





The intramolecular chaperone-mediated protein folding Yu-Jen Chen and Masayori Inouve

Some proteins have evolved to contain a specific sequence as an intramolecular chaperone, which is essential for protein folding but not required for protein function, as it is removed after the protein is folded by autoprocessing or by an exogenous protease. To date, a large number of sequences encoded as N-terminal or C-terminal extensions have been identified to function as intramolecular chaperones. An increasing amount of evidence has revealed that these intramolecular chaperones play an important role in protein folding both *in vivo* and *in vitro*. Here, we summarize recent studies on intramolecular chaperone-assisted protein folding and discuss the mechanisms as to how intramolecular chaperones play roles in protein folding.

Address

Robert Wood Johnson Medical School, Department of Biochemistry, 675 Hoes Lane, Piscataway, NJ 08854-5635, USA

Corresponding author: Inouye, Masayori (inouye@umdnj.edu)

Current Opinion in Structural Biology 2008, 18:1-6

This review comes from a themed issue on Proteins Edited by Janet Smith and Martino Bolognesi

0959-440X/\$ - see front matter Published by Elsevier Ltd.

DOI 10.1016/j.sbi.2008.10.005

Introduction

When Crick first presented the concept of 'Central Dogma' in 1957, it was assumed that the primary sequences of proteins contain all the information required to achieve the functional conformation [1]. However, it was not proved until Anfinsen demonstrated successful refolding of a denatured protein to its functionally active form *in vitro* [2]. It has been known for decades that molecular chaperones are involved in protein folding, which suggests that protein folding is a complicated process and the information contained in the protein to 'retain' its native conformation is not always sufficient enough to guarantee that the protein will efficiently achieve functionally active structure.

The first report that the intramolecular chaperone plays a crucial role in protein folding that leads to functionally active conformation was based on studies of proteases, such as subtilisin [3 $^{\circ}$], α -lytic protease [4], and carboxypeptidase Y [5]. Unlike the molecular chaperone, the

intramolecular chaperone is encoded in the primary sequence of the protein as an N-terminal or a C-terminal sequence extension and is usually termed propeptide or prosequence. Although, it is not part of the functional domain and does not contribute to the protein function, it is essential for the folding of the functional protein [6°]. Indeed, all these intramolecular chaperones are removed upon the completion of protein folding either by autoprocessing in the case of proteases or by an exogenous protease in the case of non-protease proteins. To date, a large number of propeptides from various proteins have been identified to function as an intramolecular chaperone to assist the folding of the respective functional domains [7,8]. On the basis of their roles in protein folding, intramolecular chaperones can be classified into two categories (Figure 1). The type I intramolecular chaperones include those that assist tertiary structure formation and mostly are produced as the N-terminal sequence extension, and the type II intramolecular chaperones are those that are not directly involved in tertiary structure formation but guide the assembly of quaternary structure to form the functional protein complex and are mostly located at the C-terminus of the protein.

Type I intramolecular chaperones

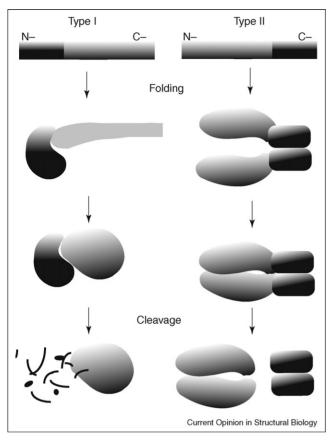
The first discovery of an intramolecular chaperone was based on the studies on subtilisin, an alkaline serine protease from *Bacillus subtilis* [3°]. An interesting aspect of the intramolecular chaperone is that one can study the molecular mechanism of protein folding by introducing amino acid substitution mutations in the propeptide region but not in the functional domain of the protein [9]. Indeed, a number of such mutations have been introduced in the subtilisin propeptide and their effect on protein folding has been investigated [10,11°].

