Acta Biochimica Polonica

Vol. 51 No. 2/2004

299–321

QUARTERLY

Review

Probing protein structure by limited proteolysis★

Angelo Fontana✉, Patrizia Polverino de Laureto, Barbara Spolaore, Erica Frare, Paola Picotti and Marcello Zambonin

CRIBI Biotechnology Centre, University of Padua, Padua, Italy

Received: 10 May, 2004; accepted: 12 May, 2004

Key words: limited proteolysis, protein domains, complementing fragments, protein flexibility, molten globule, mass spectrometry, apomyoglobin, lysozyme, α-lactalbumin, cytochrome *c*, human growth hormone

Limited proteolysis experiments can be successfully used to probe conformational features of proteins. In a number of studies it has been demonstrated that the sites of limited proteolysis along the polypeptide chain of a protein are characterized by enhanced backbone flexibility, implying that proteolytic probes can pinpoint the sites of local unfolding in a protein chain. Limited proteolysis was used to analyze the partly folded (molten globule) states of several proteins, such as apomyoglobin, α-lactalbumin, calcium-binding lysozymes, cytochrome *c* and human growth hormone. These proteins were induced to acquire the molten globule state under specific solvent conditions, such as low pH. In general, the protein conformational features deduced from limited proteolysis experiments nicely correlate with those deriving from other biophysical and spectroscopic techniques. Limited proteolysis is also most useful for isolating protein fragments that can fold autonomously and thus behave as protein domains. Moreover, the technique can be used to identify and prepare protein fragments that are able to associate into a native-like and often functional protein complex. Overall, our results underscore the utility of the limited proteolysis approach for unravelling molecular features of proteins and appear to

★Presented as invited lecture at the 29th Congress of the Federation of European Biochemical Societies, Warsaw, Poland, 26 June–1 July 2004.

✉Corresponding author: Prof. Angelo Fontana, CRIBI Biotechnlogy Centre, University of Padua, Viale G. Colombo 3, 35121 Padua, Italy; phone (39 049) 827 6156; fax (39 049) 827 6159; e-mail: angelo.fontana@unipd.it

Abbreviations: AFU, autonomous folding unit; apoMb, apomyoglobin; CD, circular dichroism; CYT, cytochrome *c*; E:S, enzyme to substrate ratio; HD, hydrogen/deuterium; hGH, human growth hormone; LA, α-lactalbumin; LYS, lysozyme; MG, molten globule; NMR, nuclear magnetic resonance; 3D, three-dimensional; nicked protein, a protein with a single peptide bond cleaved.

prompt its systematic use as a simple first step in the elucidation of structure-dynamics-function relationships of a novel and rare protein, especially if available in minute amounts.

A variety of spectroscopic techniques are available for monitoring conformational transitions of proteins in solution, the most commonly used being circular dichroism (CD). Far-UV CD measurements allow to evaluate the overall features of the secondary structure of proteins, as well as to quantify the relative proportions of α-helix, β-sheet and random coil. Tertiary structure and the microenvironment of aromatic chromophores in a protein can be monitored by near-UV CD, fluorescence emission and differential absorption in the UV region. Also infrared and Raman spectroscopy can be used to estimate conformational features and transitions of proteins. Nowadays, the most useful and informative technique for the structural elucidation of (small) proteins in solution is NMR spectroscopy. However, usually NMR remains plagued by heavy instrumentation requirements, needs a millimolar concentration of a non-aggregating protein solution and, in addition, is of limited success in the detailed analysis of partly folded and fluctuating states of proteins due to resonance broadening and/or lack of sufficient chemical shift dispersion. Even X-ray crystallography can be utilized for protein structure analysis only if suitable protein crystals are available and, moreover, it is of no use with dynamic protein systems. Therefore, no one technique is fully superior to all others and each one has advantages and drawbacks.

In this review we will summarise the results of our studies aimed to use limited proteolysis experiments for probing molecular features of proteins in their native or partly folded states. The key features of the site(s) of protein cleavage by proteolytic probes are discussed. Even the aggregation process of proteins to protofibrils and amyloid deposits can be monitored by this technique. Moreover, protein fragments that can fold autonomously and/or can associate to form native-like nicked protein species have been prepared by limited proteolysis of several proteins. It is shown that the simple biochemical technique of limited proteolysis can provide important information on the structure and dynamics of proteins, thus complementing the results that can be obtained by using other biophysical and spectroscopic techniques. We have not attempted here to cover the vast amount of literature dealing with proteolysis of proteins, so that the reader likely will find a personal selection of subjects and omissions as well. Rather we will present here what we consider the most significant results of our studies conducted over the past two decades. The reader may find additional information in our previous review articles dealing with the limited proteolysis phenomenon (Fontana, 1989; Fontana *et al.*, 1989; 1993; 1997a; 1999).

THE RATIONALE OF THE TECHNIQUE

Proteolysis of a protein substrate can occur only if the polypeptide chain can bind and adapt to the specific stereochemistry of the protease's active site (Schechter & Berger, 1967). However, since the active sites of proteases have not been designed by nature to fit the specific sequence and fixed stereochemistry of a stretch of at least 8–10 amino-acid residues of a particular globular protein, an induced-fit mechanism of adaptation of the protein substrate to the active site of the protease is required for binding and formation of the transition state of the hydrolytic reaction (Herschlag, 1988). Therefore, the native rigid structure of a globular protein cannot act as substrate for a protease, as documented by the fact that folded proteins under physiological conditions are rather resistant to proteoly-

sis. This is no longer the case when the fully unfolded state (U) of a globular protein exists at equilibrium with the native state (N). However, the N ⇔ U equilibrium is much shifted towards the native state under physiological conditions, according to the Boltzmann relationship $\Delta G = -RT\ln[U]/[N]$, where ΔG is 5–15 kcal/mol. Therefore, only a tiny fraction (10^{-6}–10^{-9}) of protein molecules is in the U state that is suitable for proteolysis. Consequently, native globular proteins are rather resistant to proteolytic degradation, as a result of the fact that the N ⇔ U equilibrium actually dictates and regulates the rate of proteolysis.

Nevertheless, as shown in the Scheme of Fig. 1, even native globular proteins can be attacked by a protease and, in a number of cases, it has been shown that the peptide bond fission occurs only at one (or a few) peptide bond(s). This results from the fact that a globular protein is not a static entity as can be inferred by a picture of its crystallographically determined 3D-structure, but instead is a dynamic system capable of fluctuations around its average native state at the level of both side chains and polypeptide backbone. Indeed, crystallographers analyze this protein mobility in terms of *B*-factor for both side chains and C_α-backbone (Frauenfelder *et al.*, 1979; Sternberg *et al.*, 1979; Ringe & Petsko, 1985). The main chain *B*-factor is a measure of average displacements of the polypeptide chain from its native structure, so that it can experience displacements leading to some local unfolding. It can be envisaged that these higher energy, locally unfolded states are those required for a native protein to be attacked by a proteolytic enzyme. Evidence for this mechanism of local unfolding required for limited proteolysis has been provided by demonstrating a close correspondence between sites of limited proteolysis and sites of higher backbone displacements in the 316-residue polypeptide chain of thermolysin (Fontana *et al.*, 1986). It is plausible to suggest that limited proteolysis derives also from the fact that a specific chain segment of the folded protein substrate is sufficiently exposed to bind at the active site of the protease. However, the notion of exposure/protrusion/accessibility is a required property, but clearly not at all sufficient to explain the selective hydrolysis of just one peptide bond, since it is evident that even in a small globular protein there are many exposed sites (the all protein surface) which could be targets of proteolysis. Instead, enhanced chain flexibility (segmental mobility) appears to be the key feature of the site(s) of limited proteolysis (Fontana *et al.*, 1986; 1997a).

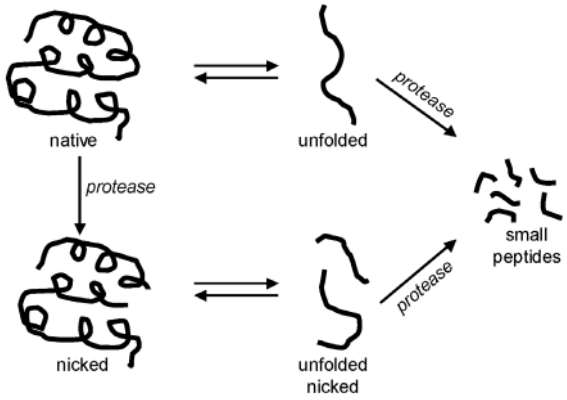

Figure 1. Schematic view of the mechanism of proteolysis of a globular protein.

A dual mechanism of protein degradation is shown, the one involving as substrate for proteolysis the (fully) unfolded protein and the other the native form of the protein. In this last case proteolysis is limited and occurs at flexible site(s), leading to a nicked protein species that can unfold and then be degraded to small peptides.

The results obtained with thermolysin (Fontana *et al.*, 1986) are in line with those derived from limited proteolysis experiments conducted on a variety of other proteins of known three-dimensional (3D) structure (Fontana *et al.*, 1993; 1997a). In many cases, limited proteolysis was observed to occur at sites of the polypeptide chain displaying high segmental mobility or poorly resolved in the electron density map, implying significant static/dynamic disorder. Therefore, it was concluded that limited proteolysis of a globu-

lar protein occurs at flexible loops and, in particular, that chain segments in a regular secondary structure (such as helices) are not sites of limited proteolysis. Indeed, Hubbard *et al.* (1991; 1994) conducted modelling studies of the conformational changes required for proteolytic cleavages and concluded that the sites of limited proteolysis require a large conformational change (local unfolding) of a chain segment of up to 12 residues (Hubbard, 1998). A possible explanation of the fact that helices and, in general, elements of regular secondary structure are not easily hydrolyzed by proteolytic enzymes can be given also on the basis of energetic considerations. If proteolysis is occurring at the centre of the helical segment, likely the helix is fully destroyed by end-effects and consequently all hydrogen bonds, which cooperatively stabilize it, are broken. On the other hand, a peptide bond fission at a disordered flexible site likely does not change much the energetics of that site, since the peptide hydrolysis can easily be compensated by some hydrogen bonds with water (Krokoszynska & Otlewski, 1996). Therefore, it can be proposed that proteolysis of rigid elements of secondary structure is thermodynamically very disadvantageous.

