UNIVERSITA' DEGLI STUDI DI PADOVA

Centro di Ricerca Interdipartimentale per le Biotecnologie Innovative

(CRIBI)

Department of Chemistry, University of Cambridge (UK)

SCUOLA DI DOTTORATO DI RICERCA IN BIOCHIMICA E BIOTECNOLOGIE

INDIRIZZO DI BIOTECNOLOGIE

XX CICLO

PROTEIN AMYLOIDOGENESIS

CHARACTERIZATION OF AGGREGATION PRONE CONFORMATIONS AND FIBRILS STRUCTURE

Direttore della Scuola : Ch.mo Prof. Lorenzo Pinna

Supervisore : Dr.ssa Patrizia Polverino de Laureto

Dottoranda : Maria Francesca Mossuto

31 gennaio 2008

CONTENTS

ABBREVIATIONS

ANS	8-aniline-1-naftalen-sulphonate
ATR-FTIR	attenuated total reflectance
CD	circular dichroism
DTT	dithiothreitol
EDTA	ethylen-diammine-tetraacetic acid
EM	electron microscopy
ESI	electrospray ionization
E:S	enzyme to substrate ratio
FTIR	Fourier-transform infrared spectroscopy
GdnHCl	guanidine hydrochloride
GdnSCN	guanidine thiocyanate
HEWL	hen egg-white lysozyme
HPLC	high-performance liquid chromatography
HuL	human lysozyme
LA	alpha-lactalbumin
LYS	human lysozyme
MG	molten globule
MRW	mean residue weight
MS	mass spectrometry
MW	molecular weight
RP	reverse-phase
RT	retention time
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
[θ]	mean residue ellipticity
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
ThT	thioflavin T
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
w/w	weight/weight

AMINO ACIDS ABBREVIATIONS

Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid (Aspartate)
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid (Glutamate)
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine

RIASSUNTO

Da qualche anno c'è un crescente interesse per lo studio delle fibrille amiloidi e questo per un sempre maggiore coinvolgimento in molti settori scientifici. Innanzitutto, le fibrille costituiscono l'elemento distintivo di importanti malattie come il morbo di Alzheimer e di Parkinson. Inoltre, dal momento che è stato dimostrato che pressochè tutte le proteine sono in grado di formare strutture amiloidi in opportune condizioni, la comprensione di come e perchè queste strutture si formino è diventato un problema centrale nella biologia strutturale. Infine, data la loro struttura altamente ordinata e ripetuta, le fibrille amiloidi costituiscono la base ideale per lo sviluppo di nanomateriali con possibili applicazioni tecnologiche. In generale il processo di formazione di fibrille amiloidi da parte di una qualsiasi proteina prevede la parziale perdita di struttura dello stato nativo e la formazione di intermedi strutturali in grado di interagire l'uno con l'altro, organizzandosi dapprima in oligomeri e quindi in fibrille ordinate. Nonostante la capacità di quasi tutte le proteine di formare fibrille amiloidi, ancora molto poco è noto riguardo la loro struttura e i fattori che ne regolano la formazione.

In questo lavoro di Tesi, è stata analizzata la formazione di fibrille amiloidi da parte di proteine globulari, studiando nella prima parte le conformazioni proteiche non native propense ad aggregare, e nella seconda parte caratterizzando più nel dettaglio la struttura delle fibrille amiloidi.

L'analisi degli stati conformazionali propensi all'aggregazione è stata effettuata su due proteine globulari modello, alfa-lattalbumina (LA) e HypF-N. Infatti, è stato precedentemente riportato che in appropriate condizioni sperimentali queste due proteine assumono conformazioni parzialmente strutturate (Uversky, 2002; Chiti et al., 2001), importanti nel processo formazione di fibrille amiloidi.

Lo studio condotto su LA si è basato sull'analisi degli effetti che il taglio proteolitico di questa proteina ha sulle sue caratteristiche conformazionali e proprietà di aggregazione. E' noto che LA è in grado di formare fibrille amiloidi morfologicamente indistinte da quelle patologiche (Uversky et al., 2002). Nel lavoro qui presentato, è stata studiata la propensione ad aggregare di derivati di LA caratterizzati uno da un singolo taglio proteolitico (1-40/41-123, denominato Th1-LA) e l'altro dalla rimozione di un segmento di 12 amino acidi corrispondenti a parte del dominio beta della proteina nativa (1-40/53-123, denominato desbeta-LA). Inoltre, le prime fasi del processo di aggregazione di questi due derivati sono state confrontate con quelle di LA intatta. La conclusione principale di questo lavoro (Polverino de Laureto et al., 2005) è che la

maggiore flessibilità dei due derivati di LA favorisce quei cambiamenti conformazionali necessari per formare la struttura delle fibrille amiloidi. E' stato inoltre ribadito che la proteolisi può essere considerata un meccanismo scatenante l'aggregazione proteica e la fibrillogenesi.