In another approach by introducing a series of mutations, the energy barrier of transition state in subtilisin was reduced allowing it to fold without the intramolecular chaperone, although at a slower rate [12°]. Addition of the propeptide in trans was shown to accelerate the folding of the engineered subtilisin. The same study also showed that a propeptide mutant adopting a more stable structure can assist the folding of the engineered subtilisin at higher efficiency when added in trans [13]. However, active subtilisin was poorly made when folded in cis with the stabilized propertide owing to a slower rate of propeptide processing and degradation [14]. A recent study of POIA1, a protease inhibitor that adopts a propeptide-like but more stable structure, also showed less efficiency in producing active subtilisin in vivo when it is fused to subtilisin [15]. Structural studies of the isolated

Current Opinion in Structural Biology 2008, 18:1-6

2 Proteins

Figure 1

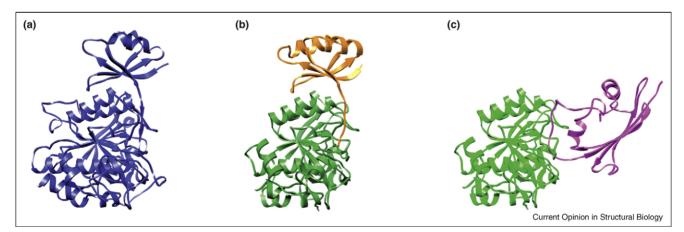


Schematic models for the functions of type I and II intramolecular chaperones. Type I intramolecular chaperones assist the folding of the functional domain to form the native tertiary structure, while type II intramolecular chaperones guide the formation of the quaternary structure of the functional domain.

propeptide of subtilisin revealed that the propeptide is unfolded under physiological conditions [16]. It is important to note that the propeptide inhibits subtilisin function, and therefore it has to be autocatalytically cleaved and degraded by the mature protease in order to produce the active protease. As the propertide is not highly stable, it can be readily degraded upon its cleavage from the functional protein [17]. In addition, previous structural studies have shown that subtilisin interacts with certain inhibitors at positions different as compared with autoprocessed propeptide-subtilisin complex although these inhibitors adopt a structure similar to that of the subtilisin propertide (Figure 2) [18–20]. It is thus speculated that the propertide functions as intramolecular chaperone and inhibitor through different mechanisms.

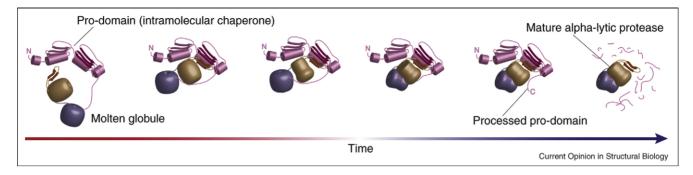
Studies of α -lytic protease suggested that the propertide folds independently from the protease domain and is required at the last stage of protein folding by converting the molten globule state of the protease domain into the active conformation [21**,22]. Interestingly, the two domains of the mature protease were folded independently, but were unable to form native structure without the propeptide. Structural studies have revealed that the N-terminal domain of the propertide interacts with the C-terminal domain of the mature protease. The interaction is mediated through a three-stranded beta sheet in the propeptide structure and a two-stranded beta sheet in the mature protease structure. It was thus suggested that α-lytic protease folds through a nucleation mechanism, in which the propeptide folds first and acts as scaffold, which stabilizes the C-terminal domain of the mature protease allowing the structural rearrangement of the two domains to pack into the native structure (Figure 3) [21**,23-26].

Figure 2



Crystal structures of subtilisins. The unprocessed tk-subtilisin (a subtilisin homolog from hyperthermophilic archaeon, *Thermococcus kodakaraensis*) precursor (PDB code: 2E1P) (a) and processed pro-subtilisin E complex (PDB code: 1SCJ) (b) adopt similar structure. Although the natural inhibitor, *Streptomyces* subtilisin inhibitor (SSI), adopts a structure similar to that of the subtilisin propeptide, it binds to subtilisin at different position (PDB code: 3SIC) (c).

Figure 3



Schematic model of α -lytic protease maturation. The N-terminal propeptide folds into a stable structure, which acts as a scaffold for the protease domain to pack into native structure (modified from [21**]).