The limited proteolysis approach for probing protein conformation implies that the proteolytic event should be dictated by the stereochemistry and flexibility of the protein substrate and not by the specificity of the attacking protease (Mihalyi, 1978; Price & Johnson, 1990; Fontana *et al.*, 1986; 1993; 1997a; 1997b; 1999; Hubbard *et al.*, 1994; Hubbard, 1998). To this aim, the most suitable proteases are those displaying broad substrate specificity, such as subtilisin, thermolysin, proteinase K and pepsin (Bond, 1990). These endopeptidases display a moderate preference for hydrolysis at hydrophobic or neutral amino-acid residues, but often cleavages occur at other residues as well. The recommended approach is to perform trial experiments of proteolysis of the protein of interest in order to find out the most useful protease, the optimal protein substrate:protease (E:S) ratio and the effect of temperature and time of incubation (Fontana *et al.*, 1999). Possible ways to control proteolysis is by using a low concentration of protease, short reaction times and low temperature. It is not easy to predict in advance the most useful experimental conditions for conducting a limited proteolysis experiment, since these depend upon the structure, dynamics, stability/rigidity properties of the protein substrate and from the actual aim of the experiment, i.e., identification of the sites of protein flexibility, isolation of the rigid core of the protein or preparation of a nicked protein (see below). In typical experiments of limited proteolysis, it has been found that an E:S ratio of 1:100 (by weight) is recommended, but occasionally both 1:20 or 1:5000 can be used. This results from the fact that there is a great variation in the rate of the selective peptide fission in a globular protein, requiring seconds or days for the limited proteolysis event (Mihalyi, 1978). Moreover, if isolation of the nicked protein resulting from the initial proteolysis is desired, both the time and temperature of reaction should be properly controlled, since the nicked species may be present only transiently in the proteolysis mixture. Indeed, a nicked protein is usually much more flexible and unstable than the native one and easily unfolds to a protein substrate that is finally degraded to small peptides (see Fig. 1).

IDENTIFICATION OF FLEXIBLE SITES OF A POLYPEPTIDE CHAIN

Limited proteolysis does not occur at the numerous sites scattered across the protein surface, but is restricted to (very) few specific locations. These sites have shown a good correlation with larger crystallographic *B*-factors, uncertain electron density and larger dispersion values of backbone angles (Fontana *et al.*, 1993; 1999). Hence, limited proteolysis occurs preferentially at those

loops which display inherent conformational flexibility, whereas the protein core remains quite rigid and thus resistant to proteolysis (Fontana, 1989; Fontana *et al.*, 1986; 1989; 1993; 1997a; 1997b; Polverino de Laureto *et al.*, 1995). Usually it is a region, rather than a specific site, the target of limited proteolysis, as given by the fact that, if several proteases are used, one observes that cleavage takes place over a stretch of peptide bonds (Fontana *et al.*, 1997a; 1997b). A survey of the cleavage sites of a variety of proteins of known 3D structure revealed that they never occur at the level of α-helices, but largely at loops (Fontana *et al.*, 1986; 1993; 1999). Therefore, limited proteolysis experiments can be used to identify the sites of enhanced flexibility or of local unfolding of a polypeptide chain (Fontana *et al.*, 1993; 1997a; Hubbard, 1998).

Limited proteolysis experiments were used to probe the structural and dynamic differences between the holo and apo form of horse myoglobin (Mb) (Fontana *et al.*, 1997a). Initial nicking of the polypeptide chain of apoMb (153 amino-acid residues, no disulfide bonds) by several proteases occurs at the level of chain segment 89–96. In contrast, holoMb is resistant to proteolytic digestion when reacted under identical experimental conditions. More recently, the conformational features of native and mutant forms of sperm-whale apoMb at neutral pH were probed by limited proteolysis experiments utilizing up to eight proteases of different substrate specifities (Picotti *et al.*, 2004). It was shown that all proteases selectively cleave apoMb at the level of chain segment 82–94 encompassing helix F in the X-ray structure of the holo form of the native protein; for example, thermolysin cleaves the Pro88–Leu89 peptide bond (Fig. 2). These results indicate that helix F is highly flexible or largely disrupted in apoMb, in full agreement with NMR (Eliezer & Wright, 1996; Lecomte *et al.*, 1996; 1999; Eliezer *et al.*, 1998) and molecular dynamics simulations (Brooks, 1992; Tirado-Rives & Jorgensen, 1993; Hirst & Brooks, 1995; Onufriev *et al.*, 2003).

Since helix F contains the helix-breaking Pro88 residue, it was conceivable to suggest that helix F is kept in place in the native holo protein by a variety of helix-heme stabilizing interactions, including the coordination of the heme iron by proximal His93. In order to modulate the stability of helix F, the Pro88Ala and Pro88Gly mutants of sperm-whale apoMb were prepared by site-directed mutagenesis and their conformational properties investigated by both far-UV CD spectroscopy and limited proteolysis (Picotti *et al.*, 2004). The helix content of the Pro88A mutant was somewhat enhanced with respect to that of both native and Pro88Gly mutant, as expected from the fact that the Ala residue is the strongest helix-inducer among the 20 amino-acid residues. The rate of limited proteolysis of the three apoMb variants by thermolysin and proteinase K was in the order native > Pro88Gly >> Pro88Ala, in agreement with the scale of helix propensity of Ala, Gly and Pro.

The clear-cut results of the proteolysis experiments conducted on apoMb (Fontana *et al.*, 1997; Picotti *et al.*, 2004) emphasize the utility of proteolytic probes of protein structure and dynamics. By using the simple biochemical approach of limited proteolysis it is possible to detect the unfolding of helix F, in agreement with the results obtained in analyzing the molecular features of apoMb at neutral pH by using NMR and computational approaches. Clearly, it is the local mobility/unfolding of the chain of the apoMb substrate that dictates the limited proteolysis phenomenon (Fontana *et al.*, 1986). Indeed, when the mobile chain region encompassing helix F in apoMb is induced to adopt a quite rigid and hydrogen-bonded structure, as that resulting from a Pro→Ala replacement, the site of limited proteolysis was rather well protected against the proteolytic attack (Picotti *et al.*, 2004).

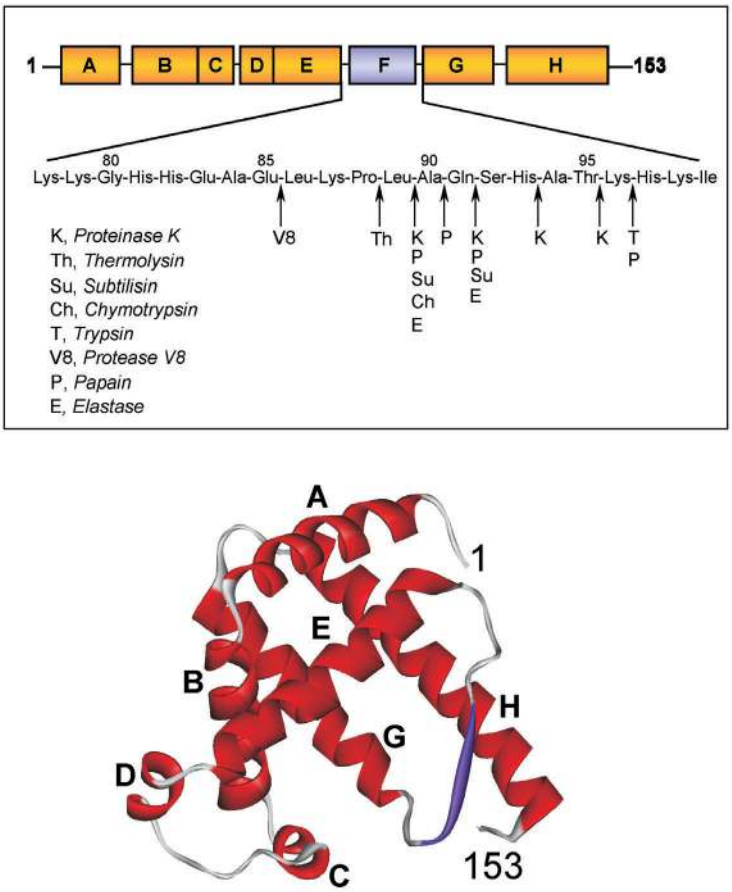

Figure 2. Limited proteolysis of apomyoglobin (apoMb) at neutral pH.

(*Top*) Schematic secondary structure of sperm-whale myoglobin. The height helices A to H are shown in colored boxes. The amino-acid sequence of the chain segment encompassing helix F is magnified and the sites of initial proteolytic cleavage by proteinase K (K), thermolysin (Th), subtilisin (Su), chymotrypsin (Ch), trypsin (T), V8-protease (V8), papain (P) and elastase (E) are indicated by arrows. (*Bottom*) Schematic 3D structure of sperm-whale apoMb. The location of the helical segments (A to H) of the 153-residue chain of holo myoglobin are shown. The chain segment encompassing helix F is shown to be disordered in apoMb (see text). The model was constructed from the X-ray structure (file 1YMB taken from the Brookhaven Protein Data Bank) using the program WebLab (Molecular Simulations Inc., San Diego, CA, U.S.A.).