Nel caso di HypF-N, invece, la caratterizzazione della conformazione propensa ad aggregare è stata condotta in ambiente acido, per permettere alla proteina di popolare un'insieme di stati parzialmente strutturati. Usando una combinazione di tecniche biofisiche e biochimiche è stato dimostrato che questa conformazione parzialmente strutturata ha tutte le caratteristiche di uno stato *pre-molten globule*, cioè più compatto di uno stato completamente non strutturato ma meno organizzato di un intermedio *native-like* o di uno stato *molten globule* (Uversky, 2002). Inoltre, è stato evidenziato che modulando la forza ionica della soluzione è possibile aumentare la velocità di aggregazione di HypF-N in queste condizioni (Campioni et al., *submitted*). In aggiunta, è stato dimostrato che l'aggregazione di HypF-N in presenza di alta forza ionica è mediata dalla formazione di oligomeri, in cui le molecole di HypF-N interagiscono mediante una particolare regione della sequenza, corrispondente al segmento proteico che in HypF-N ha sia alta idrofobicità sia la più alta propensione a formare struttura a foglietti beta e che non ha carica a pH acido. Queste tre caratteristiche ne fanno la regione ideale per mediare l'oligomerizzazione proteica.

Nella seconda parte di questa Tesi di dottorato, invece, il lavoro svolto riguarda la caratterizzazione strutturale di fibrille amiloidi. Molti studi sono stati condotti sugli aggregati amiloidi formati in diverse condizioni da peptidi come Aβ e frammenti di transtiretina e prione (Kodali and Wetzel, 2007). Tuttavia, studiare la formazione di fibrille amiloidi da parte di frammenti proteici è molto diverso che studiare lo stesso processo con proteine intatte. Infatti, la struttura fortemente impaccata che si ottiene in fibrille amiloidi formate da frammenti proteici (10-40 residui) non può essere realizzata in una proteina ($\geq$ 50 residui), a eccezione delle regioni che formano il *core* della fibrilla (Chatani and Goto, 2005). Per studiare la struttura di fibrille amiloidi formate da una proteina intatta, è stato scelto il lisozima umano (HuL) innanzitutto per il fatto che alcune delle varianti naturali di HuL sono responsabili della formazione di placche amiloidi in vivo, in una malattia chiamata amiloidosi sistemica non-neuropatica (Pepys et al., 1993; Booth et al., 1997). Inoltre, è stato possibile sfruttare la grande disponibilità di informazioni strutturali e di *folding* riguardanti HuL *wild type*, dal momento che anch'esso è in grado di formare fibrille amiloidi simili a quelle patologiche, a pH acido

ed alta temperatura (Morozova-Roche et al., 2000). Rispetto alle fibrille formate da frammenti proteici, inoltre, studiare fibrille composte da HuL è più complicato dal momento che è una catena polipeptidica di 130 residui con i limiti conformazionali dati dai suoi quattro ponti disolfuro. Negli studi qui riportati, le fibrille di HuL formate a pH acido sono state analizzate mediante esperimenti di proteolisi limitata e spettroscopia infrarossa in trasformata di Fourier (FTIR), allo scopo di caratterizzare dal punto di vista conformazionale la catena di HuL una volta riorganizzata nella fibrilla amiloide (Frare et al., 2006). Dopo la digestione con pepsina a basso pH, è stato evidenziato che le fibrille di HuL erano composte per lo più da molecole di HuL da cui erano state rimosse le estremità C e N terminali, e da un minor numero di molecole di HuL intatto completamente resistenti alla proteolisi. Gli spettri FTIR hanno confermato che le fibrille di HuL contengono un'estesa struttura a foglietto beta e alcuni elementi di struttura non beta e *random* che vengono ridotti vistosamente dopo la proteolisi. La regione di HuL completamente resistente alla proteolisi nella fibrilla corrisponde ai residui 32-108, e comprendente i foglietti beta e l'elica C di HuL nativo, regioni precedentemente dimostrate essere propense a destrutturarsi localmente sia nella molecola *wild type* che nei mutanti patogenici. Inoltre, questa regione, dimostrata essere il *core* della fibrilla, comprende la regione altamente propensa ad aggregare recentemente identificata nel lisozima di pollo (Frare et al., 2004). Questi dati indicano che la regione del HuL che si destruttura e aggrega più prontamente corrisponde alla regione di HuL resistente alla proteolisi e quindi più strutturata nelle fibrille. Infine, questi dati mostrano che la formazione di fibrille amiloidi non richiede tutta la sequenza proteica di HuL.

Le varianti patogeniche di HuL, però, data la loro maggiore instabilità rispetto al HuL *wild type*, aggregano in condizioni fisiologiche, ovvero pH 7.5 e 37°C (Dumoulin et al., 2005). Durante l'ultima parte del dottorato è stato dimostrato che anche il HuL *wild type* è in grado di aggregare in condizioni più simili a quelle patologiche, cioè a pH neutro e 60°C. Dato che HuL forma fibrille amiloidi in condizioni così diverse (pH 2.0 50°C e pH 7.5 60°C), la struttura e la stabilità di questi due tipi di fibrille sono state analizzate e confrontate. Nello studio qui riportato sono state prodotte fibrille di HuL sia a pH acido che neutro, ed entrambe hanno le tre caratteristiche delle fibrille amiloidi, ovvero legano tioflavina T, hanno struttura *cross*-beta e hanno una morfologia fibrillare visibile al microscopio. Queste fibrille sono state studiate mediante fluorescenza, valutando la capacità di interagire con un probe fluorescente (*8-aniline-1-naftalen-*

