only 1990s when the existence of these rarely-populated partially unfolded forms was experimentally confirmed by a quantitative analysis of the kinetics of native-state HX (Bai et al., 1995b).

Deciphering the physical determinants of proteolytic susceptibility was another huddle in understanding the mechanism of proteolysis. As quantitative descriptions of the thermodynamic stability of proteins became available in 1970s and 1980s, efforts were made to link the proteolytic susceptibility to the thermodynamic stability of proteins (Pace and Barrett, 1984; Parsell and Sauer, 1989; Wang and Kallenbach, 1998). Also, proteolysis kinetics were demonstrated to be consistent with unfolding kinetics of proteins under the condition where proteins are significantly destabilized and the conformational change of proteins is rate-limiting for proteolysis (Arnold and Ulbrich-Hofmann, 1997; Imoto et al., 1986). However, the complete description of proteolytic susceptibility requires a quantitative analysis of the energetics of partial unfolding, which became realized only when native-state proteolysis was introduced recently (Park and Marqusee, 2004).

What is the current view of the mechanism of proteolysis? Proteases catalyze the hydrolysis of peptide bonds in proteins. For catalysis, however, a peptide bond needs to bind to the active site of a protease in a productive way, which requires a stretch of peptide chain to be exposed to the solvent in an extended conformation (Figure 2) (Hubbard, 1998). Unless being in a disordered region, peptide bonds in a folded protein cannot bind to the active site of a protease due to the improper

FIGURE 2 | Kinetic model of native-state proteolysis. The native form is resistant against proteolysis, and proteolysis occurs after a conformational change to the cleavable form. k_{op} and k_{cl} are the forward and reverse rate constants for the conformational change, respectively. k_{int} is the rate constant for the proteolysis of the cleavable form by a protease.

conformation for binding as well as steric hindrance from the rest of the protein (Tyndall et al., 2005). This structural requirement necessitates unfolding of proteins for proteolysis (Hubbard et al., 1994). Proteolytic susceptibility, defined as the rate of proteolysis of a protein under a given condition, is a function of the energetics of the unfolding process necessary for proteolysis as well as the protease activity. Frequently proteolysis occurs through partial unfolding, and the energetics of partial unfolding is a critical factor determining the proteolytic susceptibility of a protein. The conformational energy landscape, which is encoded by the amino-acid sequence of the protein, defines the population of the molecules in the proteolytically susceptible conformations and the kinetic barriers for the conformational change to the susceptible conformations.

NATIVE-STATE PROTEOLYSIS

Native-state proteolysis is defined as proteolysis of proteins under native conditions where the most of the protein in interest is in its folded conformation. When native-state proteolysis of a protein occurs through a partially unfolded form, we can obtain valuable information on the partially unfolded form by investigating the process of proteolysis. The kinetic model of native-state proteolysis is analogous to that of native-state HX (Figure 2) (Park and Marqusee, 2004). In this kinetic model, we assume that proteolysis occurs through a non-native form, which we name a cleavable form. Proteolysis can occur through multiple cleavable forms including globally unfolded forms. However, one cleavable form contributes dominantly to the proteolysis of each protein. In the kinetic model, proteolysis of a native protein occurs through two elementary kinetic steps (Figure 2). The first step is the reversible conformational change from the native form to the cleavable form with k_{op} and k_{cl} as the first-order rate constants for the conformational change to the cleavable form (opening) and for the refolding to the native form (closing), respectively. The second step is irreversible proteolysis of the cleavable form by a protease with k_{int} as the pseudo-first-order rate constant for proteolysis of the cleavable form (intrinsic proteolysis). Because the concentration of the cleavable form is very low under native conditions, most of the protease remains free. Therefore, k_{int} corresponds to V/K in Michaelis–Menten kinetics, which also corresponds to the product of k_{cat}/K_m and the total enzyme concentration. While k_{int} is linearly dependent on the concentration of the protease, k_{op} and k_{cl} is independent of the concentration of the protease and determined by the structural properties of the substrate protein.

When a steady-state approximation is applied to the kinetic model, the rate constant for proteolysis of the protein (k_p) is expressed as

$$k_p = \frac{k_{op} k_{int}}{k_{cl} + k_{int}} \quad (1)$$

When k_p is plotted as a function of k_{int} with constant values of k_{op} and k_{cl}, the plot shows a asymptotic behavior (Figure 3A). We describe the kinetics in two kinetic regimes separately, following the description of the kinetics of native-state HX (Hvidt and Nielsen, 1966). When $k_{cl} << k_{int}$, Eq. 1 is simplified as

$$k_p = k_{op} \quad (2)$$

In this kinetic regime, the overall proteolysis rate is limited by the conformational change to the cleavable form (Figure 3B), which we call EX1-like kinetics (Ferraro et al., 2004; Park and Marqusee, 2004). EX stands for exchange in native-state HX. Because the conformational change step is rate-limiting, the proteolysis rate is independent of k_{int} and, therefore, also independent of the protease concentration (Figure 3A).

When $k_{cl} >> k_{int}$, Eq. 1 is simplified as

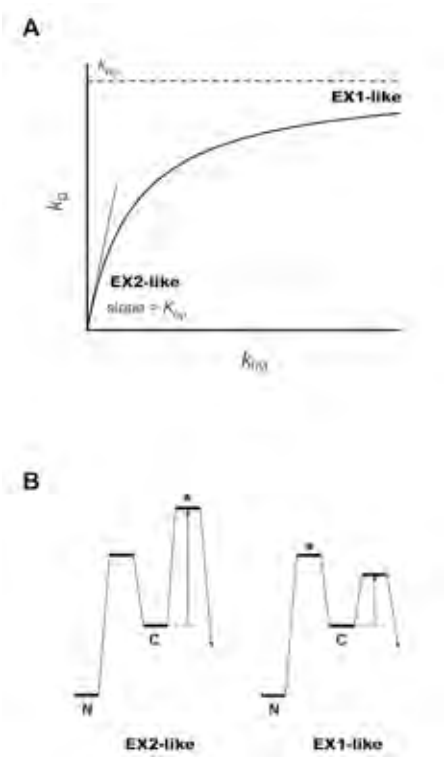

FIGURE 3 | EX1-like and EX2-like kinetics in native-state proteolysis. (A) Dependence of k_p on k_{int}. In the EX1-like kinetics, k_p converges k_{op}. In the EX2-like kinetics, k_p is linearly dependent on k_{int} with the slope corresponding to K_{op} (= k_{op}/k_{cl}). (B) Schematic reaction energy diagrams for EX1-like and EX2-like kinetics. N and C correspond to the native and cleavable forms, respectively. The transition state of the rate limiting step is indicated by an asterisk. The barrier determining k_{int} is indicated by an upward arrow. As k_{int} is increased, i.e. the barrier is lowered, the rate-limiting step changes from the proteolysis step to the conformational change step, which results in the shift from EX2-like to EX1-like kinetics.