PROBING PARTLY FOLDED STATES OF PROTEINS

Interpreting function of proteins in terms of their three-dimensional (3D) structure is dominating protein science since many decades. Nevertheless, in recent years the protein structure–function paradigm has been challenged by the observation that proteins can exist in states different from the native (N) and fully unfolded (D) state (Dunker & Obradovic, 2001; Dunker *et al.*, 2002). Indeed, there are a number of experimental observations demonstrating that proteins can adopt intermediate states that play a role in the functioning of proteins at the cellular level, such as ligand binding and protein translocation (Bychkova & Ptitsyn, 1995; Ptitsyn, 1995). Folding intermediates, usually named molten globule (MG) states, can be generated at equilibrium by exposing the protein to mild acid solutions, in the presence of moderate concentrations of protein denaturants, by removing protein-bound ligands (metal ions or prosthetic groups), as well as by chain truncation or amino acid replacements by genetic methods (Ptitsyn, 1995; Arai & Kuwajima, 2000). In some cases, the molecular features

of MGs have been found to resemble those that form transiently in kinetic experiments of protein folding (Chamberlain & Marqusee, 2000) and thus it was proposed that the MG state is an intermediate in protein folding (Ptitsyn, 1987; 1995; Ptitsyn *et al.*, 1990). The MG was defined as a dynamic compact state of a polypeptide chain, characterized by a high degree of native-like secondary structure, but lacking the fixed tertiary contacts of the native state (Ohgushi & Wada, 1983; Ptitsyn, 1987). However, the results of a variety of studies have clearly indicated that a plethora of MGs exists, ranging from those much resembling the unfolded state of a protein to those possessing substantial native-like properties, including some specific tertiary interactions (Arai & Kuwajima, 2000). Moreover, it has been reported that numerous proteins contain largely disordered chain regions of 40 or more amino-acid residues (Romero *et al.*, 2001) and that some proteins appear to be intrinsically or "natively" unfolded or only partly folded under their normal conditions in the cell (Wright & Dyson, 1999; Uversky *et al.*, 2000; Uversky, 2002). Therefore, it has become increasingly clear that proteins are dynamic systems that can adopt a variety of partly folded states or MGs and, in particular, that the function of proteins cannot be interpreted solely on the basis of their static 3D structures (Wright & Dyson, 1999; Dunker & Obradovic, 2002).

It is clear, therefore, that the structural analysis of protein intermediates or MGs is relevant for a number of biophysical and biological aspects of proteins and, in particular, for re-assessing the protein structure–function paradigm (Wright & Dyson, 1999). However, the analysis of the molecular features of MGs is not at all an easy task, since these states usually are flexible and heterogeneous, not amenable to structural elucidation by X-ray crystallography and quite difficult to analyse by NMR spectroscopy (Evans & Radford, 1994). Recent developments in hydrogen/deuterium (HD) exchange, combined with two-dimensional NMR spectroscopy, have provided useful experimental tools for analyzing folding intermediates even at the level of atomic resolution (Eliezer *et al.*, 1998). The advances in NMR measurements include the use of multi-dimensional hetero-nuclear NMR techniques utilizing ^{13}C- or ^{15}N-labeled proteins. Also the solution X-ray scattering technique has been used to obtain accurate data on the size, shape and, in some cases, even tertiary fold of compact, non-native states of proteins (Arai & Kuwajima, 2000). However, it is clear that the structural complexity of MGs requires the use of new and complementary techniques and approaches, since no one technique is fully superior to all others (Evans & Radford, 1994; Dobson, 1994; Fink, 1995; Fontana *et al.*, 1997a).

The conformational state of α-lactalbumin (LA) exposed to acid pH (A-state) is nowadays regarded as a *prototype* MG (Ptitsyn, 1987; 1995; Kuwajima, 1989; 1996; Permyakov & Berliner, 2000; Permyakov *et al.*, 2003). The A-state of LA has been investigated in great detail using a variety of experimental approaches and techniques (Kuwajima, 1996; Arai & Kuwajima, 2000). These studies were conducted on bovine, human, goat and guinea-pig LA, but the conformational features of these homologous proteins are very similar (Pike *et al.*, 1996) and thus results obtained with LA from different sources and their interpretations likely can be used interchangeably. NMR and HD-exchange measurements revealed that in acid the 123-residue chain of LA adopts a partly folded state characterized by a disordered β-domain, while the α-domain maintains substantial, albeit dynamic, helical secondary structure (Wu *et al.*, 1995). The α-domain is a discontinuous domain given by the N-terminal segment 1–37 and the C-terminal segment 85–123 and comprises all helical segments of the protein, while the β-domain is given by the remainder of the protein chain encompassing the β-strands and a coil region (Pike *et al.*, 1996).

Alternatively, the term β-subdomain was used to indicate the chain segment from residue 34 to 57 encompassing the three β-strands of the protein (Polverino de Laureto *et al.*, 1995; 2001; 2002a; 2002b).

Polverino de Laureto *et al.* (1995) were first in using the limited proteolysis approach for unravelling molecular features of MG states of proteins. Proteolysis of LA in its A-state at low pH by pepsin results in the initial cleavage at peptide bonds 52–53 and 40–41 located in the β-subdomain of the protein, thus implying that this region is flexible or unfolded. Subsequently, pepsin cleaves also at peptide bond 103–104 and, therefore, fragments 53–103 and 1–40/104–123 accumulate in the proteolysis mixture. Overall, it was concluded that the proteolytic probe detects, in the A-state of LA, the flexibility or unfolding of the β-subdomain, in full agreement with the results of other physicochemical studies. Indeed, the proposal was advanced that the MG of LA has a "bipartite structure" given by a structured α-domain and a disordered β-domain (Peng & Kim, 1994; Wu *et al.*, 1995; Schulman *et al.*, 1995; 1997).

The conformational features of the calcium-depleted form of LA are not yet defined in such a detail as those of the A-state. Actually, there are discrepancies in the reported experimental results and conflicting proposals regarding the conformational state of apo-LA, ranging from a classical MG devoid of a cooperative thermal transition (Yutani *et al.*, 1992) to a partly folded state with some native-like properties and displaying instead cooperativity (see Kuwajima, 1996, for a discussion and references). The thermal unfolding transition of apo-LA to a MG state was monitored by differential calorimetry, second-derivative UV spectroscopy, fluorescence emission, Raman spectroscopy and near-UV CD. Depending upon the technique and the solvent conditions (pH, ionic strength), different figures for the melting temperature (T_m) of apo-LA have been reported, from about 10°C to 42°C (Griko *et al.*, 1994; Griko & Remeta, 1999; Permyakov & Berliner, 2000). Often it is assumed that apo-LA, as obtained for example by dissolving the protein at 20°C in Tris buffer, pH 8.0, containing a calcium chelating agent, adopts a MG state (Kataoka *et al.*, 1997), but this may be not be true without specifying the ionic strength and the temperature of the protein solution. It is clear, therefore, that the conformational state adopted by apo-LA is strongly influenced by the specific solvent conditions (Griko & Remeta, 1994). Recently, both X-ray crystallography (Chrysina *et al.*, 2000) and NMR spectroscopy (Wijesinha-Bettoni *et al.*, 2001) have been used to analyze the structure of apo-LA in the presence of salt (e.g., 0.5 M NaCl) at neutral pH. These studies revealed that the conformational features of the apo-form of LA are much similar to those of the calcium-loaded protein and that the structural differences between the apo and holo form are mainly confined at the level of the calcium-binding sites.

Limited proteolysis of bovine apo-LA at neutral pH under moderate heating occurs at the same β-subdomain region as in the A-state of LA at low pH (Polverino de Laureto *et al.*, 2002b). All sites of limited proteolytic cleavage by the voracious and non-specific proteinase K (Lebherz *et al.*, 1986) occur at the level of the β-subdomain encompassing the three β-strands S1, S2 and S3 of native LA (Fig. 3). Thus, proteolysis data indicate that the heat-mediated MG of apo-LA at neutral pH shares similar overall conformational and dynamical features to those of the A-state of the protein and, in particular, that the β-subdomain is disordered. Visual inspection of the 3D-model of LA (see Fig. 3) reveals that the β-subdomain is rather well separated from the protein core and thus it can be considered a structural domain or subdomain (Wetlaufer, 1973; 1981). Removal of the β-subdomain from the MG of apo-LA by proteolysis leads to the gapped species desβ-LA, given by fragments 1–34 and 54–123 or 57–123 covalently linked by the disulfide

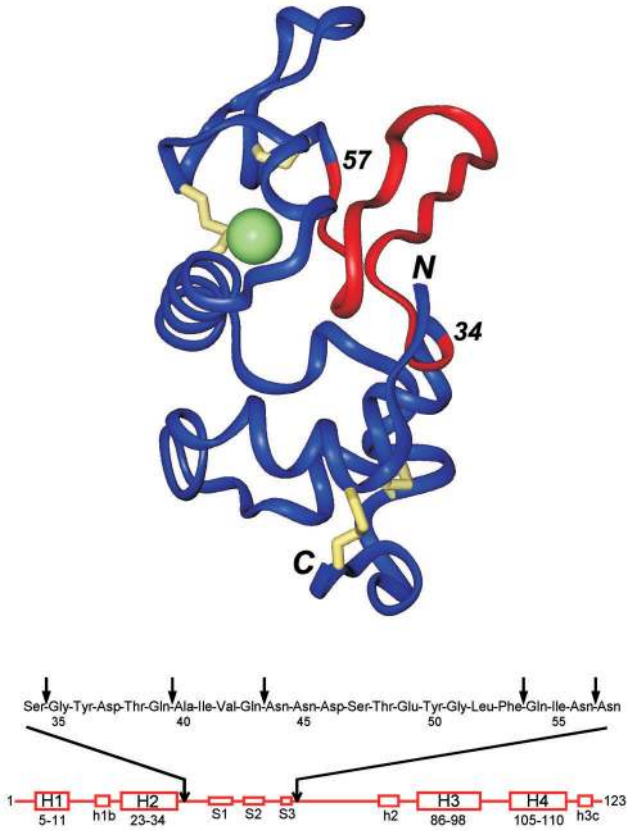

Figure 3. Limited proteolysis of α-lactalbumin (LA) in its partly folded (molten globule) state.