sulphonate, ANS), spettroscopia FTIR e diffrazione a raggi X per evidenziare le loro possibili differenze strutturali. Sono stati condotti anche studi di stabilità, analizzando i due tipi di fibrilla mediante esperimenti di dissociazione indotta da agenti denaturanti, quali guanidinio. I risultati ottenuti indicano chiaramente che le condizioni usate per l'aggregazione di HuL promuovono la formazione di fibrille con diverse caratteristiche strutturali e diverse proprietà di stabilità, dovute al diverso ripiegamento della catena polipeptidica di HuL nei due tipi di struttura fibrillare.

In conclusione, il lavoro di ricerca condotto durante il dottorato ha permesso la comprensione di importanti aspetti riguardanti le conformazioni assunte da alcune proteine globulari che le rendono propense all'aggregazione. Inoltre, nuove ed importanti informazioni sono state ottenute sulla struttura e sul polimorfismo delle fibrille amiloidi.

Questo lavoro è stato svolto principalmente al CRIBI (Università di Padova), presso il laboratorio di Chimica delle Proteine, collaborando in parte con altri gruppi di ricerca, come il laboratorio del Prof. Chiti presso il Dipartimento di Biochimica dell'Università di Firenze e il laboratorio del Prof. Dobson presso il Dipartimento di Chimica dell'Università di Cambridge, dove è stata condotta la caratterizzazione strutturale delle fibrille di HuL a pH neutro.

SUMMARY

Current interest in studying amyloid fibrils arises from their involvement in different fields (Chiti and Dobson, 2006). First, they play a crucial role in disorders such as Alzheimer's and Parkinson's diseases. Second, since it has been demonstrated that all polypeptide chains form fibrils under appropriate conditions, the understanding of why and how this process happens has become central problem in protein knowledge. Last, the ordered ultrastructure characterizing amyloid fibrils may be thought as a basis for nanomaterials with possible technological applications. However, despite the ability of most proteins to form amyloid fibrils, very little is known about their structures and the factors that govern their formation.

The process of amyloid formation requires the partial unfolding of protein molecules into such conformations able to interact to each others and reorganize into well-ordered structured aggregates, named amyloid fibrils.

In this Thesis the amyloid formation by globular proteins has been analyzed from different points of view, focusing in the first part of the research work on the elucidation of some conformational features promoting protein aggregation. The second part was concentrated on the characterization of the final supramolecular structure of amyloid fibrils.

In order to study the partially folded state, two globular proteins have been analyzed, alpha-lactalbumin (LA) and HypF-N. Indeed, under specific conditions, these proteins populate a not fully folded state previously shown to play an important role in the amyloid formation (Uversky, 2002; Chiti et al., 2001).

The study conducted on LA has been based on the effects of the proteolytic dissection of the molecule on its conformational features and aggregation properties. It was previously shown that LA is able to form fibrils morphologically indistinct from the pathological ones (Uversky et al., 2002). Here, we have studied the aggregation propensities of LA derivatives characterized by a single peptide bond fission (1-40/41-123, named Th1-LA) or a deletion of a chain segment of 12 amino acid residues located at the level of the β-subdomain of the native protein (1-40/53-123, named desβ-LA). We have also compared the early stages of the aggregation process of these LA derivatives with those of intact LA. The main conclusion of this work was that the inherent flexibility of the LA derivatives allows the large conformational changes required to form the cross-β-structure of the amyloid fibrils. It has been emphasized that

proteolysis can be considered a causative mechanism of protein aggregation and fibrillogenesis (Polverino de Laureto et al., 2005).

In the other case, the conformational characterization of an amyloidogenic state of HypF-N has been performed at acid pH, in order to allow the protein to populate a partially unfolded ensemble. Combining different biophysical and biochemical techniques, it has been shown that this partially unfolded structure has all the hallmarks of a pre-molten globule state, i.e. it is more compact than a random coil-like state but less organized than a native-like intermediate or a MG state (Uversky, 2002). Furthermore, it is shown that a modulation of the total ionic strength of the solution allows enhancing the apparent rate of aggregation of HypF-N under these conditions. This increased rate of aggregation has been shown to be mediated by the interaction of monomers to form initial oligomers, through a particular region in the sequence, corresponding to the sequence part having highly hydrophobicity, the highest beta-sheet propensity and with no net charge at acid pH, representing the ideal segment suitable to mediate protein oligomerization.

From all these studies, it is clear that, except the unique native state of globular proteins wherein the side chains pack together in a unique manner, every state of a polypeptide molecule is a broad ensemble of often diverse conformations. It is not surprising, therefore, that even the fibrillar products of aggregation processes are characterized by morphological and structural diversity, representing variations on a common theme.