$$k_p = \frac{k_{op}}{k_{cl}} k_{int} \quad (3)$$

In this kinetic regime, the overall proteolysis rate is limited by the proteolysis step catalyzed by a protease (Figure 3B), which we call EX2-like kinetics (Bai et al., 1995a; Park and Marqusee, 2004). The proteolysis rate is linearly dependent on k_{int} and, therefore, also dependent on the protease concentration (Figure 3A). When the equilibrium constant for the conformational change step is defined as K_{op}, Eq. 3 can be rewritten as

$$k_p = K_{op} k_{int} \quad (4)$$

Using Eq. 4, one can calculate K_{op} from experimentally determined k_p and k_{int}, and also the free energy required for the conformational change to the cleavable form ($\Delta G_{op}°$) from K_{op}. Due to the utility of $\Delta G_{op}°$ in investigations of partial unfolding, most applications of native-state proteolysis are performed under the conditions that ensure EX2-like kinetics. A simple test for EX2-like kinetics is to determine k_p at varying concentration of a protease. A linear dependence of k_p on the protease concentration confirms that the reaction occurs through EX2-like kinetics.

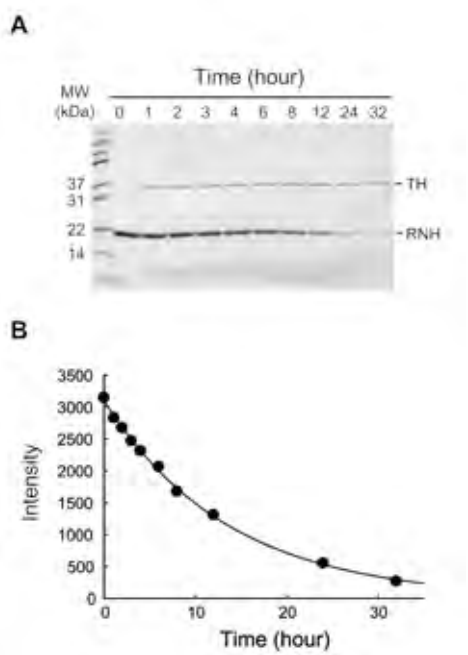

FIGURE 4 | Determination of k_p. (A) Monitoring proteolysis by SDS PAGE. Proteolysis of RNase H by thermolysin is monitored from the disappearance of the intact protein on SDS PAGE. MW indicates the molecular weight markers. TH and RNH indicate the bands corresponding to thermolysin and RNase H, respectively. (B) Kinetics of proteolysis. The change in the band intensity of the intact RNase H is plotted against time. The plot was fitted with a first-order rate equation. Images are adapted from (Park and Marqusee, 2004).

The common way to determine k_p is to monitor the disappearance of the intact protein in a proteolysis reaction by SDS PAGE (Figure 4A) (Park and Marqusee, 2004). In native-state proteolysis, we typically investigate only the proteolysis of the intact protein and do not examine the subsequent proteolysis of the cleaved products. The proteolysis of the cleavage products does not provide any valuable information on the partial unfolding of the protein itself, simply because the cleaved products are distinct entities from the native protein. At designated time points, a proteolysis reaction is quenched and the relative quantity of the remaining intact protein in the reaction is determined from the intensity of the corresponding band on a SDS PAGE gel. The band intensities are plotted against the time when the reaction is quenched, and the plot is fitted with a first-order rate equation with k_p as the rate constant (Figure 4B) (Park and Marqusee, 2004). The reaction kinetics may deviate from a first-order rate equation due to product inhibition especially when the concentration of the substrate protein is high (Kasper et al., 2014a).

We approximate k_{int} using k_{cat}/K_m values determined with unstructured peptide substrates under identical conditions to the native-state proteolysis reaction. If the cleavage site in the protein substrate is known, we design a peptide substrate with an identical sequence to the cleavage site. If the cleavage site is unknown, we use a generic peptide substrate for the specific protease. Proteolysis of *E. coli* ribonuclease H (RNase H) by thermolysin produces two cleavage products discernable on an SDS PAGE gel at early time points (Figure 4A) (Park and Marqusee, 2004). By N-terminal sequencing and mass spectrometry, we identified the cleavage products to be formed from hydrolysis of the peptide bond between Thr92 and Ala93. In this case, we approximated k_{int} with the k_{cat}/K_m value determined with a tetrapeptide substrate (Lys-Thr-Ala-Asp) that we designed based on the sequence of the cleavage site (Park and Marqusee, 2004). However, *E. coli* maltose binding protein (MBP) (Chang and Park, 2009) and *E. coli* dihydrofolate reductase (DHFR) (Kasper et al., 2014b) are digested by thermolysin without producing any discernable early cleavage products that may reveal the initial cleavage site. In these cases, we approximated k_{int} with the k_{cat}/K_m value determined with a generic peptide substrate (Ala-Gly-Leu-Ala) for thermolysin. When a generic peptide substrate is used to estimate k_{int}, the uncertainty in k_{int} can be a source of a systematic error in $\Delta G_{op}°$.

Nonspecific proteases are most suitable for native-state proteolysis to probe partial unfolding in proteins without any restriction from sequence specificity. Thermolysin is a primary choice for native-state proteolysis because the protease is nonspecific but still prefers hydrophobic residues to polar residues (Keil, 1992). This selectivity for hydrophobic residues makes thermolysin a useful probe for unfolding (Hecht et al., 1983; Heinrikson, 1977), as unfolding of proteins exposes buried hydrophobic residues. Moreover, the exceptional stability and

the broad pH optimum of thermolysin (Dahlquist et al., 1976; Feder and Schuck, 1970) are also valuable properties for the application of native-state proteolysis under various conditions. $\Delta G_{op}°$ determined by native-state proteolysis may be dependent on the choice of the protease because a cleavable form probed by a protease might not be cleavable by another protease. This dependence on the choice of the protease would be less significant if partial unfolding exposes an extended region that contains cleavage sites for proteases with different specificities. We have previously determined $\Delta G_{op}°$ of MBP with both thermolysin and trypsin, and confirmed that the result was not significantly dependent on the choice of protease (Chang and Park, 2009).