Nerve growth factor (NGF) is also produced as proNGF with an N-terminal extension. Both proNGF and NGF are dimeric proteins that carry out different biological functions [27]. In contrast to the propertide of subtilisin, the propeptide of NGF contains limited secondary structure and is removed by the prohormone convertase rather than by autoprocessing. The hydrogen-deuterium exchange experiments and the spectroscopic studies showed that the propertide acts as a competitive inhibitor for the receptor binding of the mature NGF dimer [28,29]. Distinct from subtilisin and α -lytic protease, the NGF propertide is required to be covalently linked to mature NGF [30]. Structural studies showed that the NGF forms a cysteine knot by virtue of three intramolecular disulfide bonds [31]. It was thus suggested that the propeptide assists the folding of NGF via stabilizing a folding intermediate allowing the three disulfide bonds to be properly formed. It is worth to note that the isolated propeptide is monomeric, but both proNGF and mature NGF are dimeric, suggesting that the cleavage of the propeptide occurs after dimerization [28]. It is likely that the quaternary structure may stabilize the tertiary structure; however, the role of the propertide in the quaternary structure remains to be elucidated.

The C-peptide of proinsulin is a unique and intriguing example of intramolecular chaperone, as it is located at the central part of its primary sequence, and is removed by trypsin after disulfide bridges are formed between the A-peptides and B-peptides. The active insulin consists of two polypeptide chains, A and B, linked by two interchain disulfide bonds [32]. The C-peptide is required to ensure the disulfide bridges correctly formed [33]. Isolated C-peptide folds independently from the rest of the protein [34], and acts as a scaffold positioning N-terminal segment (A-peptide) and C-terminal segment (B-peptide) for proper disulfide bond formation. Although the C-peptide has independent physiological activities, such as stimulating Na⁺, K⁺-ATPase [35], which suggest multiple functions, it undoubtedly functions as an intramolecular chaperone for folding of insulin.

Type II intramolecular chaperone

In some proteins, the intramolecular chaperone is not directly involved in folding tertiary structure but instead plays a key role in quaternary structure assembly. For example, the tail spike of E. coli K1-specific bacteriophages, which exists as a homotrimer of endosialidases, is produced with a C-terminal domain (CTD), which is not part of the functional trimer [36]. Notably, when the CTD is removed, the truncated endosialidase becomes insoluble and functionally inactive. Further, it was reported that the CTD folds independently from the enzymatic domain and forms a hexamer [37,38]. These studies suggest that the CTD is able to associate with each other and likely to initiate the trimerization of endosialidases.

Another example is fibril-forming collagen, which contains both an N-terminal and a C-terminal propeptide. The C-terminal propeptide of the collagen prevents premature fibril formation [39]. It was also reported that the C-terminal propeptide of collagen plays a crucial role in the triple helix formation [40]. Recent structural studies suggest that it may be acting as an intramolecular chaperone [41]. Other studies suggested its role in network-forming collagen for superstructure formation. In this case, however, the C-terminal propeptide remains uncleaved and becomes essential for the network structure retention [42]. It was also proposed that the Nterminal propeptide is important in fibril association of the collagen triple helix [40]. Therefore, it is likely that the C-propeptide acts as intramolecular chaperone for collagen quaternary structure formation, while the Nterminal propeptide is required for higher order structure. Recently, the von Willebrand factor (VWF) was reported to contain an N-terminal propeptide, which functions as an intramolecular chaperone [43]. The propeptide of VWF consists of two homologous D domains. The functional VWF consists of a large multimeric protein complex. The propeptide is proteolytically processed in the functional multimer. It was proposed that VWF forms a dimer through an intermolecular disulfide

4 Proteins

bridge at the C-terminal domain followed by multimerization mediated by the N-terminal propeptide [44]. A single mutation at the propeptide (Tyr87Ser) caused VWF to form a dimer with partial function [43,45]. However, it remains to be elucidated if the N-terminal propeptide functions as an intramolecular chaperone for VWF folding.

Discussion

All structural studies, which demonstrate the requirement of the intramolecular chaperones for protein folding to date, have been carried out *in vitro* using entire proteins containing the intramolecular chaperones. However, it is important to note that in the cell, the protein folding is likely to be initiated before its synthesis is complete. Therefore, it remains to be elucidated if the intramolecular chaperone-mediated protein folding studied in vitro is identical to that in vivo. In this regard, it is intriguing to study the folding of premature protein fragments containing the intramolecular chaperone. N-terminal intramolecular chaperone is possibly involved in tertiary structure formation at an early stage of protein folding, through structure nucleation or simply through prevention of misfolding and aggregation. As they assist the tertiary structure formation, most N-terminal intramolecular chaperones are classified into the type I category. However, a C-terminal intramolecular chaperone has been identified in aminopeptidase A (APA) [46] and belongs to the type I category. Since folding of APA involves dimerization, further structural studies are needed to elucidate the precise mechanism of its action. On the contrary, the formation of the higher order structure using type II intramolecular chaperones may be a prerequisite for the completion of the tertiary structure formation. As expected, most type II intramolecular chaperones are found at the C-terminus of the functional domain.