The second part of my PhD Thesis deals with the structural characterization of fibrils. Many studies have been conducted on amyloid aggregates formed under different conditions by peptides, such as Aβ, TTR and prion fragments (Kodali and Wetzel, 2007). Indeed, the problem of amyloid formation by a full-length protein is more complex, since the dense packing reachable in amyloid fibrils made of peptides (10-40 residues) could not be accomplished in all the amino acid residues of a full-length protein, except in the core regions (Chatani and Goto, 2005). The object of my study was human lysozyme due, most of all, to the fact that some natural variants of human lysozyme (HuL) are responsible for the formation of amyloid plaques *in vivo*, in a so called familial non-neuropathic systemic amyloidosis (Pepys et al., 1993; Booth et al., 1997). Moreover, it has been possible to exploit the available wealth of structural and folding information about wild type HuL, since it has been shown to be able to form fibrils quite similar to the pathological ones, under acidic conditions and high

temperature (Morozova-Roche et al, 2000). With respect to fibrils made of peptides, besides, studying amyloid fibrils conformations from HuL is more challenging because it is a 130 amino acid chain with the structural constrains given by the four disulfide bridges present in the lysozyme molecule. This study can also give some insights into the complex problem of strains diversity, such as for prion diseases, helping the clarification of the structural principles of amyloid fibrils which can produce multiple and distinct amyloid conformations from one protein sequence.

In the presented study, fibrils of wild-type HuL formed at low pH have been analyzed by limited proteolysis experiments and Fourier-transform infrared (FTIR) spectroscopy, in order to map conformational features of the 130 residue chain of lysozyme when embedded in the amyloid aggregates (Frare et al., 2006). After digestion with pepsin at low pH, the lysozyme fibrils were found to be composed primarily of N and C-terminally truncated protein species encompassing residues 26–123 and 32–108, although a minority of molecules was found to be completely resistant to proteolysis under these conditions. FTIR spectra provide evidence that lysozyme fibrils contain extensive β-sheet structure and substantial elements of non β-sheet or random structure that are reduced significantly in the fibrils after digestion. The sequence 32–108 includes the β-sheet and helix C of the native protein, previously found to be prone to unfold locally in human lysozyme and its pathogenic variants. Moreover, this core structure of the lysozyme fibrils encompasses the highly aggregation-prone region of the sequence recently identified in hen lysozyme (Frare et al., 2004). The present data indicate that the region of the lysozyme molecule that unfolds and aggregates most readily corresponds to the most highly protease-resistant and thus highly structured region of the majority of mature amyloid fibrils. Overall, the data show that amyloid formation does not require the participation of the entire lysozyme chain.

HuL variants, however, aggregate in a physiological environment, roughly at pH 7-7.5 at 37 °C, because of their instability (Dumoulin et al., 2005). In my work, it has been demonstrated that also HuL is able to aggregate under conditions similar to the pathological ones, presumably neutral pH and 37 °C. Considering that HuL forms amyloid fibrils in such different conditions (pH 2.0 50°C and pH 7.5, 60°C), a comparison of the structure and the stability of fibrils obtained under these different conditions has been conducted. In this study HuL fibrils were produced at acidic and at neutral pH, leading both to the formation of fibrils having the three hallmarks of amyloid, that are cross-beta structure, binding of ThT and an overall amyloid fiber

morphology. These fibrils have been studied by means of ANS binding, FTIR and X-ray fiber diffraction in order to characterize the differences in the structure. Guanidinium-induced fibrils dissociation, instead, has been applied in order to test the chemical stability of the two kinds of fibrils. The results clearly indicate that the solution conditions used for lysozyme aggregation promote the formation of fibrils with different structural features and stability properties, due to the diverse rearrangements of the lysozyme polypeptide chain into the fibril structure.

In conclusion, the research work conducted in this Thesis allowed the comprehension of important aspects of the unfolding of some globular proteins leading to amyloid fibrils. In addition, original data have been obtained on the structural polymorphism of amyloid fibrils.

The work has been conducted mainly at CRIBI (University of Padua), also in collaboration with other laboratories, such as Biochemistry Department of University of Florence and Department of Chemistry at Cambridge University, where in particular the structural characterization of HuL fibrils obtained at neutral pH have been performed.

References

Booth DR, Sunde M, Bellotti V, Robinson CV, Hutchinson WL and Fraser PE et al. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis, Nature 27, 787–793

Chatani E, Goto Y. (2005) Structural stability of amyloid fibrils of beta(2)-microglobulin in comparison with its native fold. Biochim Biophys Acta. 1753, 64-75.

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

Chiti F, Bucciantini M, Capanni C, Taddei N, Dobson CM, Stefani M. (2001) Solution conditions can promote formation of either amyloid protofilaments or mature fibrils from the HypF N-terminal domain. Protein Sci. 10, 2541-2547.

Dumoulin M, Canet D, Last AM, Pardon E, Archer DB and Muyldermans S et al. (2005) Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. J. Mol. Biol. 25, 773–788.

Frare E, Mossuto MF, Polverino de Laureto P, Dumoulin M, Dobson CM, Fontana A. (2006) Identification of the core structure of lysozyme amyloid fibrils by proteolysis. J Mol Biol. 361, 551-561.