INVESTIGATION OF THE ENERGETICS OF PARTIAL UNFOLDING

To investigate transient partial unfolding of a protein by native-state proteolysis, one needs to confirm first that proteolysis of the protein occurs through partial unfolding, not global unfolding. In other words, it should be confirmed that the cleavable form (Figure 2) is a partially unfolded form. A simple but crude way to check the identity of the cleavable form is to compare $\Delta G_{op}°$ with $\Delta G_{unf}°$ (Figure 1). If proteolysis occurs through partial unfolding, $\Delta G_{op}°$ will be smaller than $\Delta G_{unf}°$. If proteolysis occurs through global unfolding, $\Delta G_{op}°$ will be similar to $\Delta G_{unf}°$. Table 1 lists $\Delta G_{op}°$ and $\Delta G_{unf}°$ for the proteins that have been investigated with native-state proteolysis. The $\Delta G_{op}°$ values for RNase H (5.5 kcal/mol) and MBP (8.3 kcal/mol) were significantly smaller than the $\Delta G_{unf}°$ values of each protein (10.6 kcal/mol and 15.7 kcal/mol, respectively), clearly demonstrating that proteolysis occurred through partial unfolding. The $\Delta G_{op}°$ value for DHFR (4.9 kcal/mol) was smaller than its $\Delta G_{unf}°$ (6.6 kcal/mol). However, the difference between $\Delta G_{op}°$ and $\Delta G_{unf}°$ was not large enough to demonstrate that proteolysis of DHFR occurs through partial unfolding, considering the potential systematic error in the absolute value of $\Delta G_{op}°$ we described above. The $\Delta G_{op}°$ value for the Src SH2 domain (8.1 kcal/mol) was somewhat greater than $\Delta G_{unf}°$ (7.2 kcal/mol), suggesting that the proteolysis occurred through global unfolding.

A more definitive way to figure out the identity of the cleavable form is to determine the effect of urea on $\Delta G_{op}°$ (Park and Marqusee, 2004). The free energy for a conformational change in a protein has been empirically known to be linearly dependent on the concentration of a chemical denaturant. The dependence (*m*-value) is proportional to the change in the solvent accessible surface area accompanying a conformational change (Myers et al., 1995). When proteolysis occurs through global unfolding, the slope of the plot of $\Delta G_{op}°$ versus urea concentration is close to the *m*-value of global unfolding. A slope smaller than the *m*-value for global unfolding indicates that proteolysis occurs through partial unfolding. This approach to determine the degree of unfolding in a partially unfolded form is commonly employed in native-state HX also (Bai et al., 1995b; Chamberlain et al., 1996). Figure 5 shows an example of the urea effect on $\Delta G_{op}°$ in a hypothetical protein. At low concentrations of urea, the slope of $\Delta G_{op}°$ is less than that of global unfolding, suggesting that proteolysis occurs through partial unfolding. At higher concentrations of urea, $\Delta G_{op}°$ converges with $\Delta G_{unf}°$ and also the slope of $\Delta G_{op}°$ becomes indistinguishable from that of global unfolding, suggesting that proteolysis occurs through global unfolding. In this example, the switch over of the proteolysis mechanism occurs because the globally unfolded form becomes more accessible than the partially unfolded form at higher concentrations of urea. This type of behavior has been also observed in HX (Bai, 2006; Chamberlain et al., 1996). Therefore, the *m*-value of $\Delta G_{op}°$ offers a definitive way to determine whether proteolysis occurs through partial unfolding or not. Moreover, the ratio of the *m*-value of partial unfolding to that of global unfolding tells us the degree of unfolding in the partially unfolded form quantitatively.

Table 1 lists *m*-values determined by native-state proteolysis. For RNase H, the *m*-value of $\Delta G_{op}°$ is not significantly different from 0 while its *m*-value for global unfolding is 2.2 kcal/mol. The negligible *m*-value of RNase H suggests that unfolding for proteolysis occurs locally, probably involving only the extended loop region where the cleavage site is located (Figure 6A). A negligible $\Delta C_p°$ determined from the temperature effect on $\Delta G_{op}°$ also confirmed the localized unfolding for proteolysis in RNase H (Youn and Park, 2009). For MBP, the *m*-value of $\Delta G_{op}°$ is 0.7 kcal/mol while the *m*-value for global unfolding is 5.5 kcal/mol, suggesting that proteolysis occurs through a partially unfolded form that exposes ~10% of the total change in the solvent accessible surface area upon global unfolding. For DHFR, the *m*-value of $\Delta G_{op}°$ is 0.7 kcal/mol while the *m*-values for

TABLE 1 | Energetics of unfolding probed by native-state proteolysis

Protein	Native-state proteolysis		Equilibrium unfolding	
	$\Delta G_{op}°$ (kcal/mol)	*m*-value (kcal $M^{-1}mol^{-1}$)	$\Delta G_{unf}°$ (kcal/mol)	*m*-value (kcal $M^{-1}mol^{-1}$)
E. coli ribonuclease H[a]	5.5	0	10.6	2.2
E. coli maltose binding protein[b]	8.3	0.7	15.7	5.5
E. coli dihydrofolate reductase[c]	4.9	0.7	6.6	2.3
Chicken c-Src SH2 domain[d]	8.1	1.2	7.2	1.2

[a](Park and Marqusee, 2004), [b](Chang and Park, 2009), [c](Kasper et al., 2014b), [d](Wildes et al., 2006)

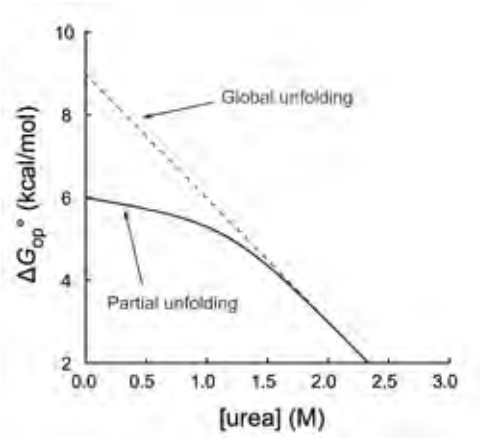

FIGURE 5 | Effect of urea on ΔG_{op}°. The plot shows the dependence of ΔG_{op}° on urea (solid line) in case that proteolysis occurs through a partially unfolded form with ΔG_{op}° of 6.0 kcal/mol and the *m*-value for partial unfolding of 0.5 kcal $mol^{-1}M^{-1}$. The protein has the global stability of 9.0 kcal/mol and the *m*-value for global unfolding of 3.0 kcal $mol^{-1}M^{-1}$. The change in the global stability in varying concentrations of urea (dashed line) is also shown for comparison. Proteolysis occurs through partial unfolding at lower concentrations of urea but through global unfolding at higher concentrations of urea.