Since various human proteins involved in diseases are found to contain the sequence extensions, which probably function as intramolecular chaperones, it is important to decipher their precise roles in protein folding. In particular, it is important to note that mutations in the intramolecular chaperone can cause misfolding of the functional domain resulting in distortion of their function leading to human diseases, even if the primary structures of the mature functional domains are identical. Such mutations termed protein-memory mutations have been discovered for subtilisin [11°]. More recently, a mutation (Val66Met) in the propertide of brain-derived neurotrophic factor (BDNF) caused loss of memory in humans even if the patients have BDNF, with a primary structure identical to that of the wild-type BDNF [47°]. Such diseases caused by mutations in intramolecular chaperones may be called in general as protein-memory diseases, many of which may be wildly prevailing and not always identified. Existence of various types of intramolecular chaperones, which work through different mechanisms, makes it difficult to propose a single general mechanism governing intramolecular chaperone-mediated protein folding. However, further structural and fundamental studies of individual intramolecular chaperones will provide important clues to our understanding of the general molecular mechanisms of protein folding.

Conclusions

The study of the role of intramolecular chaperones in protein folding is unique and important, as it provides clues for our understanding of not only the basic principle of protein folding but also the etiology of some human diseases caused by protein-memory mutations. Recent advances in technologies for protein structural studies, such as single protein production (SPP) system [48], incell NMR [49,50], and cotranslational structural studies [51**] now open an avenue to study protein structures and folding in the cell. This, in turn, offers great foundation for investigating the mechanisms of intramolecular chaperone-mediated protein folding *in vivo*.

Acknowledgements

We thank Sangita Phadtare for the critical reading. Our work is supported by NIH grant GM056419.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- · of special interest
- •• of outstanding interest
- Crick FH: On protein synthesis. Symp Soc Exp Biol 1958, 12:138-163.
- 2. Anfinsen CB: Principles that govern the folding of protein chains. *Science* 1973, **181**:223-230.
- 3. Ikemura H, Takagi H, Inouye M: Requirement of pro-sequence
- for the production of active subtilisin E in Escherichia coli. J Biol Chem 1987, 262:7859-7864.

The propeptide-mediated protein folding was described for the first time with the use of prosubtilisin.

- Silen JL, Frank D, Fujishige A, Bone R, Agard DA: Analysis of prepro-alpha-lytic protease expression in Escherichia coli reveals that the pro region is required for activity. J Bacteriol 1989, 171:1320-1325.
- Winther JR, Sorensen P: Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity. Proc Natl Acad Sci U S A 1991, 88:9330-9334
- 6. Inouye M: Intramolecular chaperone: the role of the pro-
- peptide in protein folding. *Enzyme* 1991, **45**:314-321. The studies of propeptide-mediated protein folding were reviewed and the term, intramolecular chaperon, was mentioned for the first time and defined.
- Shinde U, Inouye M: Intramolecular chaperones: polypeptide extensions that modulate protein folding. Semin Cell Dev Biol 2000, 11:35-44.
- Eder J, Fersht AR: Pro-sequence-assisted protein folding. Mol Microbiol 1995. 16:609-614.
- Kobayashi T, Inouye M: Functional analysis of the intramolecular chaperone. Mutational hot spots in the subtilisin pro-peptide and a second-site suppressor mutation within the subtilisin molecule. J Mol Biol 1992, 226:931-933.

- 10. Shinde U, Fu X, Inouye M: A pathway for conformational diversity in proteins mediated by intramolecular chaperones. J Biol Chem 1999, 274:15615-15621.
- 11. Shinde UP, Liu JJ, Inouye M: Protein memory through altered folding mediated by intramolecular chaperones. Nature 1997,

The 'protein-memory' phenomenon was discovered for the first time.

12. Almog O, Gallagher T, Tordova M, Hoskins J, Bryan P, Gilliland GL: Crystal structure of calcium-independent subtilisin BPN' with restored thermal stability folded without the prodomain. Proteins 1998, 31:21-32.