Frare E, Polverino de Laureto P, Zurdo J, Dobson CM and Fontana A (2004) A highly amyloidogenic region of hen lysozyme. J. Mol. Biol. 23, 1153–1165.

Kodali R and Wetzel R (2007). Polymorphism in the intermediates and products of amyloid assembly. Current Opinion in Structural Biology 17, 48-57.

Morozova-Roche LA, Zurdo J, Spencer A, Noppe W, Receveur V and Archer DB et al. (2000) Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants. J. Struct. Biol. 130, 339–351.

Pepys MB, Hawkins PN, Booth DR, Vigushin DM, Tennent GA and Soutar AK et al. (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis, Nature 362, 553–557

Polverino de Laureto P, Frare E, Battaglia F, Mossuto MF, Uversky VN, Fontana A. (2005) Protein dissection enhances the amyloidogenic properties of alpha-lactalbumin. FEBS J. 272, 2176-2188.

Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Proetin Science 11, 739-756.

CHAPTER 1.
FROM PROTEIN MISFOLDING TO AMYLOID AGGREGATION

A broad range of human diseases arises from the failure of a specific peptide or protein to adopt, or remain in, its native functional conformational state (Chiti and Dobson, 2006). These pathological conditions are generally referred to as protein misfolding (or protein conformational) diseases. They include pathological states in which impairment in the folding efficiency of a given protein results in a reduction in the quantity of the protein that is available to play its normal role. This reduction can arise as the result of a posttranslational process, such as an increased probability of degradation via the quality control system of the endoplasmic reticulum, as occurs in cystic fibrosis (Amaral, 2004), or the improper trafficking of a protein, as seen in early-onset emphysema (Lomas and Carrel, 2002). The largest group of misfolding diseases, however, is associated with the conversion of specific peptides or proteins from their soluble functional states ultimately into highly organized fibrillar aggregates. These structures are generally described as amyloid fibrils or plaques when they accumulate extracellularly, whereas the term “intracellular inclusions” has been suggested as more appropriate when fibrils morphologically and structurally related to extracellular amyloid form inside the cell (Westermark et al., 2005).

Current interest in studying amyloid fibrils arises from their involvement in different fields. First, they play a crucial role in disorders such as Alzheimer’s and Parkinson’s diseases. Second, since it has been proved that all polypeptide chains form this kind of fibrils under appropriate conditions, the understanding of why and how this process happen has become central problem in protein knowledge. Last, the ordered ultrastructure characterizing amyloid fibrils may be thought as a basis for nanomaterials with possible technological applications. However, despite the ability of most proteins to form amyloid fibrils, very little is known about their structures and the factors that govern their formation.

1.1. Amyloid diseases and amyloid fibril structure

A number of human diseases, including amyloidoses and many other neurodegenerative diseases, originate from the deposition of stable, ordered, filamentous protein aggregates, commonly referred to as amyloid fibrils. In each of these pathological states, a specific protein or protein fragment changes from its natural

soluble form into insoluble fibrils, which accumulate in a variety of organs and tissues (Kelly, 1998; Dobson, 1999; Rochet and Lansbury, 2000). Although approximatively 20 different proteins are known to be involved in amyloid diseases (extracellular deposits) (see Table I), they are unrelated in terms of sequence or structure. In addition, a number of diseases also involve the deposition of fibrillar intracellular deposits, as well as non-fibrillar deposits. Prior to fibrillation, amyloidogenic polypeptides may be rich in beta-sheet, alpha-helix, beta-helix, or contain both alpha-helix and beta-sheets (see Table I). They may be globular proteins with rigid 3D-structure or belong to the class of natively unfolded (or intrinsically unstructured) proteins (Wright and Dyson, 1999; Uversky, 2002). Despite these differences, the fibrils from different pathologies display many common properties including a core cross-beta-sheet structure in which continuous beta-sheets are formed with beta-strands running perpendicular to the long axis of the fibrils. Fibrils have similar morphologies, and mature fibrils usually have a twisted, rope-like structure, reflecting a filamentous substructure (Fig. 1). Typically mature fibrils consist of two to six unbranched protofilaments, 2-5 nm in diameter, associated laterally or twisted together to form fibrils with 4-13 nm diameter (Fandrich, 2007).

Amyloid fibrils have been formed *in vitro* from disease-associated as well as from disease-unrelated proteins and peptides (Table I). Moreover, there is an increasing belief that the ability to form fibrils is a generic property of the polypeptide chain, i.e. many proteins, perhaps all, are potentially able to form amyloid fibrils under appropriate conditions (Dobson, 1999, 2001; Fandrich et al., 2001).