(*Top*) Schematic representation of the crystal structure of bovine LA generated from the coordinates deposited in Brookhaven Protein Data Bank (1hfz) using the program Insight II version 97.0 (Molecular Simulations, San Diego, CA, U.S.A.). The chain segment from residue 35 to 56 encompassing the three-stranded antiparallel β-pleated sheets is indicated in red and the chain segments 1–34 and 57–123 in blue. Calcium is represented by a solid sphere in green. The connectivities of the four disulfides (represented by sticks) along the 123-residue chain of LA are 6–120, 28–111, 61–77 and 73–91. (*Bottom*) Scheme of the secondary structure of the 123-residue chain of bovine LA (Pike *et al.*, 1996). The four α-helices (H1–H4) along the protein chain are indicated by major boxes and below them the corresponding chain segments are given. The three β-strands (S1, 41–44; S2, 47–50; S3, 55–56) are indicated by small boxes. The short 3_{10} helices (h1b, 18–20; h2, 77–80; h3c, 115–118) are also shown by small boxes. At variance from LA derived from other sources, the chain segment encompassing helix H4 (residues 105–110) in the crystal structure of bovine LA exhibits a variety of distinct conformers, including a distorted α-helical conformation (Pike *et al.*, 1996). The amino-acid sequence of the chain region (residues 34–57) of apo-LA is explicitly shown. The sites of peptide bond fission of apo-LA in its MG state at neutral pH in the presence of oleic acid by proteinase K are indicated by arrows (see text).

bridges of the protein. This gapped desβ-LA species is sufficiently stable and rigid to resist further proteolysis, so that, during proteolysis, accumulates in the reaction mixture as the major protein species. Conformational and stability features of desβ-LA have been reported (Polverino de Laureto *et al.*, 2001).

It has been found that LA interacts with a variety of hydrophobic compounds, including fatty acids (Cawthern *et al.*, 1997; Permyakov *et al.*, 2003). A most intriguing observation was reported by Håkansson *et al.* (1995; 1999; 2000) and Svensson *et al.* (1999; 2000). They found a form of LA isolated from the casein fraction of milk that can induce apoptosis in tumor cells, but not in healthy cells. The LA variant inducing apoptosis, yet not thoroughly characterized, appears to require the binding of oleic acid (18:1) as cofactor and to possess spectroscopic properties related to those of the MG state of LA in acid (Svensson *et al.*, 2000). The authors gave to this LA vari-

ant the fancy name HAMLET (human α-lactalbumin made lethal to tumor cells) (Svensson *et al.*, 2000) and found it to possess several interesting biological properties (Köhler *et al.*, 1999; 2001), including a bactericidal activity (Håkansson *et al.*, 2000). It is not clear if HAMLET is monomeric or oligomeric, since Svanborg and co-workers named the oleic acid complex of LA also MAL, i.e., multimeric LA (Köhler *et al.*, 1999; Håkansson *et al.*, 1999; see: also Permyakov *et al.*, 2003).

The simple addition of oleic acid (7.5 equivalents) to a solution of bovine apo-LA at neutral pH induces the formation of the MG state of the protein, as demonstrated by far- and near-UV CD (Polverino de Laureto *et al.*, 2002b). The oleic acid-induced MG behaves towards the proteolytic probe proteinase K in the same way as that of the protein upon moderate heating. Previously, the oleic acid/apo-LA complex or HAMLET has been prepared by a chromatographic procedure (Svensson *et al.*, 2000). The freshly prepared oleic acid/apo-LA complex, obtained by simple mixing the protein with the fatty acid, is monomeric and shows spectroscopic properties identical to those of the MG of LA (Polverino de Laureto *et al.*, 2002b). The protein species desβ-LA, given by the N-terminal fragment 1–34 linked *via* disulfide bridges to the C-terminal fragment 54–123 or 57–123, was isolated from the proteolysis mixture of the oleic acid/LA complex (see Fig. 3). Therefore, seemingly the same MG state of apo-LA can be obtained at neutral pH at 37–45°C or by mixing apo-LA with oleic acid at room temperature. This MG of LA maintains a native-like tertiary fold, characterized by a structured α-domain and a disordered β-subdomain 34–57 (Polverino de Laureto *et al.*, 2002b).

We have analyzed also the partly folded states of protein members of the lysozyme/lactalbumin (LYS/LA) superfamily by CD measurements and limited proteolysis experiments (Polverino de Laureto *et al.*, 2002a). Hen, horse, dog and pigeon lysozyme (LYS) and bovine LA were used for this study. These are related proteins of 123–129 amino-acid residues with similar 3D structures, but notable differences among them reside in their calcium-binding properties and capability to adopt partly folded or MG states in acid solution or upon depletion of calcium at neutral pH (apo-state). Far- and near-UV CD measurements revealed that, while the structures of hen and dog LYS are rather stable in acid at pH 2.0 or at neutral pH in the absence of calcium, conformational transitions to various extents occur with all other LYS/LA proteins. The most significant perturbation of tertiary structure in acid was observed with bovine LA and LYS from horse milk and pigeon egg-white. While hen LYS at pH 2.0 was fully resistant to proteolysis by pepsin, the other members of the LYS/LA superfamily were cleaved at different rates at few sites of the polypeptide chain, thus leading to rather large protein fragments. The apo-form of bovine LA, horse LYS and pigeon LYS were attacked by proteinase K at pH 8.3, while dog and hen LYS were resistant to proteolysis when reacted under identical experimental conditions. Briefly, it has been found that the proteolysis data correlate well with the extent of conformational transitions inferred from CD spectra and with existing structural informations regarding the proteins investigated and mainly derived from NMR and HD-exchange measurements. The sites of initial proteolytic cleavages in the LYS variants occur at the level of the β-subdomain (approximately chain region 34–57), in analogy to those observed with bovine LA (Polverino de Laureto *et al.*, 1995; 2001; 2002b). Overall, proteolysis data indicated that the MG of the LYS/LA proteins is characterized by a structured α-domain and a largely disrupted β-subdomain.

MONITORING PROTEIN AGGREGATION

The problem of protein aggregation was considered in the past just a nuisance in protein research, but recently the analysis of the molecular features of the protein aggregates constituting amyloid fibrils has attracted considerable attention and research effort by numerous investigators. Amyloid fibrils are self-assembled filaments formed by the spontaneous aggregation of a wide variety of peptides and proteins and these aggregates are associated with severe debilitating diseases, such as Alzheimer's disease, type 2 diabetes, prion diseases, Parkinson's disease, senile systemic amyloidosis and Huntington's disease (Sacchettini & Kelly, 2002; Dobson, 2003; Stefani & Dobson, 2003; Selkoe, 2003). The interest of protein scientists arises from fundamental questions regarding the mechanisms by which amyloid fibrils form from monomeric or oligomeric species and the nature of the interactions that make amyloid fibrils a stable structural state for polypeptide chains. Peculiar characteristics of amyloid fibrils include the presence of the cross-β structural motif (Sunde & Blake, 1997) and the unusual resistance to proteolytic degradation. Information about the molecular structures and stabilizing interactions of amyloid fibrils, as well as of the mechanisms of fibril formation, is likely to be useful for the development of therapeutic strategies and drugs for amyloid diseases (Cohen & Kelly, 2003).

The results of a variety of recent studies have indicated that amyloid fibril formation from native proteins occurs *via* a conformational change leading to the formation of partly folded intermediates and subsequent association of these protein species to form pre-fibrillar species given by soluble oligomers that susequently associate into well-ordered mature fibrils (Dobson, 2003). Formation of protein intermediates appear to be critical for the onset of fibril formation, since these species are capable of strong intermolecular interactions due to the exposure of hydrophobic patches otherwise buried in the overall fold of the native protein (Fink, 1998). This view led to the proposal that likely all polypeptide chains can form amyloid aggregates under proper experimental conditions that promote formation of a partly denatured state of the protein chain (Dobson, 1999; 2003).

However, it should be emphasized that a large proportion of physiologically relevant amyloid deposits in the tissue are given by protein fragments deriving from relatively larger proteins (Sacchettini & Kelly, 2002; Dobson, 2003). For example, the Alzheimer's fibrils derive from the fragment(s) produced by limited proteolysis of the amyloid precursor protein (APP). Other examples of aggregating fragments include those of serum amyloid A, gelsolin, apolipoprotein A1, prolactin, amylin, calcitonin, β2-microglobulin, transthyretin, medin, fibrinogen and others. Protein fragments derived by limited proteolysis of proteins are particularly unstable, because usually they can only adopt partly folded states and cannot establish the long-range interactions present in the intact native protein. In particular, a peculiar property of fragments is that they can display hydrophobic patches that can cause protein aggregation (Fink, 1998). Therefore, it seems that protein fragmentation by limited proteolysis can be a causative mechanism of several amyloid diseases.

Considering our research interests in protein folding intermediates, protein fragments and mechanisms of proteolysis, we made an effort to contribute to the problem of protein fibrillogenesis. First of all, we attempted to follow the aggregation phenomenon from protofibrils to mature fibrils of a SH3 domain by using far-UV CD spectroscopy, electron microscopy and limited proteolysis experiments (Polverino de Laureto *et al.*, 2003). The SH3 domains are small protein modules of 60–85 amino-acid residues that are found in many proteins involved in intracellular signal

transduction. The SH3 domain of the p85α subunit of bovine phosphatidyl-inositol 3′-kinase (PI3-SH3) under acidic conditions readily forms amyloid fibrils. The process of PI3-SH3 aggregation at low pH was monitored by using pepsin as proteolytic probe. Remarkably, the protein aggregates that are formed initially display enhanced susceptibility to proteolysis, suggesting that the protein becomes more unfolded and/or flexible in the early stages of aggregation. By contrast, the more defined amyloid fibrils that are formed over longer periods of time are completely resistant to proteolysis. This observation appears to indicate that a rather unfolded/flexible state of the protein is that required to transform protofibrils into the final well-ordered fibrillar structures. Moreover, considering that the initial pre-fibrillar aggregates are substantially more cytotoxic than mature fibrils (Bucciantini *et al.*, 2002), it was proposed that the more unfolded early aggregates could expose specific regions of the protein to the external environment, causing inappropriate binding to membranes and other cellular components and leading to impairment of cellular viability and ultimately cell death.