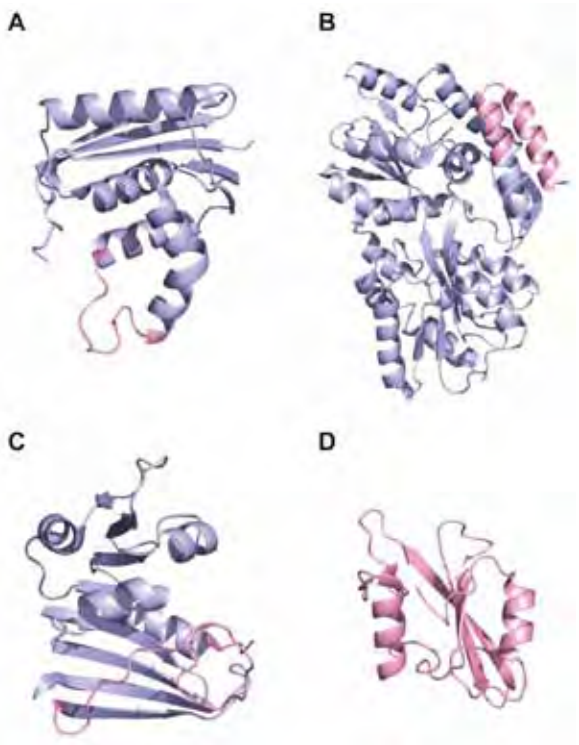

FIGURE 6 | Structures of partially unfolded proteins determined by native-state proteolysis. The unfolded regions of the partially unfolded proteins are shown in pink on the ribbon representation of the crystallographic structures of of RNase H (PDB CODE: 1F21) (A), MBP (PDB CODE: 1OMP) (B), DHFR (PDB CODE: 5DFR) (C), and the Src SH2 domain (PDB CODE: 1SPR) (D). The protein structures are not to scale.

global unfolding are 2.3 kcal/mol, suggesting that the partially unfolded form exposes ~30% of total change in the solvent accessible surface area upon global unfolding. Further studies of the structure of the partially unfolded form confirmed that the estimated degree of unfolding using the *m*-values is a reasonable approximation (Figures 6B and 6C; also see the next section). For the Src SH2 domain, the *m*-value of ΔG_{op}° (1.2 kcal/mol) is identical to the *m*-value for global unfolding (1.2 kcal/mol), confirming that proteolysis occurs through global unfolding even in the absence of a chemical denaturant (Figure 6D). The lack of a partially unfolded form detectable by proteolysis indicates that the SH2 domain is a cooperative structural domain without any partially unfolded forms. This observation is consistent with the finding that the SH2 domain folds in a single dominant kinetic step without any detectable folding intermediate (Wildes et al., 2006).

Investigation of urea effect on ΔG_{op}° also offers a way to assess the systematic error in ΔG_{op}°, which may result from the influence of product inhibition on the determination of k_p values and the uncertainty in k_{int}. As shown in Figure 5, at higher urea concentrations proteolysis may occur through global unfolding. The similarity of ΔG_{op}° to ΔG_{unf}° under this condition confirms that the systematic error in ΔG_{op}° is not significant, which we have demonstrated with RNase H (Park and Marqusee, 2004) and DHFR (Kasper et al., 2014b). One caveat is that ΔG_{op}° for proteolysis through global unfolding might be slightly greater than ΔG_{unf}° determined from equilibrium unfolding due to the lack of proline isomerization in transiently populated globally unfolded forms, which has been also observed in native-state HX (Huyghues-Despointes et al., 1999). The systematic error in ΔG_{op}° is not a significant concern in the measurement of the *m*-values because the *m*-value is determined by the change in ΔG_{op}°, not by the absolute value of ΔG_{op}°.

INVESTIGATION OF THE STRUCTURE OF PARTIALLY UNFOLDED FORMS

Native-state proteolysis provides information on not only the energetics of partial unfolding but also the structure of the partially unfolded form. In cases as with RNase H, we may determine the initial cleavage site by characterizing the products from the initial cleavage. This approach is applicable only when the initial cleavage produces a stable intermediate (a nicked protein) that accumulates significantly. However, initial cleavage products tend to be more susceptible to proteolysis than their native proteins due to the loss of the structural integrity and do not accumulate to detectable levels. This type of proteolysis has been previously dubbed a 'one-by-one' or 'all or none' mechanism (Linderstrøm-Lang, 1950; Tiselius and Eriksson-Quensel, 1939), because each molecule is digested completely to the final products once cleaved. In this case, one cannot determine the initial cleavage site by analyzing cleavage products.

We developed a general approach to determine the structure

of the cleavable form without relying on the presence of initial cleavage products. In this approach, we probe the structure of a partially unfolded form by investigating the effects of point mutations on the stabilities of the native and cleavable forms. From the effects of mutations, we determine a parameter dubbed 'φ_c value' (Figure 7) (Chang and Park, 2009). This value is analogous to φ value, which is commonly used to probe the structure of proteins in the transition state during folding (Matouschek et al., 1989). A φ_c value is defined as

$$\varphi_C = \frac{\Delta\Delta G_{U\text{-}C}^{\circ}}{\Delta\Delta G_{U\text{-}N}^{\circ}} \quad (5)$$

where $\Delta\Delta G_{U\text{-}N}^{\circ}$ is the change in the global stability upon mutation, and $\Delta\Delta G_{U\text{-}C}^{\circ}$ is the change in the stability of the cleavable form upon mutation. As shown in Figure 7, if the mutated residue belongs to the folded region of the cleavable form, the mutation would destabilize the native form and the cleavable form similarly, and the φ_c value would be close to 1. However, if the mutated residue belongs to the unfolded region of the cleavable form, the mutation would destabilize the native form but would not destabilize the cleavable form, and the φ_c value would be close to 0. Therefore, φ_c values indicate the intactness of the native contacts of the mutated residue in the cleavable form. $\Delta G_{U\text{-}N}^{\circ}$ is determined by monitoring unfolding of proteins in varying concentrations of chemical denaturant. $\Delta G_{U\text{-}C}^{\circ}$ (=$\Delta G_{U\text{-}N}^{\circ}$ – $\Delta G_{C\text{-}N}^{\circ}$) is determined with $\Delta G_{U\text{-}N}^{\circ}$ and $\Delta G_{C\text{-}N}^{\circ}$ (=ΔG_{op}°) from native-state proteolysis. The systematic error in $\Delta G_{C\text{-}N}^{\circ}$ from native-state proteolysis is not a significant concern in determining φ_c values because we use the change in $\Delta G_{C\text{-}N}^{\circ}$, not on the absolute value of $\Delta G_{C\text{-}N}^{\circ}$, caused by a mutation.