By introducing stabilizing mutations and removal of the calcium-binding site, the authors demonstrated successful refolding of subtilisin without the intramolecular chaperone despite the slow folding rate. The structure of the refolded subtilisin variant is shown in this article.

- 13. Ruan B, Hoskins J, Bryan PN: Rapid folding of calcium-free subtilisin by a stabilized pro-domain mutant. Biochemistry 1999. **38**:8562-8571.
- Takagi H, Koga M, Katsurada S, Yabuta Y, Shinde U, Inouye M, Nakamori S: Functional analysis of the propeptides of subtilisin E and aqualysin I as intramolecular chaperones. FEBS Lett 2001, 508:210-214.
- 15. Kojima S, Iwahara A, Yanai H: Inhibitor-assisted refolding of protease: a protease inhibitor as an intramolecular chaperone. FEBS Lett 2005, 579:4430-4436.
- 16. Buevich AV, Shinde UP, Inouye M, Baum J: Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. J Biomol NMR 2001,
- 17. Falzon L, Patel S, Chen YJ, Inouye M: Autotomic behavior of the propeptide in propeptide-mediated folding of prosubtilisin E. J Mol Biol 2007, 366:494-503.
- Tanaka S, Saito K, Chon H, Matsumura H, Koga Y, Takano K, Kanaya S: Crystal structure of unautoprocessed precursor of subtilisin from a hyperthermophilic archaeon: evidence for Ca2+-induced folding. J Biol Chem 2007, 282:8246-8255
- Jain SC, Shinde U, Li Y, Inouye M, Berman HM: The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 Å resolution. J Mol Biol 1998,
- 20. Takeuchi Y, Noguchi S, Satow Y, Kojima S, Kumagai I, Miura K, Nakamura KT, Mitsui Y: **Molecular recognition at the active site** of subtilisin BPN': crystallographic studies using genetically engineered proteinaceous inhibitor SSI (Streptomyces subtilisin inhibitor). Protein Eng 1991, 4:501-508.
- 21. Anderson DE, Peters RJ, Wilk B, Agard DA: alpha-lytic protease precursor: characterization of a structured folding intermediate. Biochemistry 1999, 38:4728-4735.

A model was proposed to depict the molecular mechanism of the intramolecular chaperone-mediated protein folding on the basis of the structural studies on α -lytic protease.

- Cunningham EL, Jaswal SS, Sohl JL, Agard DA: Kinetic stability as a mechanism for protease longevity. Proc Natl Acad Sci U S A 1999, **96**:11008-11014.
- Cunningham EL, Mau T, Truhlar SM, Agard DA: The pro region Nterminal domain provides specific interactions required for catalysis of alpha-lytic protease folding. Biochemistry 2002,
- 24. Cunningham EL, Agard DA: Interdependent folding of the N- and C-terminal domains defines the cooperative folding of alphalytic protease. Biochemistry 2003, 42:13212-13219.
- 25. Truhlar SM, Agard DA: The folding landscape of an alpha-lytic protease variant reveals the role of a conserved beta-hairpin in the development of kinetic stability. Proteins 2005, 61:105-114.
- Sauter NK, Mau T, Rader SD, Agard DA: Structure of alpha-lytic protease complexed with its pro region. Nat Struct Biol 1998, 5:945-950.
- Ibanez CF: Jekyll-Hyde neurotrophins: the story of proNGF. Trends Neurosci 2002, 25:284-286.