Hen egg-white lysozyme (LYS) was found to form amyloid fibrils when incubated *in vitro* at pH 2.0 for several days (Krebs *et al.,* 2000). We have analyzed the amyloid fibril formation by hen LYS, since this well-characterized protein appears to be an excellent experimental model to study the determinants of protein aggregation. Amyloid fibrils from LYS, obtained under rather harsh solution conditions by incubating the protein for several days at pH 2.0 and 65°C, are mainly composed of protein fragments, primarily encompassed by chain region 49–101 of the 129-residue chain of the protein (Frare *et al.,* 2004). In order to gain further insights into the aggregation properties of LYS, the propensities of different fragments of the protein to aggregate were examined. Several fragments have been prepared by limited proteolysis of hen LYS by pepsin at pH 0.9 and in the presence of 2 M guanidine hydrochloride. Under these solvent conditions the protein adopts a partly folded state (Sasahara *et al.,* 2000). Fragments 57–107 and 1–38/108–129 were abundant species in the proteolytic mixture, the last fragment being a two chain species constituted by the N-terminal fragment 1–38 and the C-terminal fragment 108–129 covalently linked by the two disulfide bridges Cys6–Cys127 and Cys30–Cys115. The propensity of these fragments to aggregate in solution was studied in detail by CD, thioflavine T binding and electron microscopy. It was shown that fragment 57–107, but not fragment 1–38/108–129, is able to generate well-structured amyloid fibrils when incubated at pH 2.0, 37°C for 2–6 days, i.e., under the same solvent conditions that cause fibrillar assembly of the full-length LYS. These findings indicate that the polypeptide chain encompassing the β-domain and C-helix of LYS is a highly amyloidogenic region of the protein. Of interest, this chain region was shown previously to locally unfold in the amyloidogenic variant D67H of human lysozyme causing systemic amyloidosis and thus populating a partly folded protein species that initiates aggregation events (Canet *et al.,* 2002).

IDENTIFICATION AND PREPARATION OF PROTEIN DOMAINS

Relatively large globular proteins are assemblies of compactly folded substructures, usually called domains or modules. Protein domains are almost invariably seen with proteins made up of more than about 100 amino-acid residues and often their existence can be recognized simply by visual inspection of the 3D model of a protein molecule (Wetlaufer, 1973; 1981). Many large proteins, such as those involved in cell adhesion, clotting, fibrinolysis and signalling, are composed of a series of functionally, distinct, autonomously folding protein domains. Often, identification

of domains or modules in a large protein can be reached simply from sequence analysis of the databases and, in recent years, there has been a rapid advance in identifying families of multi-module proteins, such as the immunoglobulins, fibronectins, kringles, SK2 and SK3 domains, EGF-containing proteins and others (Campbell & Downing, 1994).

Wetlaufer (1973) proposed that protein domains could represent intermediates in the folding process of globular proteins. It is conceivable to suggest that, in a multidomain protein, specific segments of the unfolded polypeptide chain first refold to individual domains and that they subsequently associate and interact with each other to give the final tertiary structure of the protein, much the same as do subunits in oligomeric proteins. The major implication of this hierarchical model of protein folding by a mechanism of modular assembly is that isolated protein fragments corresponding to domains in the intact protein are expected to fold into a native-like structure, thus resembling in their properties a small globular protein. The conformation of the individual domains may be different from that attained in the final native protein, since in the assembled domains some novel mutually stabilizing interactions may be operative. Therefore, a strategy to overcome the complexity of the protein folding problem is to cleave a multi-domain protein into fragments that correspond to domains in the native protein and to study their conformational features in isolation. This protein dissection strategy has been used with several protein systems and in a number of studies it has been demonstrated that some protein fragments can behave as autonomous folding units (AFU) (Wetlaufer, 1981; Wu *et al.*, 1994; Peng & Wu, 2000). However, the problem remains how to choose and prepare suitable protein fragments to be studied for their folding properties.

The current rapid growth in the number of even large protein structures solved by X-ray methods, together with the fact that domains are the constituent building blocks of these proteins, has prompted both theoretical and experimental research on protein domains (Peng & Wu, 2000). Several computer algorithms were developed for the identification of protein domains utilizing the C_α-coordinates of protein structures derived from X-ray analyses (Rose, 1979; Janin & Wodak, 1983, Zehfus, 1987; Siddiqui & Barton, 1995). These algorithms, based on principles such as interface area minimization, plane cutting, clustering, distance mapping, specific volume minimization and compactness, allow a description of globular proteins in terms of a hierarchic architecture given by elements of secondary structure (helices, strands), subdomains (supersecondary structures or folding units), domains and whole protein molecule. Therefore, the concept of protein domains appears to be a convenient way both to simplify the description and classification of protein structures and to study protein folding. In current literature there is no strict, universally accepted definition of a protein domain, but a consensus view of a domain involves a compact, local and independent unit relatively well separated from the rest of the protein molecule.

Limited proteolysis appears to be the best experimental technique for splicing out a fragment that can fold autonomously. The success of the technique resides in the fact that limited (specific) proteolysis of a globular protein occurs at the “hinge” regions or connecting segments between domains. These hinge regions between domains are usually more flexible than the rest of the polypeptide chain forming the globular units of the domains. Thus, a peptide bond fission at the flexible or unfolded chain region likely does not hamper overall structure and stability of the individual protein domains. Indeed, Neurath (1980) emphasized that proteolysis is expected to occur at the flexible hinges between protein domains. Over the years, limited proteolysis was found a most suitable procedure to produce from large proteins in-

dividual, autonomously folding domains for further structural and functional characterization (Wetlaufer, 1981; Neurath, 1980; Peng & Wu, 2000). The most interesting applications of the limited proteolysis approach for producing protein domains were with multifunctional proteins, allowing the isolation of fragments capable of independent folding and displaying some of the activities of the parent protein, thus helping to clarify both structure and mechanism of biological activity of complex proteins. For example, proteolytic cleavage of DNA polymerase leads to a large fragment which catalyzes 3′–5′ exonuclease action on both single stranded and unpaired regions of double-stranded DNA, and to a small fragment which retains only 5′–3′ exonuclease activity (Klenow & Hennigsen, 1970). Other examples are the separation and isolation of kringles from prothrombin, plasminogen, urokinase and plasminogen activator (Patthy *et al.*, 1984) and the isolation of the γ-carboxyl-glutamic acid (Gla) domains from coagulation factors (Esmon *et al.*, 1983). Limited proteolysis has been used also to define the boundaries of a domain by removing its flexible and unstructured parts (Dalzoppo *et al.*, 1985; Darby *et al.*, 1996). Finally, the chain flexibility notion explains the fact that, in certain proteins, the rather loose parts can be removed by limited proteolysis, allowing the isolation of the "core" of the protein as the most proteolysis-resistant moiety and thus as a stable protein entity (Vindigni *et al.*, 1994).

In our laboratory we have conducted systematic conformational studies on protein fragments derived from thermolysin, a two-domain metallo-protease (Vita & Fontana, 1982; Fontana *et al.*, 1983; Vita *et al.*, 1984). It was found that several C-terminal fragments of thermolysin are capable of folding into a native-like structure independently from the rest of the polypeptide chain, thus possessing protein domain properties. These fragments of thermolysin were initially prepared by cyanogen bromide cleavage of the protein at the level of methionine residues in position 120 and 205 of the 316-residue chain of the protein (Vita *et al.*, 1982; 1984). Of course the dimensions of the fragments, for example those of the most studied fragment 206–316, were dictated by the location of the Met residue in the protein chain. With the aim to define the minimum size of a thermolysin C-terminal fragment capable to acquire a stable native-like conformation, Dalzoppo *et al.* (1985) digested the thermolysin fragment 206–316 by means of several proteases. From the kinetics of proteolysis digestion and analysis of the isolated subfragments, it was found that the rather short fragment 255–316 was quite resistant to further proteolysis, implying a tightly folded conformation. Indeed, both CD spectroscopy (Dalzoppo *et al.*, 1985) and, in particular, NMR measurements (Rico *et al.*, 1994) provided a clear-cut evidence of a stable, native-like structure of this small 62-residue fragment.

PREPARATION OF COMPLEMENTING PROTEIN FRAGMENTS

The utility of peptide fragments for analyzing features of protein structure and folding has been recognized long time ago, as given by the pioneering studies on fragments of ribonuclease A and staphylococcal nuclease (Anfinsen & Scheraga, 1975). A specific aim of these early studies was to develop suitable experimental conditions to produce rather long protein fragments by limited proteolysis of the protein and to reconstitute a folded, functionally active "nicked" protein, i.e., the noncovalent complex of (usually two) protein fragments (Taniuchi *et al.*, 1986). Over the years a number of nicked proteins, or complementing fragment systems, has been described and the results of these studies provided information about principles underlying protein structure, folding and dynamics (Fischer & Taniuchi, 1992; dePrat-Gay, 1996).

In general, it was proposed that fragment association proceeds through the formation in the individual fragments of quite unstable "native formats" or folding subdomains that, upon association, acquire a more rigid and stable 3D structure (Anfinsen & Scheraga, 1975; Tsai *et al.*, 1998).

It has been found that functionally active, nicked proteins can be formed by a combination of fragments in such a way that the discontinuity of the polypeptide chain occurs at exposed and flexible sites (usually loops) of the folded protein, outside regions of regular secondary structure (Taniuchi *et al.*, 1986). For this reason, limited proteolysis of globular proteins proved to be the most suitable technique to produce nicked proteins resulting from usually two protein fragments capable of autonomous folding and self association. This success derives from the fact that limited proteolysis occurs at flexible sites along the protein chain and that stable, autonomous folded fragments are more resistant to further proteolytic degradation than relatively unstructured fragments (Fontana *et al.*, 1997a; 1999). Recently, we have applied the limited proteolysis approach to produce sets of complementing fragments from human growth hormone (hGH) (Spolaore *et al.*, 2004), horse apoMb (Musi *et al.*, 2004) and horse cytochrome *c* (CYT) (Spolaore *et al.*, 2001). With these proteins experimental conditions have been devised to cleave them at a single peptide bond, thus producing for each protein a set of two fragments covering the entire protein chain. Besides the identification of the flexible sites along the protein chain in the intact proteins, our studies led to the discovery of three novel systems of complementing fragments. Each of these two-fragment systems (1–88/89–153 of apoMb, 1–44/45–191 of hGH and 1–56/57–104 of CYT) produce a stable protein complex (nicked protein) possessing a native-like 3D structure. Conformational studies conducted by CD spectroscopy on the isolated fragments revealed that they adopt partially folded states and that only upon complementation they acquire the specific long-range interactions that are essential for attaining the structure of the intact protein. Overall, our results indicate that protein fragment complementation is a valuable tool to study the fundamental steps of the folding process, thus simplifying the difficult problem of protein folding (Tsai *et al.*, 1998). Here, we summarize the results of our studies conducted on nicked hGH, apoMb and CYT.