In both MBP and DHFR, proteolysis occurs without any

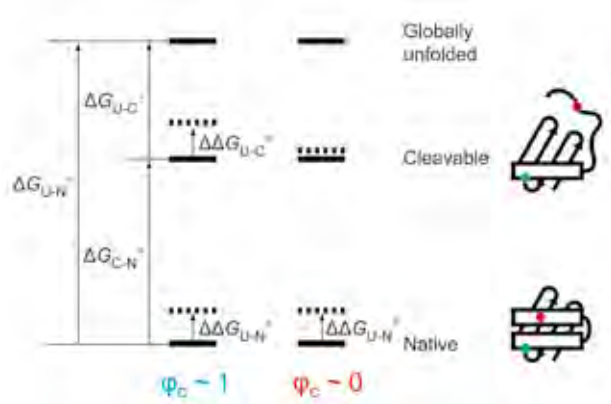

FIGURE 7 | φ_c value analysis. If a residue is in the folded region of the cleavable form (blue dot), the mutation of the residue destabilizes both native and cleavable forms similarly, which results in a φ_c value close to 1. If a residue is in the unfolded region of the cleavable form (red dot), the mutation of the residue destabilizes the native form much more than the cleavable form, which results in a φ_c value close to 0. $\Delta G_{U\text{-}N}^{\circ}$ is the free energy for global unfolding (= ΔG_{unf}°), $\Delta G_{C\text{-}N}^{\circ}$ is the free energy for partial unfolding to the cleavable form (= ΔG_{op}°), and $\Delta G_{U\text{-}C}^{\circ}$ is the free energy for the unfolding of the cleavable form.

detectable early cleavage products, and the initial cleavage sites were unknown. Still, we have successfully determined the structures of the cleavable form in MBP and DHFR using the φ_c value analysis (Figure 8) (Chang and Park, 2009; Kasper et al., 2014b). In MBP, the residues with low φ_c values cluster in the hydrophobic core formed by four helices including two C-terminal helices (Figure 8A). This distribution of φ_c values strongly suggests that a large portion of the two C-terminal helices is unfolded and the hydrophobic core is disrupted in the cleavable form. The structure of the cleavable form indicates that the unfolding of the C-terminal helices is uncoupled with the unfolding of the rest of the protein. To test if the region that remains folded in the cleavable form can fold without one or both of the C-terminal helices, we constructed two deletion mutants, Δ334 and Δ354, which have deletions after residues 334 and 354, respectively (Chang and Park, 2009). Therefore, Δ334 lacks both C-terminal helices (C1 and C2 helices), and Δ354 lacks only the last C-terminal helix (C1 helix). Though the thermodynamic stabilities are significantly compromised by the deletions, both deletion mutants still fold into well-defined structures (Chang and Park, 2009), confirming that the rest of protein remains folded even when the C-terminal helices are unfolded.

The interaction with maltose provides another piece of information on the structure of the cleavable form of MBP. The rate of proteolysis of MBP was not affected in the presence of 100 μM maltose (Chang and Park, 2009). MBP is fully saturated with maltose under this condition as the concentration is much greater than the known K_d value of MBP-maltose complex (~1 μM) (Marvin and Hellinga, 2001; Telmer and Shilton, 2003). This insensitivity of proteolytic susceptibility to the binding of the ligand suggests that the cleavable form still retains its affinity to maltose. Trp340, which is located at the N-terminal half of C2 helix, is known to be critical for binding to maltose. When this residue is mutated to alanine, the binding affinity to maltose is significantly compromised (K_d > 1 mM) (Martineau et al., 1990). The observation that the cleavable form retains the binding affinity to maltose implies that Trp340 is still in its native conformation in the cleavable form. Therefore, it is likely that C2 helix is only partially unfolded in the cleavable form (Figure 6B).

The unfolding of the C-terminal helices in MBP provides an insight on the folding of this protein in cells. Being a periplasmic protein, MBP is synthesized in the cytosol as a preprotein and exported into the periplasmic space through a protein channel (Tomkiewicz et al., 2008). Folding of MBP without the C-terminal helices suggests that the protein may acquire certain folded structures before the C-terminal helices are translocated completely during the process of export. This partial folding of MBP must be beneficial because the acquisition of the structure before the completion of the translocation will minimize the occurrence of harmful consequences such as aggregation or degradation from the exposure of unfolded proteins.

In DHFR, residues with low φ_c values cluster on one side of the central β-sheet that interacts with two extended loops, the Met20

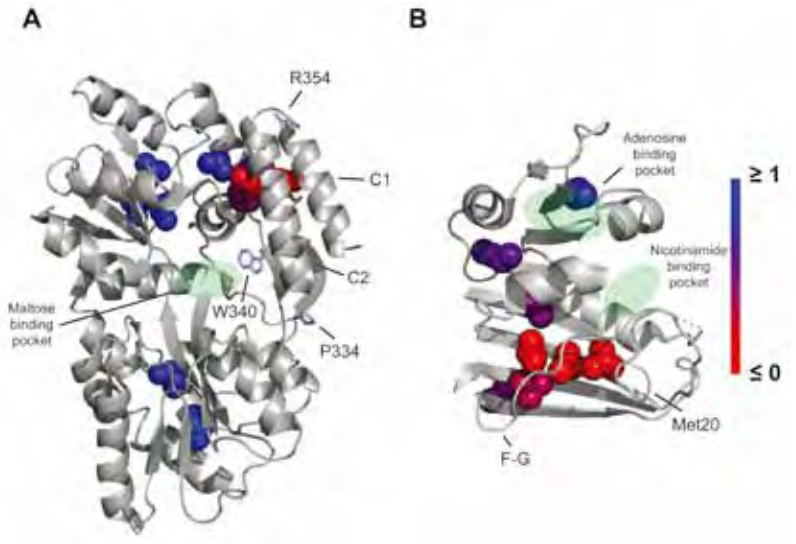

FIGURE 8 | φ_c values of selected residues in maltose binding protein (A) and dihydrofolate reductase (B). The side chains of the mutated residues are shown in color spectrum according to their φ_c values (red for 0 and blue for 1). For clarity, φ_c values greater than 1 or less than 0 were shown as 1 and 0, respectively. The images are adapted from (Chang and Park, 2009) and (Kasper et al., 2014b).