Table I Amyloidogenic proteins and amyloid-based clinical disorders

Diseases-associated amyloidogenic proteins			
Amyloidogenic protein	Type of structure	disease	Tissue distribution of protein deposits
Prion protein and its fragments	N-terminal fragment (23–121) is natively unfolded; C-terminal domain (121– 230) is a-helical (predominantly)	Creutzfeld– Jacob disease (CJD) Gerstmann–Straussler–Schneiker syndrome (GSS) Fatal familial insomnia (FFI) Kuru Bovine spongiform encephalopathy (BSE) and scrapie	Brain
Amyloid-β and its fragments	Natively unfolded	Alzheimer's disease (AD)	Brain
		Dutch hereditary cerebral hemorrhage with amyloidosis (HCHWA, also known as cerebrovascular amyloidosis)	
		Congophilic angiopathy	Brain, spinal cord
ABri	Natively unfolded	Familial British dementia	Brain
Cystatin C	α/ β	Hereditary cystatin c amyloid angiopathy (HCCAA)	Brain
Huntingtin	α-Helical (but exon 1 is unfolded and forms fibrils)	Huntington Disease	Brain
Androgen receptor protein	Ligand-binding (LBD) and DNA-binding domains (DBD) are a-helical; amino-terminal domain (NTD) is natively unfolded	Spinal and bulbar muscular atrophy (SBMA)	Brain, scrotal skin, dermis, kidney, heart, and testis, spinal cord
Ataxin-1	Unknown (likely natively unfolded)	Spinocerebellar ataxia (SCA) Brain, spinal cord Neuronal intranuclear inclusion disease (NIID)	Central and peripheral nervous system
DRPLA protein (atrophin-1)	Unknown (likely natively unfolded)	Hereditary dentatorubral-pallidoluysian atrophy (DRPLA)	Brain
Serum amyloid A and its fragments	α/ β	AA amyloidosis (inflammation-associated reactive systemic amyloidosis)	Bladder, stomach, thyroid, kidney, liver, spleen, gastrointestinal tract
Medin (245–294 fragment of lactadherin)	β -Sheet	Aortic medial amyloidosis	Aortic smooth muscle
Islet amyloid polypeptide (IAPP, Amylin)	Natively unfolded	Pancreatic islet amyloidosis in late-onset diabetes (type II diabetes mellitus)	Pancreas
Calcitonin	Natively unfolded	Medullary Carcinoma of the Thyroid (MCT)	Thyroid
Lysozyme	α + β	Hereditary systemic amyloidosis	Several visceral organs and tissues
Gelsolin	α/ β	Hereditary systemic amyloidosis Finnish-type familial amyloidosis	Several visceral organs and tissues
Transthyretin	β-Sheet (predominantly)	Senile systemic amyloidosis (SSA) (or senile cardiac amyloidosis)	Almost all organs and tissues, including heart, gland, arteries, bones, liver, digestive tract, etc.
		Familial amyloid polyneuropathy (FAP)	Various organs and tissues.
Apolipoprotein A1	α-Helical	Hereditary systemic amyloidosis	Eyes
β2-Microglobulin	β-Sheet	Amyloid associated with hemodialysis (AH or Ah2M) (athropathy in hemodialysis)	Musculoskeletal tissues (large and medium-sized joints, bones, muscles), peripheral nervous system, gastrointestinal tract, tongue, heart, urogeniteal tract
Tau protein	Probably natively unfolded	Alzheimer's disease (AD), Pick's disease	Brain

Table I (continued)

Diseases-associated amyloidogenic proteins			
Amyloidogenic protein	Type of structure	disease	Tissue distribution of protein deposits
α-Synuclein	Natively unfolded	Parkinson's disease (PD) Diffuse Lewy bodies disease (DLBD) Lewy bodies variant of Alzheimer's disease (LBVAD) Dementia with Lewy bodies (DLB) Multiple system atrophy (MSA) Hallervorden– Spatz disease	Brain
NAC (central fragment of α-synuclein)	Natively unfolded	Alzheimer's disease (AD)	Brain
Fibrinogen and its fragments	β-Sheet	Hereditary renal amyloidosis	Kidney
Atrial natriuretic factor	''Small protein''	Atrial amyloidosis	Heart
Insulin	Predominantly a-helical	Injection-localized amyloidosis	Skin, muscles

Non-disease-related amyloidogenic proteins and peptides			
Protein (peptide)	Type of structure	Protein (peptide)	Type of structure
Betabellins 15D and 16D	β -Sandwich	Prothymosin a	Natively unfolded
Cytochrome c552	α-Helical	Myoglobin	α-Helical
Methionine aminopeptidase	α-Helical	Muscle acylphosphatase	α/β
Phosphoglycerate kinase	α/ β	Hen egg white lysozyme	α + β
Hen egg white lysozyme, β-domain	β-Sheet	Acidic fibriblast growth factor	β-Barrel
PI3-SH3 domain	β-Barrel	OspA protein, BH9–10 peptide	β-Turn
β-Lactoglobulin	β-Sheet (predominantly)	De novo ata peptide	α-Helix-turn-a-helix
Monellin	α/β	Lung surfactant protein C	α-Helix
Immunoglobulin light chain LEN	β-Sheet	α-Lactalbumin	α + β
HypF, N-terminal domain	α/ β	VL domain of mouse antibody F11	β-Sheet
Human complement receptor 1, 18– 34 fragment	Unfolded	Apolipoprotein C-II	Natively unfolded
Human stefin B	α/ β	Cold shock protein A	β-Barrel
GAGA factor	Natively unfolded	Protein G, B1 Ig-binding domain	Four-stranded β-sheet with a flanking α-helix
Yeast prion Ure2p	α- Helical/unfolded	Cold shock protein B, 1 –22 fragment	Unfolded
Herpes simplex virus glycoprotein B fragment	β-Structural	De novo proteins from combinatorial library	β-Structural
The fiber protein of adenovirus, 355–396 peptide from shaft	Fibrillar	Soluble homopolypeptides: poly-L-lysine poly-L-glutamic acid poly-L-threonine	Unordered