- 28. Kliemannel M, Rattenholl A, Golbik R, Balbach J, Lilie H, Rudolph R, Schwarz E: The mature part of proNGF induces the structure of its pro-peptide. FEBS Lett 2004, 566:207-212.
- 29. Nykjaer A, Lee R, Teng KK, Jansen P, Madsen P, Nielsen MS, Jacobsen C, Kliemannel M, Schwarz E, Willnow TE et al.: Sortilin is essential for proNGF-induced neuronal cell death. Nature 2004, **427**:843-848.
- Rattenholl A, Ruoppolo M, Flagiello A, Monti M, Vinci F, Marino G, Lilie H, Schwarz E, Rudolph R: **Pro-sequence assisted folding** and disulfide bond formation of human nerve growth factor. J Mol Biol 2001, 305:523-533.
- 31. McDonald NQ, Lapatto R, Murray-Rust J, Gunning J, Wlodawer A, Blundell TL: **New protein fold revealed by a 2.3-A resolution** crystal structure of nerve growth factor. Nature 1991, **354**:411-414.
- 32. Brems DN, Brown PL, Heckenlaible LA, Frank BH: Equilibrium denaturation of insulin and proinsulin. Biochemistry 1990, 29:9289-9293
- 33. Min CY, Qiao ZS, Feng YM: Unfolding of human proinsulin. Intermediates and possible role of its C-peptide in folding/ unfolding. Eur J Biochem 2004, 271:1737-1747.
- 34. Munte CE, Vilela L, Kalbitzer HR, Garratt RC: Solution structure of human proinsulin C-peptide. FEBS J 2005, 272:4284-4293.
- Ohtomo Y. Bergman T. Johansson BL. Jornvall H. Wahren J: Differential effects of proinsulin C-peptide fragments on Na+, K+-ATPase activity of renal tubule segments. Diabetologia 1998. 41:287-291.
- Stummeyer K, Dickmanns A, Muhlenhoff M, Gerardy-Schahn R, Ficner R: Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. Nat Struct Mol Biol 2005, 12:90-96
- 37. Muhlenhoff M, Stummeyer K, Grove M, Sauerborn M, Gerardy-Schahn R: Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. J Biol Chem 2003,
- 38. Schwarzer D, Stummeyer K, Gerardy-Schahn R, Muhlenhoff M: Characterization of a novel intramolecular chaperone domain conserved in endosialidases and other bacteriophage tail spike and fiber proteins. J Biol Chem 2007, 282:2821-2831.
- 39. Kadler KE, Holmes DF, Trotter JA, Chapman JA: Collagen fibril formation. Biochem J 1996, 316(Pt 1):1-11.
- Hulmes DJ: Building collagen molecules, fibrils, and suprafibrillar structures. J Struct Biol 2002, 137:2-10.
- Khoshnoodi J, Cartailler JP, Alvares K, Veis A, Hudson BG: Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. J Biol Chem 2006, **281**:38117-38121.
- 42. Sundaramoorthy M, Meiyappan M, Todd P, Hudson BG: Crystal structure of NC1 domains. Structural basis for type IV collagen assembly in basement membranes. J Biol Chem 2002, **277**:31142-31153.
- 43. Rosenberg JB, Haberichter SL, Jozwiak MA, Vokac EA, Kroner PA Fahs SA, Kawai Y, Montgomery RR: The role of the D1 domain of the von Willebrand factor propeptide in multimerization of VWF. Blood 2002, 100:1699-1706.
- Voorberg J, Fontijn R, Calafat J, Janssen H, van Mourik JA, Pannekoek H: Assembly and routing of von Willebrand factor variants: the requirements for disulfide-linked dimerization reside within the carboxy-terminal 151 amino acids. J Cell Biol 1991, **113**:195-205.
- 45. Haberichter SL, Fahs SA, Montgomery RR: von Willebrand factor storage and multimerization: 2 independent intracellular processes. Blood 2000, 96:1808-1815.
- Rozenfeld R, Muller L, El Messari S, Llorens-Cortes C: The Cterminal domain of aminopeptidase A is an intramolecular chaperone required for the correct folding, cell surface expression, and activity of this monozinc aminopeptidase. J Biol Chem 2004, 279:43285-43295.

6 Proteins

- 47. Bath KG, Lee FS: Variant BDNF (Val66Met) impact on brain
- structure and function. Cogn Affect Behav Neurosci 2006, 6:79-85. This review article describes the effects of a single mutation in the propeptide domain of brain-derived neurotrophic factor (BDNF) on human brain. Although further structure studies were needed, this may be the first example of protein-memory diseases.
- 48. Suzuki M, Zhang J, Liu M, Woychik NA, Inouye M: Single protein production in living cells facilitated by an mRNA interferase. Mol Cell 2005, **18**:253-261.
- Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dotsch V: **High-resolution macromolecular NMR**

- spectroscopy inside living cells. J Am Chem Soc 2001,
- 50. Burz DS, Shekhtman A: In-cell biochemistry using NMR spectroscopy. PLoS ONE 2008, 3:e2571.
- 51. Hsu ST, Fucini P, Cabrita LD, Launay H, Dobson CM,
 Christodoulou J: Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. Proc Natl Acad Sci USA 2007, 104:16516-16521.

This paper describes the first structural and dynamics study of a newly synthesized polypeptide chain in a cell-free system, which is bound to a