loop and the F-G loop (Figure 8B) (Kasper et al., 2014b). This distribution of φ_c values suggests that the two extended loops are unfolded and the hydrophobic core between the central β-sheet and the loops is disrupted in the cleavable form (Figure 6C). Also, fractional φ_c values in the rest of the protein suggest that even the region that remains folded in the cleavable form may experience loosening of the packing. The structure of the cleavable form of DHFR provides valuable insight on the folding mechanism of this protein. DHFR is known to fold through two major folding intermediates, I_{BP} (BP stands for "burst phase") and I_{HF} (HF stands for "high fluorescence") (Jennings et al., 1993; Kuwajima et al., 1991). However, the structures of the intermediates have been largely unknown. The comparison of the cleavable form we identified by native-state proteolysis with the two known folding intermediates revealed that the cleavable form resembles I_{HF} (Kasper et al., 2014b). The high fluorescence of I_{HF} is known to result from the folding of a hydrophobic core formed by Trp47 and Trp74 (Kuwajima et al., 1991). Ile61, which is in contact with Trp47 and Trp74, showed a relatively high φ_c value (Kasper et al., 2014b), suggesting the hydrophobic core formed by Trp47 and Trp74 remains mostly intact in the cleavable form. In addition, 70% burial of the solvent accessible surface area in the cleavable form (Kasper et al., 2014b) is quite consistent with about 65% of the solvent accessible surface area is buried in I_{HF}. The structure of the cleavable form offers a chance to infer the structure of I_{HF} with much greater details than before. Based on the structure of the cleavable form, we suggested that the folding from I_{HF} to the native form involves the organization of the extended loops, which seems to be coupled with the compaction of the rest of the protein.

The effect of $NADP^+$ binding to the partial unfolding of DHFR is also consistent with the structure of the cleavable form we determined by the φ_c value analysis. $NADP^+$ is one of the products of DHFR, and the ligand binds to the native form of DHFR with the K_d value of 24 μM (Fierke et al., 1987). $NADP^+$ suppresses the partial unfolding of DHFR, but the degree of suppression reaches a plateau at high concentration of $NADP^+$, which suggests that, though the affinity is much lower (K_d ~ 2 mM) than to the native form, $NADP^+$ still binds to the cleavable form of DHFR (Kasper and Park, 2014). The structure of the complex of DHFR and $NADP^+$ provides the molecular basis for the binding of the ligand to the cleavable form (Sawaya and Kraut, 1997). The adenosine moiety of $NADP^+$ interacts mostly with the adenosine-binding domain of DHFR, which remains folded in the cleavable form, while the nicotinamide moiety of $NADP^+$ interacts with the Met20 loop, which is unstructured in the cleavable form. Moreover, the nicotinamide moiety of $NADP^+$ shows two different conformations in the complex; in one conformation, the nicotinamide moiety is bound in the binding cleft formed by the Met20 loop, and in the other conformation, the nicotinamide moiety is outside of the binding cleft, pointing to the bulk solvent (Sawaya and Kraut, 1997). This heterogeneity in the conformation of the nicotinamide moiety in the complex suggests that the interaction between DHFR and the nicotinamide moiety does not contribute much to the total binding affinity. This structural characteristic is consistent with our observation that $NADP^+$ still binds to the cleavable form in spite of the disruption of the nicotinamide-binding site.

The successful applications of the φ_c value analysis in determining the structure of the cleavable forms in MBP and DHFR demonstrate the utility of this approach. Combination of site-directed mutagenesis with native-state proteolysis offers a facile way to identify the region unfolded in the cleavable form, even in case that proteolysis does not produce any detectable early cleavage products. Therefore, native-state proteolysis can be used to reveal the structure of the partially unfolded form as well as to determine the energetics of partial unfolding.

COMPARISON WITH NATIVE-STATE HYDROGEN/DEUTERIUM EXCHANGE

The development of native-state proteolysis has been greatly influenced by the principle of native-state HX. Both methods are based on the same kinetic model (Figure 2). The effect of urea on the energetics of partial unfolding (*m*-value) has a critical utility in estimating the degree of unfolding in both methods (Figure 5).

TABLE 2 | Comparison of native-state HX and proteolysis

	Native-state HX	Native-state proteolysis
Conformations to be probed	Multiple conformations	Conformation with the lowest energy (the cleavable form)
Applicable pH range	Low pH (pH<6)	Broad
Modulation of k_{int}	By changing pH	By changing protease concentrations
Size of proteins	Limited when NMR is used	Not limited
Minimum unfolding for detection	Local fluctuation	Unfolding involving ~10 residues
Instrumentation	Heavy	Very light

As native-state HX has already been established as a powerful method to investigate transient partial unfolding in proteins, why do we still need native-state proteolysis? I will address this question by listing the advantages and the disadvantages of each method (Table 2).

When the native-state HX of a protein is monitored by two-dimensional NMR, the method reveals the energetics of unfolding at a residue-specific level. The residue-specific information can be used to determine the stabilities and the structures of multiple partially unfolded forms. However, native-state proteolysis probes only one partially unfolded form that contributes dominantly to proteolysis (the cleavable form). Still, because the cleavable form is the most accessible partially unfolded form under native conditions, native-state proteolysis is likely to provide information on the partially unfolded form of the most importance in function and regulation of the protein. In other words, native-state proteolysis reveals the Achilles' heel of a protein structure.

Though native-state proteolysis provides information on only one partially unfolded form, this method is more versatile in the choice of experimental conditions than native-state HX. Determination of $\Delta G_{op}°$ by native-state HX commonly requires acidic conditions (typically pH<6). The intrinsic exchange rate (k_{int}) of native-state HX is strongly dependent on pH, and EX2 kinetics are typically achieved only under acidic conditions. At pH>6, k_{int} is commonly too large to achieve EX2 kinetics. In native-state proteolysis, EX2-like kinetics can be achieved simply by changing the concentration of the protease to a proper level (Figure 3). Especially, by choosing a protease active at the pH of the experimental condition, one can employ native-state proteolysis virtually at any pH. Also, while native-state HX by NMR is only applicable to small proteins amenable for NMR, native-state proteolysis is not limited by the size of proteins. Mass spectrometry instead of NMR is frequently employed to monitor native-state HX in larger proteins (Englander et al., 2003; Mandell et al., 1998). However, because mass spectrometry monitors native-state HX at a peptide level, not at a residue level, it is quite complicate to extract quantitative energetics information from complex kinetics data generated by native-state HX by mass spectrometry.

Simplicity in analysis is another merit of native-state proteolysis. Frequently, native-state HX in a protein is dominated by local fluctuation, a brief exposure of amide protons to the solvent without a significant conformational change (Maity et al., 2003). To obtain information on partial unfolding from native-state HX kinetics, signals from trivial local fluctuations need to be filtered out carefully. Because proteolysis requires a minimum of 8-12 residues in an extended formation to form a protease-substrate complex (Hubbard et al., 1994), trivial local fluctuations are basically invisible in native-state proteolysis. Therefore, native-state proteolysis allows us to focus on a non-trivial partially unfolded form. Instrumental requirements are significantly lighter in native-state proteolysis than native-state HX. While native-state HX requires NMR spectrometer or mass spectrometer, native-state proteolysis requires only electrophoresis equipment. The simple instrumentation allows parallel monitoring of partial unfolding under multiple conditions, facilitating a rapid survey of partial unfolding in proteins. Also, monitoring proteolysis by electrophoresis allows applications of native-state proteolysis to unpurified proteins in crude lysates (Chang and Park, 2009; Park et al., 2007).