Limited proteolysis of hGH by pepsin at low pH 4.0 occurs at the level of the Phe44–Leu45 peptide bond, leading to the production of fragments 1–44 and 45–191 (Spolaore *et al.*, 2004) (see Fig. 4). Thus, proteolysis data indicate that in acid solution hGH adopts a partly folded state characterized by a local unfolding of the first mini-helix (residues 38–47) encompassing the Phe44–Leu45 peptide bond. Fragment 1–44 was shown to retain little secondary and tertiary structure at neutral pH, while fragment 45–191 independently folds into a highly helical secondary structure. The two peptidic fragments are able to associate into a stable and native-like hGH complex 1–44/45–191. Of interest, hGH has both insulin-like and diabetogenic effects and two fragments of hGH occur *in vivo* and exert these two opposite activities, namely fragment 1–43 showing an insulin-potentiating effect and fragment 44–191 a diabetogenic activity. It was suggested that the conformational changes of hGH induced by an acidic pH promote the generation of the two physiologically relevant fragments by proteolytic processing of the hormone. Likely, limited proteolysis of hGH at low pH is physiologically relevant, since the hormone is exposed to an acidic environment in the cell. The study of Spolaore *et al.* (2004) reports for the first time the analysis of the conformational features of the two individual functional domains of hGH and of their complex.

Proteolysis of the 153-residue chain of horse apoMb by thermolysin results in the selective cleavage of the peptide bond Pro88–Leu89

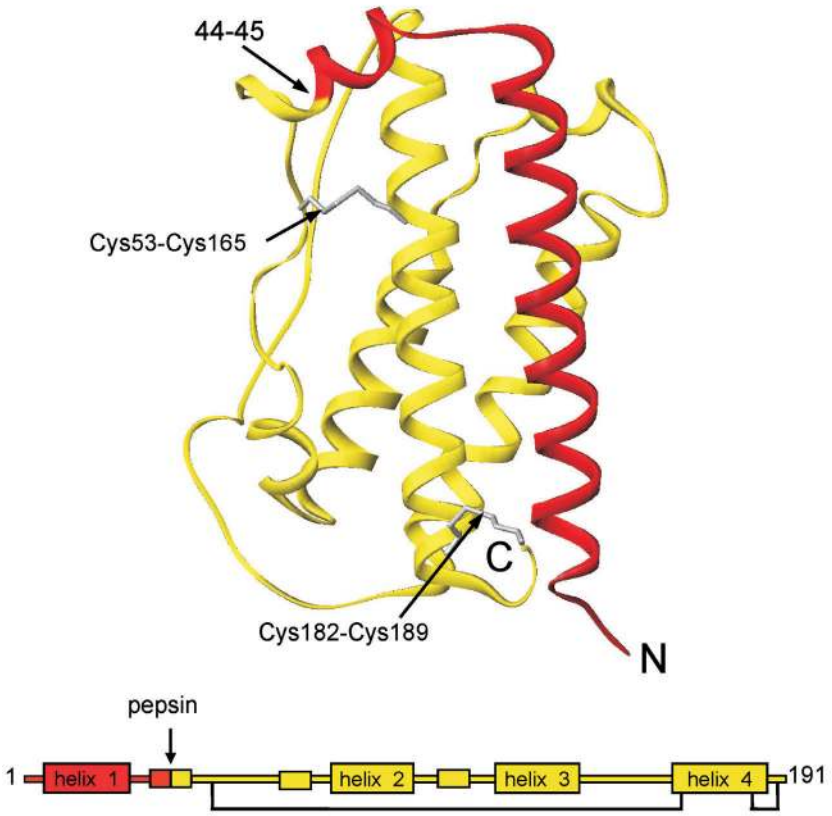

Figure 4. Complementing fragments of human growth hormone (hGH).

(*Top*) Schematic 3D structure of hGH. The four major helices and the three minor helices are shown as ribbons, the remaining residues are represented as a string and the two disulfide bridges (Cys53–Cys165 and Cys182–Cys189) are indicated by grey sticks. The site of the Phe44–Leu45 peptide bond cleavage by pepsin at pH 4.0 is indicated by an arrow. Segments of the 3D structure of hGH corresponding to fragment 1–44 and fragment 45–191 are shown in red and yellow, respectively. The model was constructed from the X-ray structure of hGH (PDB file 3HHR) using the program WebLab Viewer Pro 4.0 (Molecular Simulations Inc., San Diego, CA, U.S.A.). (*Bottom*) Scheme of the secondary structure of hGH. The main boxes indicate the helical segments of the four-helix bundle in hGH, smaller boxes indicate the three short helical segments, whereas disulfide bonds are represented by a solid line. Fragment 1–44 (red) and fragment 45–191 (yellow) are colored as in the 3D model of hGH (*Top*).

(see Fig. 2). The N-terminal (residues 1–88) and C-terminal (residues 89–153) fragments of apoMb were isolated to homogeneity and their conformational and association properties investigated in detail. Far-UV CD measurements revealed that both fragments in isolation acquire a high content of helical secondary structure, while near-UV CD indicated the absence of tertiary structure. A 1:1 mixture of the fragments leads to a tight noncovalent protein complex (1–88/89–153, nicked apoMb), characterized by secondary and tertiary structures similar to those of intact apoMb. The apoMb complex binds heme in a native-like manner, as given by CD measurements in the Soret region. Moreover, in analogy to intact apoMb, the nicked protein binds the hydrophobic dye 1-anilinonaphthalene-8-sulfonate. It was concluded that the two proteolytic fragments 1–88 and 89–153 of apoMb adopt partly folded states characterized by sufficiently native-like conformational features that promote their specific association and mutual stabilization into a nicked protein species much resembling in its structural features intact apoMb. It was suggested that the formation of a noncovalent complex upon fragment complementation can mimic the protein folding process of the entire protein chain, with the

difference that the folding of the complementary fragments 1–88 and 89–153 is an intermolecular process (Tsai *et al.*, 1998). Considering that apoMb has been extensively used as a paradigm in protein folding studies since few decades, the novel fragment complementing system of apoMb appears to be very useful for investigating the initial as well as late events in protein folding.

Limited proteolysis experiments have been used by Spolaore *et al.* (2001) to monitor the folding of a polypeptide chain from a rather unstructured state to a folded, helical state. The N- and C-terminal fragment 1–56 and 57–104, respectively, of horse CYT were used. It was shown that the folding of the polypeptide chain, as given by far-UV CD measurements, resulting from the association of the two fragments into a folded native-like complex, can be monitored by using proteinase K as proteolytic probe. It has been demonstrated that the simple biochemical technique of limited proteolysis can provide useful protein structural data, complementing those obtained by the commonly used CD technique. A specific interest of the complementing fragment system 1–56/57–104 of CYT resides in the fact that the individual fragments correspond exactly to the two exon products of the CYT gene.

CONCLUDING REMARKS

The sites of limited proteolysis (nicksites) in globular proteins of known 3D structure are characterized by enhanced chain flexibility or segmental mobility (Fontana *et al.*, 1986; 1993; 1999). Often, the sites of specific fission are located at regions even devoid of structure, i.e., at disordered protein regions for which no recognizable signal appears in the electron density maps. This is in keeping with the general notion that the protein substrate should suffer considerable structural change in order to properly bind at the precise stereochemistry of the protease's active site in order to form the idealized transition state of the hydrolytic reaction (induced fit mechanism; Herschlag, 1988). Proteolytic enzymes, therefore, can be used to pinpoint, in globular proteins, sites (loops or turns) or regions characterized by local unfolding.

A specific advantage of using proteolytic probes for probing protein structure and dynamics is that they can provide data on the solution structure of a protein, even if these data do not reach the high resolution level given by other physicochemical techniques (NMR, X-ray). The limited proteolysis technique is simple to use and modest in demands for protein sample requirements, instrumentation and experimental efforts. In the past, the technique was somewhat difficult to apply, since the analytical methods required to isolate and characterize protein fragments were labor intensive and not sensitive enough. The present availability of automatic, efficient and highly sensitive techniques of protein sequencing and, in particular, the recent dramatic advances in mass spectrometry in analyzing peptides and proteins, likely will prompt a much systematic use of the limited proteolysis approach as a simple first step in the elucidation of molecular features of a novel and rare protein, especially if available in minute amounts.

We wish to emphasize again that quite often in past and current literature limited proteolysis events are wrongly interpreted in terms of "exposure" of the site(s) of cleavage (Novotny & Bruccoleri, 1987). Of course, the notion of accessibility is a required property of the sites of cleavage in order that the bimolecular reaction between the protease and the protein substrate can take place, but not at all sufficient to explain the *selective* proteolysis of one single peptide bond among hundred(s) bonds, as often observed in limited proteolysis experiments. There are plenty of exposed peptide bonds in a globular protein, but the one that is cleaved should be embedded in a highly flexible or unstructured chain region (Fontana *et al.*, 1986; 1993; 1997a; 1997b;

1999). Therefore, aiming to probe the "surface topography" of a protein by the limited proteolysis approach is simply unfounded. Instead, the approach is eminently suitable to pinpoint in a globular protein the sites of chain flexibility or local unfolding. The correlation between the sites of limited proteolysis and the mobile sites detected by X-ray or NMR methods (Fontana *et al.*, 1986; 1997a; 1997b), as well as by molecular dynamics simulations (Stella *et al.*, 1999; Falconi *et al.*, 2002), has been amply documented. In this respect, we may mention that, in a recent study, disorderd chain regions in protein structures were identified by both proteolysis experiments and predictions of their location along the polypeptide chain by the neural network program PONDR (Prediction of Natural Disordered Regions; Iakoucheva *et al.*, 2001). Predictions nicely correlated with the results of limited proteolysis, thereby indicating that chain disorder or flexibility is the key parameter dictating limited proteolysis events in proteins.

REFERENCES

Anfinsen CB, Scheraga HA. (1975) Experimental and theoretical aspects of protein folding. *Adv Protein Chem.*; **29**: 205–300.

Arai M, Kuwajima K. (2000) Role of the molten globule state in protein folding. *Adv Protein Chem.*; **53**: 209–82.