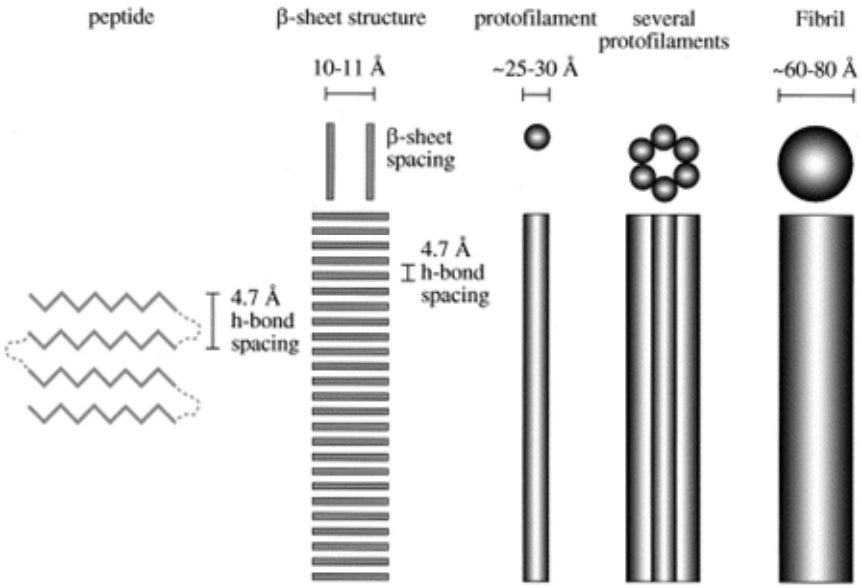


Fig. 1 Structural hierarchy of amyloid fibrils: from a protein in a beta-pleated sheet to amyloid protofilaments and fibrils. Views shown looking down the axis of the fibre and axially. The predominantly beta-sheet structure of the proto-filaments is given by beta-strands running perpendicular to the fibre axis and held together by hydrogen bonding. The so formed beta-sheets, parallel or antiparallel, assemble into the protofilament. A number of protofilaments, usually from 2 to 6, interacts along their length and twists together to form mature fibrils (Data from the Aβ peptide, adapted from Serpell, 2000)

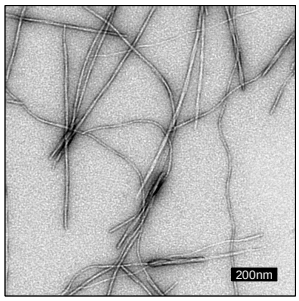
BOX 1

Hallmarks of amyloid fibrils

Amyloid is a non-covalent polymer of extended, intermolecularly hydrogen bonded beta-sheets that laterally self-assemble to yield twisted fibers.

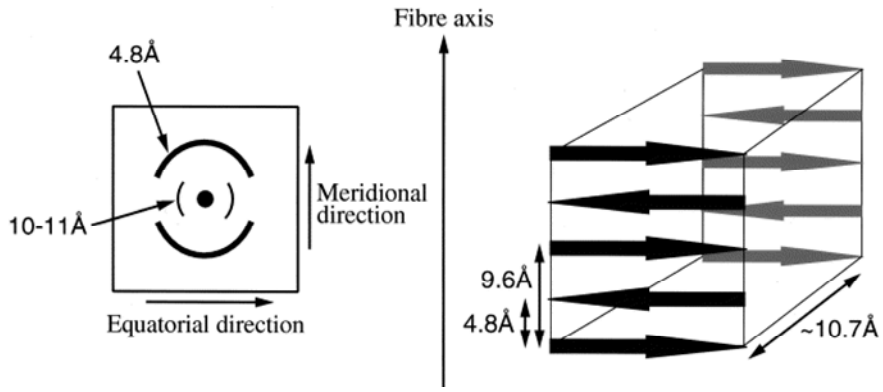
Morphology

Observing amyloid fibril with TEM, amyloid fibrils appear to be formed from unbranched protofilaments (2-5 nm wide) that twist together or associate laterally to form fibrils of higher width (generally 7-13 nm but even more).



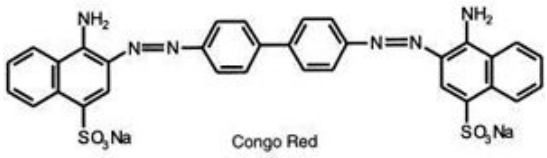
Cross-beta structure

Within individual protofilaments a β-sheet structure is present, characterized by β-strands running parallel to each other and perpendicular to the axis of the fibrils. Amyloid has a distinct X-ray diffraction pattern owing to this repeated structure. The meridional reflection at 4.7-4.8 Å results from the interstrand repeats (the inter beta-strands distance), and the 10-11 Å equatorial reflection arises from intersheet packing.