CONCLUSION

Native-state proteolysis offers a versatile and efficient way to investigate transient partial unfolding in proteins. Though native-state proteolysis may be considered as a coarse-grained approach, the versatility and simplicity make this method a valuable alternative to native-state HX. Native-state proteolysis reports not only the energetics of partial unfolding but also the structure of the partially unfolded form. The partially unfolded form probed by native-state proteolysis is the most populated partially unfolded form, revealing valuable information on the conformational energy landscape of a protein. The partially unfolded form may suggest how the protein acquires its structure (folding) and how the protein loses its structure (aggregation or degradation). A proteomic survey demonstrated a surprising diversity in the proteolytic susceptibility of proteins in a proteome and also the functional relevance of the energetics of partial unfolding probed by proteolysis (Park et al., 2007), suggesting the value of cleavable forms in probing conformational energy

landscapes of proteins. The investigation of partial unfolding in proteins by native-state proteolysis allows us to add another dimension to our understanding of protein structure, which is not easily accessible by other methodologies in structural biology. Getting a better understanding of conformational energy landscapes of proteins by native-state proteolysis will help us elucidate how evolution has shaped conformational energy landscapes for the proper functions of proteins in cells. Also, information on the conformational energy landscape of a protein is valuable in engineering proteins with better structural properties, such as robustness for applications under harsh conditions.

ACKNOWLEGDEMENTS

I thank Chen Chen and Nathan Gardner for helpful comments on this manuscript. The work was supported by US National Science Foundation Grant 1021652.

AUTHOR INFORMATION

The authors declare no potential conflicts of interest.

Original Submission: Nov 26, 2014
Revised Version Received: Dec 3, 2014
Accepted: Dec 5, 2014

REFERENCES

Arnold, U., and Ulbrich-Hofmann, R. (1997). Kinetic and thermodynamic thermal stabilities of ribonuclease A and ribonuclease B. *Biochemistry* **36**, 2166-2172.

Bai, Y. (2006). Protein folding pathways studied by pulsed- and native-state hydrogen exchange. *Chem Rev* **106**, 1757-1768.

Bai, Y., Englander, J.J., Mayne, L., Milne, J.S., and Englander, S.W. (1995a). Thermodynamic parameters from hydrogen exchange measurements. *Methods Enzymol* **259**, 344-356.

Bai, Y., Sosnick, T.R., Mayne, L., and Englander, S.W. (1995b). Protein folding intermediates: Native-state hydrogen exchange. *Science* **269**, 192-197.

Bernheim, F., Neurath, H., and Erickson, J.O. (1942). The denaturation of proteins and its apparent reversal IV: Enzymatic hydrolysis of native, denatured, and apparently reversibly denatured proteins. *J Biol Chem* **144**, 259-264.

Calvery, H.O. (1933). Crystalline egg albumin: The hydrolysis of crystalline egg albumin by pepsin, papain-hydrocyanic acid, and pancreatic proteinase and the subsequent action of some other enzymes on the hydrolysis products produced by these enzymes. *J Biol Chem* **102**, 73-89.

Chamberlain, A.K., Handel, T.M., and Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nat Struct Biol* **3**, 782-787.

Chamberlain, A.K., and Marqusee, S. (1997). Touring the landscapes: Partially folded proteins examined by hydrogen exchange. *Structure* **5**, 859-863.

Chang, Y., and Park, C. (2009). Mapping transient partial unfolding by protein engineering and native-state proteolysis. *J Mol Biol* **393**, 543-556.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366.

Chittenden, R.H. (1895). On Digestive Proteolysis (New Haven, CT: Tuttle, Morehouse & Taylor).

Dahlquist, F.W., Long, J.W., and Bigbee, W.L. (1976). Role of calcium in the thermal stability of thermolysin. *Biochemistry* **15**, 1103-1111.

Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-19.

Dobson, C.M. (2003). Protein folding and misfolding. *Nature* **426**, 884-890.

Englander, J.J., Del Mar, C., Li, W., Englander, S.W., Kim, J.S., Stranz, D.D., Hamuro, Y., and Woods, V.L., Jr. (2003). Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc Natl Acad Sci USA* **100**, 7057-7062.

Englander, S.W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* **29**, 213-238.

Englander, S.W., Mayne, L., Bai, Y., and Sosnick, T.R. (1997). Hydrogen exchange: the modern legacy of Linderstrom-Lang. *Protein Sci* **6**, 1101-1109.

Feder, J., and Schuck, J.M. (1970). Studies on the *Bacillus subtilis* neutral-protease- and *Bacillus thermoproteolyticus* thermolysin-catalyzed hydrolysis of dipeptide substrates. *Biochemistry* **9**, 2784-2791.

Feng, Z., Butler, M.C., Alam, S.L., and Loh, S.N. (2001). On the nature of conformational openings: Native and unfolded-state hydrogen and thiol-disulfide exchange studies of ferric aquomyoglobin. *J Mol Biol* **314**, 153-166.

Ferraro, D.M., Lazo, N., and Robertson, A.D. (2004). EX1 hydrogen exchange and protein folding. *Biochemistry* **43**, 587-594.

Fierke, C.A., Johnson, K.A., and Benkovic, S.J. (1987). Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **26**, 4085-4092.

Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997). Probing the partly folded states of proteins by limited proteolysis. *Fold Des* **2**, R17-26.

Frokjaer, S., and Otzen, D.E. (2005). Protein drug stability: a formulation challenge. *Nat Rev Drug Discov* **4**, 298-306.

Hecht, M.H., Nelson, H.C., and Sauer, R.T. (1983). Mutations in lambda repressor's amino-terminal domain: Implications for protein stability and DNA binding. *Proc Natl Acad Sci USA* **80**, 2676-2680.

Heinrikson, R.L. (1977). Applications of thermolysin in protein structural analysis. *Methods Enzymol* **47**, 175-189.

Hilser, V.J., Garcia-Moreno, E.B., Oas, T.G., Kapp, G., and Whitten, S.T. (2006). A statistical thermodynamic model of the protein ensemble. *Chem Rev* **106**, 1545-1558.

Hubbard, S.J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta* **1382**, 191-206.

Hubbard, S.J., Eisenmenger, F., and Thornton, J.M. (1994). Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci* **3**, 757-768.