Bond JS. (1990) Commercially available proteases. In *Proteolytic Enzymes: A Practical Approach.* Beynon RJ, Bond JS, eds, pp 232–40. IRL Press, Oxford.

Brooks CL. (1992) Characterization of "native" apomyoglobin by molecular dynamics simulation. *J Mol Biol.*; **227**: 375–80.

Bucciantini M, Giannoni F, Chiti F, Baroni F, Formigli L, Zurdo J, *et al.* (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature.*; **416**: 507–11.

Bychkova VE, Ptitsyn OB. (1995) Folding intermediates are involved in genetic diseases? *FEBS Lett.*; **359**: 6–8.

Campbell D, Downing AK. (1994) Building protein structure and function from modular units. *Trends Biotechnol.*; **12**: 168–74.

Canet D, Last AM, Tito P, Sunde M, Spencer A, Arche DB, Redfield C, Robinson CV, Dobson CM. (2002) Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nature Struct Biol.*; **9**: 308–14.

Cawthern KM, Narayan M, Chaudhuri D, Permyakov EA, Berliner LJ. (1997) Interactions of α-lactalbumin with fatty acids and spin label analogs. *J Biol Chem.*; **272**: 30812–6.

Chamberlain AK, Marqusee S. (2000) Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Adv Protein Chem.*; **53**: 283–327.

Chrysina ED, Brew K, Acharya KR. (2000) Crystal structures of apo- and holo-bovine α-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions. *J Biol Chem.*; **275**: 37021–9.

Cohen FE, Kelly JW. (2003) Therapeutic approaches to protein-misfolding diseases. *Nature.*; **426**: 905–9.

Dalzoppo D, Vita C, Fontana A. (1985) Folding of thermolysin fragments: Identification of the minimum size of a carboxyl-terminal fragment that can fold into a stable native like structure. *J Mol Biol.*; **182**: 331–40.

Darby NJ, Kemmink J, Creighton TE. (1996) Identifying and characterizing a structural domain of protein disulfide isomerase. *Biochemistry.*; **35**: 10517–28.

de Prat-Gay G. (1996) Association of complementary fragments and the elucidation of protein folding pathways. *Protein Eng.*; **9**: 843–7.

Dobson CM. (1994) Solid evidence for molten globules. *Curr Biol.*; **4**: 636–40.

Dobson CM. (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci.*; **24**: 329–32.

Dobson CM. (2003) Protein folding and disease: A view from the first Horizon Symposium. *Nature Rev Drug Discov.*; **2**: 154–60.

Dunker AK, Obradovic Z. (2001) The protein trinity: Linking function and disorder. *Nature Biotechnol.*; **19**: 805–6.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva CM, Obradovic Z. (2002) Intrinsic disorder and protein function. *Biochemistry.*; **41**: 6573–80.

Eliezer D, Wright PE. (1996) Is apomyoglobin a molten globule? Structural characterization by NMR. *J Mol Biol.*; **263**: 531–8.

Eliezer D, Yao J, Dyson HJ, Wright PE. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nature Struct Biol.*; **5**: 148–55.

Esmon NL, DeBault LE, Esmon CT. (1983) Proteolytic formation and properties of gamma-carboxyglutamic acid-domainless protein C. *J Biol Chem.*; **258**: 5548–53.

Evans PA, Radford SE. (1994) Probing the structure of folding intermediates. *Curr Opin Struct Biol.*; **4**: 100–6.

Falconi M, Parrilli L, Battistoni A, Desideri A. (2002) Flexibility in monomeric Cu,Zn superoxide dismutase detected by limited proteolysis and molecular dynamics simulation. *Proteins Struct Funct Genet.*; **47**: 513–20.

Fink AL. (1995) Compact intermediate states in protein folding. *Annu Rev Biomol Struct.*; **24**: 495–522.

Fink AL. (1998) Protein aggregation: Folding intermediates, inclusion bodies and amyloid. *Folding Des.*; **3**: 9–23.

Fisher A, Taniuchi H. (1992) A study of core domains, and the core domain-domain interaction of cytochrome c fragment complex. *Arch Biochem Biophys.*; **296**: 1–16.

Fontana A. (1989) Limited proteolysis of globular proteins occur at exposed and flexible loops. In *Highlights of Modern Biochemistry.* Kotyk A, Skoda J, Pacek V, Kostka W, eds, pp 1711–26. VSP Int Publ, Amsterdam.

Fontana A, Vita C, Chaiken IM. (1983) Domain characteristics of the carboxyl-terminal fragment 206–316 of thermolysin. *Biopolymers.*; **22**: 69–78.

Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, Zambonin M. (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry.*; **25**: 1847–51.

Fontana A, Vita C, Dalzoppo D, Zambonin M. (1989) Limited proteolysis as a tool to detect structure and dynamic features of globular proteins: Studies on thermolysin. In *Methods in Protein Sequence Analysis.* Wittman-Liebold B, ed, pp 315–24. Springer-Verlag, Berlin.

Fontana A, Polverino de Laureto P, De Filippis V. (1993) Molecular aspects of proteolysis of globular proteins. In *Protein Stability and Stabilization.* van den Tweel W, Harder A, Buitelear M, eds, pp 101–10. Elsevier Science Publ, Amsterdam.

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

Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. (1997b) Probing the conformational state of apomyoglobin by limited proteolysis. *J Mol Biol.*; **266**: 223–30.

Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. (1999) Limited proteolysis in the study of protein conformation. In *Proteolytic Enzymes: Tools and Targets.* Sterchi EE, Stöcker W, eds, pp 257–84. Springer Verlag, Heidelberg.

Frare E, Polverino de Laureto P, Zurdo J, Dobson CM, Fontana A. (2004) A highly amyloidogenic region of hen lysozyme. *J Mol Biol.*; in press.

Frauenfelder H, Petsko GA, Tsernoglou D. (1979) Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature.*; **280**: 558–63.

Griko YV, Remeta DP. (1999) Energetics of solvent and ligand-induced conformational changes in α-lactalbumin. *Protein Sci.*; **8**: 554–61.

Griko YV, Freire E, Privalov PL. (1994) Energetics of α-lactalbumin states: A calorimetric and statistical thermodynamics study. *Biochemistry.*; **33**: 1889–901.

Håkansson A, Zhivotovsky B, Orrenius S, Sabharwal H, Svanborg C. (1995) Apoptosis induced by a human milk protein. *Proc Natl Acad Sci USA.*; **92**: 8064–8.

Håkansson A, Andreasson J, Zhivotovsky B, Karpman D, Orrenius S, Svanborg C. (1999) Multimeric α-lactalbumin from human milk induces apoptosis through a direct effect on cell nuclei. *Exp Cell Res.*; **246**: 451–60.

Håkansson A, Svensson M, Mossberg AK, Sabharwal H, Linse S, Lazon I, Lönnerdal B, Svanborg C. (2000) A folding variant of α-lactalbumin with bactericidal activity against *Streptococcus pneumoniae*. *Mol Microbiol.*; **35**: 589–600.

Herschlag D. (1988) The role of induced fit and conformational changes of enzymes in specificity and catalysis. *Bioorganic Chem.*; **16**: 62–96.

Hirst JD, Brooks CL. (1995) Molecular dynamics simulations of isolated helices of myoglobin. *Biochemistry.*; **34**: 7614–21.

Hubbard SJ, Campbell SF, Thornton JM. (1991) Molecular recognition: conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J Mol Biol.*; **220**: 507–30.

Hubbard SJ. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta.*; **1382**: 191–206.

Hubbard SJ, Eisenmenger F, Thornton JM. (1994) Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci.*; **3**: 757–68.

Iakoucheva LM., Kimzey AL, Masslon CD, Bruce JE, Garner EC, Brown CJ, Dunker AK, Smith RD, Ackerman EJ. (2001) Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci.*; **10**: 560–71.

Janin J, Wodak SJ. (1983) Structural domains in proteins and their role in the dynamics of protein function. *Prog Biophys Mol Biol.*; **42**: 21–78.

Kataoka M, Kuwajima K, Tokunaga F, Goto Y. (1997) Structural characterization of the molten globule of α-lactalbumin by solution X-ray scattering. *Protein Sci.*; **6**: 422–30.

Klenow H, Henningsen I. (1970) Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proc Natl Acad Sci USA.*; **65**: 168–75.

Köhler C, Håkansson A, Svanborg C, Orrenius S, Zhivotovsky B. (1999) Protease activation in apoptosis induced by MAL. *Exp Cell Res.*; **249**: 260–8.

Köhler C, Gogradze V, Håkansson A, Svanborg C, Orrenius S, Zhivotovsky B. (2001) A folding variant of human α-lactalbumin induces mitochondrial permeability transition in isolated mitochondria. *Eur J Biochem.*; **268**: 186–91.

Krebs MR, Wilkins DK, Chung EW, Pitkeathly MC, Chamberlain AK, Zurdo J, Robinson CV, Dobson CM. (2000) Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the beta-domain. *J Mol Biol.*; **300**: 541–9.

Krokoszynska I, Otlewski J. (1996) Thermodynamic stability effects of single peptide bond hydrolysis in protein inhibitors of serine proteases. *J Mol Biol.*; **256**: 793–802.

Kuwajima K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular protein structure. *Proteins Struct Funct Genet.*; **6**: 87–103.

Kuwajima K. (1996) The molten globule state of α-lactalbumin. *FASEB J.*; **10**: 102–9.

Lebherz HG, Burke T, Shackelford JE, Strickler JE, Wilson KJ. (1986) Specific proteolytic modification of creatine kinase isoenzymes. Implication of C-terminal involvement in enzymic activity but not in subunit-subunit recognition. *Biochem J.*; **233**: 51–6.

Lecomte JTJ, Kao YH, Cocco MJ. (1996) The native state of apomyoglobin described by proton NMR spectroscopy: The A-B-G-H interface of wild-type sperm whale apomyoglobin. *Proteins Struct Funct Genet.*; **25**: 267–85.

Lecomte JTJ, Sukits SF, Bhattacharya S, Falzone CJ. (1999) Conformational properties of native sperm whale apomyoglobin in solution. *Protein Sci.*; **8**: 1484–91.