Huyghues-Despointes, B.M., Scholtz, J.M., and Pace, C.N. (1999). Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol* **6**, 910-912.

Hvidt, A., and Linderstrøm-Lang, K. (1954). Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochim Biophys Acta* **14**, 574-575.

Hvidt, A., and Nielsen, S.O. (1966). Hydrogen exchange in proteins. *Adv Protein Chem* **21**, 287-386.

Imoto, T., Yamada, H., and Ueda, T. (1986). Unfolding rates of globular proteins determined by kinetics of proteolysis. *J Mol Biol* **190**, 647-649.

Jennings, P.A., Finn, B.E., Jones, B.E., and Matthews, C.R. (1993). A reexamination of the folding mechanism of dihydrofolate reductase from *Escherichia coli*: Verification and refinement of a four-channel model. *Biochemistry* **32**, 3783-3789.

Kasper, J.R., Andrews, E.C., and Park, C. (2014a). Product inhibition in native-state proteolysis. *PLoS ONE* **9**, e111416.

Kasper, J.R., Liu, P.F., and Park, C. (2014b). Structure of a partially unfolded form of Escherichia coli dihydrofolate reductase provides insight into its folding pathway. *Protein Sci* **23**, 1728-1737.

Kasper, J.R., and Park, C. (2014). Ligand Binding to a High-Energy Partially Unfolded Protein. *Protein Science*, In press.

Keil, B. (1992). Specificity of Proteolysis (New York: Springer-Verlag).

Kelly, J.W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* **8**, 101-106.

Kim, P.S., and Baldwin, R.L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* **51**, 459-489.

Kim, P.S., and Baldwin, R.L. (1990). Intermediates in the folding reactions

of small proteins. *Annu Rev Biochem* **59**, 631-660.

Kuwajima, K., Garvey, E.P., Finn, B.E., Matthews, C.R., and Sugai, S. (1991). Transient intermediates in the folding of dihydrofolate reductase as detected by far-ultraviolet circular dichroism spectroscopy. *Biochemistry* **30**, 7693-7703.

Linderstrøm-Lang, K. (1950). Structure and enzymatic break-down of proteins. *Cold Spring Harbor Symp Quant Biol* **14**, 117-126.

Linderstrøm-Lang, K., Hotchkiss, R.D., and Johansen, G. (1938). Peptide bonds in globular proteins. *Nature* **142**, 996.

Maity, H., Lim, W.K., Rumbley, J.N., and Englander, S.W. (2003). Protein hydrogen exchange mechanism: Local fluctuations. *Protein Sci* **12**, 153-160.

Mandell, J.G., Falick, A.M., and Komives, E.A. (1998). Measurement of amide hydrogen exchange by MALDI-TOF mass spectrometry. *Anal Chem* **70**, 3987-3995.

Martineau, P., Szmelcman, S., Spurlino, J.C., Quiocho, F.A., and Hofnung, M. (1990). Genetic approach to the role of tryptophan residues in the activities and fluorescence of a bacterial periplasmic maltose-binding protein. *J Mol Biol* **214**, 337-352.

Marvin, J.S., and Hellinga, H.W. (2001). Manipulation of ligand binding affinity by exploitation of conformational coupling. *Nat Struct Biol* **8**, 795-798.

Matouschek, A., Kellis, J.T., Jr., Serrano, L., and Fersht, A.R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122-126.

Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138-2148.

Ottesen, M. (1967). Induction of biological activity by limited proteolysis. *Annu Rev Biochem* **36**, 55-76.

Ottesen, M., and Wollenberger, A. (1952). Stepwise degradation of the peptides liberated in the transformation of ovalbumin to plakalbumin. *Nature* **170**, 801-802.

Pace, C.N., and Barrett, A.J. (1984). Kinetics of tryptic hydrolysis of the arginine-valine bond in folded and unfolded ribonuclease T1. *Biochem J* **219**, 411-417.

Park, C., and Marqusee, S. (2004). Probing the high energy states in proteins by proteolysis. *J Mol Biol* **343**, 1467-1476.

Park, C., Zhou, S., Gilmore, J., and Marqusee, S. (2007). Energetics-based protein profiling on a proteomic scale: Identification of proteins resistant to proteolysis. *J Mol Biol* **368**, 1426-1437.

Parsell, D.A., and Sauer, R.T. (1989). The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. *J Biol Chem* **264**, 7590-7595.

Richards, F.M. (1958). On the enzymic activity of subtilisin-modified ribonuclease. *Proc Natl Acad Sci USA* **44**, 162-166.

Rupley, J.A. (1967). Susceptiblity to attack by proteolytic enzymes. *Methods Enzymol* **11**, 905-917.

Sanchez-Ruiz, J.M. (2010). Protein kinetic stability. *Biophys Chem* **148**, 1-15.

Sawaya, M.R., and Kraut, J. (1997). Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: Crystallographic evidence. *Biochemistry* **36**, 586-603.

Silverman, J.A., and Harbury, P.B. (2002). Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J Biol Chem* **277**, 30968-30975.

Tanford, C., and Reynolds, J. (2001). Nature's Robots: A History of Proteins (New York: Oxford University Press).

Telmer, P.G., and Shilton, B.H. (2003). Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. *J Biol Chem* **278**, 34555-34567.

Tiselius, A., and Eriksson-Quensel, I.B. (1939). Some observations on peptic digestion of egg albumin. *Biochem J* **33**, 1752-1756.

Tomkiewicz, D., Nouwen, N., and Driessen, A.J. (2008). Kinetics and energetics of the translocation of maltose binding protein folding mutants. *J Mol Biol* **377**, 83-90.

Tyndall, J.D., Nall, T., and Fairlie, D.P. (2005). Proteases universally recognize beta strands in their active sites. *Chem Rev* **105**, 973-999.

Van Slyke, D.D. (1911). A method for quantitative determination of aliphatic amino groups. Applications to the study of proteolysis and proteolytic products. *J Biol Chem* **9**, 185-204.

Wang, L., and Kallenbach, N.R. (1998). Proteolysis as a measure of the free energy difference between cytochrome c and its derivatives. *Protein Sci* **7**, 2460-2664.

Wildes, D., Anderson, L.M., Sabogal, A., and Marqusee, S. (2006). Native state energetics of the Src SH2 domain: Evidence for a partially structured state in the denatured ensemble. *Protein Sci* **15**, 1769-1779.

Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995). Navigating the folding routes. *Science* **267**, 1619-1620.

Youn, K., and Park, C. (2009). Investigating the effect of temperature on transient partial unfolding by proteolysis. *Protein Pept Lett* **16**, 1093-1097.