Mihalyi E. (1978) *Application of Proteolytic Enzymes to Protein Structure Studies*. CRC Press, Boca Raton, FL.

Musi V, Spolaore B, Picotti P, Zambonin M, De Filippis V, Fontana A. (2004) Nicked apomyoglobin: A noncovalent complex of two polypeptide fragments comprising the entire protein chain. *Biochemistry.*; **43**: 6230–40.

Neurath H. (1980) Limited proteolysis, protein folding and physiological regulation. In *Protein Folding*. Jaenicke R, ed, pp 501–4. Elsevier/North Holland Biomedical Press, Amsterdam-New York.

Novotny J, Bruccoleri RE. (1987) Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Lett.*; **211**: 185–9.

Ohgushi M, Wada A. (1983) Molten globule state: A compact form of globular proteins with mobile side-chain. *FEBS Lett.*; **164**: 21–4.

Onufriev A, Case DA, Bashford D. (2003) Structural details, pathways and energetics of unfolding apomyoglobin. *J Mol Biol.*; **325**: 555–67.

Patthy L, Trexler M, Vali Z, Banyai L, Varadi A. (1984) Kringles, modules specialized for protein binding: Homology of the gelatin-binding region of fibronectin with the kringle structures of proteases. *FEBS Lett.*; **171**: 131–6.

Peng ZY, Kim PS. (1994) A protein dissection study of a molten globule. *Biochemistry.*; **33**: 2136–41.

Peng ZY, Wu LC. (2000) Autonomous protein folding units. *Adv Protein Chem.*; **53**: 1–47.

Permyakov EA, Berliner LJ. (2000) α-Lactalbumin: structure and function. *FEBS Lett.*; **473**: 269–74.

Permyakov EA, Permyakov SE, Berliner LJ. (2003) α-Lactalbumin: Ca^{2+}-binding protein with multiple functions. In *Protein Structures: Kaleidoscope of Structural Properties and Functions*. Uversky VN, ed, pp 475–97. Research Signpost, Kerala, India.

Picotti P, Marabotti A, Negro A, Musi V, Spolaore B, Zambonin M, Fontana A. (2004) Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements. *Protein Sci.*; **13**: 1572–85.

Pike ACW, Brew K, Acharya KR. (1996) Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure.*; **4**: 691–703.

Polverino de Laureto P, De Filippis V, Di Bello M, Zambonin M, Fontana A. (1995) Probing the molten globule state of α-lactalbumin by limited proteolysis. *Biochemistry.*; **34**: 12596–604.

Polverino de Laureto P, Vinante D, Scaramella E, Frare E, Fontana A. (2001) Stepwise proteolytic removal of the beta subdomain in α-lactalbumin: The protein remains folded and can form the molten globule in acid solution. *Eur J Biochem.*; **268**: 4324–33.

Polverino de Laureto P, Frare E, Gottardo R, van Dael H, Fontana A. (2002a) Partly folded states of members of the lysozyme/lactalbumin superfamily: A comparative study by circular dichroism spectroscopy and limited proteolysis. *Protein Sci.*; **11**: 2932–46.

Polverino de Laureto P, Frare E, Gottardo R, Fontana A. (2002b) Molten globule of bovine α-lactalbumin at neutral pH induced by heat, trifluoroethanol and oleic acid: a comparative analysis by circular dichroism and limited proteolysis. *Proteins Struct Funct Genet.*; **49**: 385–97.

Polverino de Laureto P, Taddei N, Frare E, Capanni C, Costantini S, Zurdo J, Chiti F, Dobson CM, Fontana A. (2003) Protein ag-

gregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *J Mol Biol.*; **334**: 129–41.

Price NC, Johnson CM. (1990) Proteinases as probes of conformation of soluble proteins. In *Proteolytic Enzymes: A Practical Approach.* Beynon RJ, Bond, JS, eds, pp 163–80. IRL Press, Oxford.

Ptitsyn OB. (1987) Protein folding: hypotheses and experiments. *J Protein Chem.*; **6**: 273–93.

Ptitsyn OB. (1995) Molten globule and protein folding. *Adv Protein Chem.*; **47**: 83–229.

Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulgaev OI. (1990) Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett.*; **262**: 20–4.

Rico M, Jimenez MA, Gonzalez C, De Filippis V, Fontana A. (1994) NMR solution structure of the C-terminal fragment 255–316 of thermolysin: a dimer formed by subunits having the native structure. *Biochemistry.*; **33**: 14834–47.

Ringe D, Petsko GA. (1985) Mapping protein dynamics by X-ray diffraction. *Prog Biophys Mol Biol.*; **45**: 197–235.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. (2001) Sequence complexity of disordered proteins. *Proteins.*; **42**: 38–48.

Rose GD. (1979) Hierarchic organization of domains in globular proteins. *J Mol Biol.*; **134**: 447–70.

Sacchettini JC, Kelly JW. (2002) Therapeutic strategies for human amyloid diseases. *Nat Rev Drug Discov.*; **4**: 267–75.

Sasahara K, Demura M, Nitta K. (2000) Partially unfolded equilibrium state of hen lysozyme studied by circular dichroism spectroscopy. *Biochemistry.*; **39**: 6475–82.

Schechter I, Berger A. (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun.*; **27**: 157–62.

Schulman BA, Redfield C, Peng ZY, Dobson CM, Kim PS. (1995) Different subdomains are most protected from hydrogen exchange in molten globule and native states of human α-lactalbumin. *J Mol Biol.*; **253**: 651–7.

Schulman BA, Kim PS, Dobson CM, Redfield C. (1997) A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat Struct Biol.*; **4**: 630–4.

Selkoe DJ. (2003) Folding proteins in fatal ways. *Nature.*; **426**: 900–4.

Siddiqui AS, Barton GJ. (1995) Continuous and discontinuous domains: An algorithm for the automatic generation of reliable protein domain definitions. *Protein Sci.*; **4**: 872–84.

Spolaore B, Bermejo R, Zambonin M, Fontana A. (2001) Protein interactions leading to conformational changes monitored by limited proteolysis: Apo form and fragments of horse cytochrome *c*. *Biochemistry.*; **40**: 9460–8.

Spolaore B, Polverino de Laureto P, Zambonin M, Fontana A. (2004) Limited proteolysis of human growth hormone at low pH: Isolation, characterization and complementation of the two biologically relevant fragments 1–44 and 45–191. *Biochemistry.*; **43**: 6576–86.

Stefani M, Dobson CM. (2003) Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *J Mol Med.*; **81**: 678–99.

Stella L, Di Iorio EE, Nicotra M, Ricci G. (1999) Molecular dynamics simulations of human glutathione transferase P1-1: Conformational fluctuations of the apo structure. *Proteins Struct Funct Genet.*; **37**: 10–9.

Sternberg MJE, Grace DEP, Phillips DC. (1979) Dynamic information from protein crystallography: An analysis of temperature factors from refinement of the hen egg-white lysozyme. *J Mol Biol.*; **130**: 231–53.

Sunde M, Blake C. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffration. *Adv Protein Chem.*; **50**: 123–59.

Svensson M, Sabharwal H, Håkansson A, Mossberg AK, Lipniunas P, Leffler H, Svanborg C, Linse S. (1999) Molecular characterization of α-lactalbumin folding variants

that induce apoptosis in tumor cells. *J Biol Chem.*; **274**: 6388–96.

Svensson M, Håkansson A, Mossberg AK, Linse S, Svanborg C. (2000) Conversion of α-lactalbumin to a protein inducing apoptosis. *Proc Natl Acad Sci USA.*; **97**: 4221–6.

Taniuchi H, Parr GR, Juillerat MA. (1986) Complementation in folding and fragment exchange. *Methods Enzymol.*; **131**: 185–217.

Tirado-Rives J, Jorgensen WL. (1993) Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry.*; **32**: 4175–84.

Tsai CJ, Xu D, Nussinov A. (1998) Protein folding *via* binding and *viceversa*. *Folding Des.*; **3**: R71–80.

Uversky VN. (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.*; **11**: 739–56.

Uversky VN, Gillespie JR, Fink AL. (2000) Why are "natively unfolded" proteins unstructured under physiological conditions? *Proteins.*; **41**: 415–27.

Vindigni A, De Filippis V, Zanotti G, Visco C, Orsini G, Fontana A. (1994) Probing the structure of hirudin from *Hirudinaria manillensis* by limited proteolysis: Isolation, characterization and thrombin-inhibitory properties of N-terminal fragments. *Eur J Biochem.*; **226**: 323–33.

Vita C, Fontana A. (1982) Domain characteristics of the carboxyl-terminal fragment 206–316 of thermolysin: unfolding thermodynamics. *Biochemistry.*; **21**: 5196–202.

Vita C, Dalzoppo D, Fontana A. (1984) Independent folding of the carboxyl-terminal fragment 228–316 of thermolysin. *Biochemistry.*; **23**: 5512–9.

Wetlaufer DB. (1973) Nucleation, rapid folding and globular intrachain regions in proteins. *Proc Natl Acad Sci USA.*; **70**: 697–701.

Wetlaufer DB. (1981) Folding of protein fragments. *Adv Protein Chem.*; **34**: 61–92.

Wijesinha-Bettoni R, Dobson CM, Redfield C. (2001) Comparison of the structural and dynamical properties of holo and apo bovine α-lactalbumin by NMR spectroscopy. *J Mol Biol.*; **307**: 885–98.

Wright PE, Dyson HJ. (1999) Intrinsically unstructured proteins: Reassessing the protein structure-function paradigm. *J Mol Biol.*; **203**: 321–31.

Wu LC, Grandori R, Carey J. (1994) Autonomous subdomains in protein folding. *Protein Sci.*; **3**: 369–71.

Wu LC, Peng ZY, Kim PS. (1995) Bipartite structure of the α-lactalbumin molten globule. *Nat Struct Biol.*; **2**: 281–6.

Yutani K, Ogasahara K, Kuwajima K. (1992) Absence of the thermal transition in apo-α-lactalbumin in the molten globule state: A study by differential scanning microcalorimetry. *J Mol Biol.*; **228**: 347–50.

Zehfus MH. (1987) Continuous compact protein domains. *Proteins Struct Funct Genet.*; **2**: 90–110.