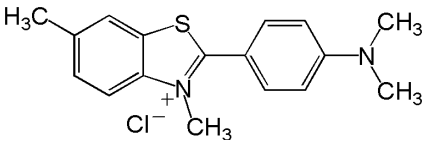


Dyes binding properties

Amyloid fibrils interact with specific dyes, such as Congo Red and Thioflavin T (ThT). Binding selectively to amyloid beta-sheets , ThT fluorescence emission undergoes a red shift (from 445 to 482 nm) and an increase in intensity. Samples of amyloi fibrils labeled with Congo Red, instead, give a typical green birefrangence under cross-polarized light.



Congo Red



Thioflavin T

1.2. Molecular basis of amyloid fibrils formation

To explain the molecular basis of amyloid fibril formation, it has been proposed that fibrillation can occur when the rigid native structure of a protein is destabilized, favouring partial unfolding and culminating in the formation of a partially unfolded conformation or intermediate (Rochet and Lansbury, 2000; Fink, 1998; Dobson, 2001). The logic behind this hypothesis is that since all fibrils, independently of the original structure of the given amyloidogenic protein, have a common cross-beta structure, considerable conformation rearrangements have to occur for this to happen. Such changes cannot take place in the typical tightly packed native protein conformation, due to the constraints of the tertiary structure. Thus, formation of non-native partially unfolded conformation is required. Presumably, such a partially unfolded conformation enables specific intermolecular interactions, including electrostatic attraction, hydrogen bonding and hydrophobic contacts, which are necessary for oligomerization and fibrillation (Fink, 1998).

This model, however, cannot directly apply to intrinsically unstructured (natively unfolded) proteins, as they are devoid of secondary structure to start with. Instead, the primary step of their fibrillogenesis has been shown to be the stabilization of a partially folded conformation, i.e. partial folding rather than unfolding occurs in such cases (Uversky et al., 2001; Goers et al., 2002). Thus, by taking the intrinsically unstructured protein into consideration, a general hypothesis of fibrillogenesis might be formulated as follows: structural transformation of a polypeptide chain into a partially folded conformation represents an important prerequisite for protein fibrillation.

The formation of amyloid-like fibrils is not the only pathological hallmark of "conformational" or protein deposition diseases. In several disorders (as well as in numerous *in vitro* experiments) protein deposits are composed of amorphous aggregates, without local order. Similarly, soluble oligomers represent another alternative final product of the aggregation process. The choice between the three aggregation pathways, fibrillation, amorphous aggregate formation and oligomerization, is determined by the amino acid sequence and by the details of the protein environment. Fig. 2 represents a simplified model of protein aggregation and illustrates the idea of that aggregation is an extremely complex process, which can be divided into three major steps. The model shows that proteins with different types of structure (alpha-helical, beta-structural, natively unfolded, single-domain or multi-domain, etc.) are equally subject to aggregation (Merlini and Bellotti, 2003).

The structural transformations of these diverse soluble proteins into the “sticky” aggregation-prone precursor or intermediate represent the first stage of the aggregation process. A single conformation of the intermediate is shown for the sake of brevity only. In fact, these aggregation-prone intermediates would be structurally different for different proteins. Furthermore, the intermediate might contain different amount/types of ordered structure even for the same protein undergoing different types of aggregation.

The formation of different oligomers (protofibrils or protoaggregates) represents the second stage of the aggregation process, which is usually considered as a nucleation step, in which formation of the nucleus is a kinetically disfavoured event, and leads to the lag period preceding significant formation of aggregates (Merlini and Bellotti, 2003). However, once a critical nucleus has been generated, the conditions change in favour of a rapid increase in size. As a result, any available aggregation-prone conformation quickly becomes entrapped in the fibrils, soluble oligomers or amorphous aggregate.

Several recent investigations lend support to the idea that fibrillar proteins (as in senile plaques in AD or Lewy bodies in PD) are not necessarily the toxic entity, but rather the formation of some “protofibrillar” structures is responsible for the toxicity (Lashuel et al., 2002).

1.3. Conformational states accessible to polypeptide chains during fibrillogenesis

The different features of the aggregation processes, described in the previous sections, reveal that polypeptide chains can adopt a multitude of conformational states and interconvert between them on a wide range of timescales. The network of equilibria, which link some of the most important of such states is schematically illustrated in Fig. 3 (Chiti and Dobson, 2006). Following biosynthesis on a ribosome, a polypeptide chain is initially unfolded. It can then populate a wide distribution of conformations, each of which contains little persistent structure, as in the case of natively unfolded proteins, or fold to a unique compact structure, often through one or more partly folded intermediates. In such a conformational state, the protein can remain as a monomer or associate to form oligomers or higher aggregates, some of which are functional with characteristics far from those of amyloid structures, such as in actin, myosin, and microtubules. Sooner or later, the vast majority of proteins will be

degraded, usually under very carefully controlled conditions as part of normal biochemical processes, with their amino acids often being recycled.

This description of normal functional behaviour, honed by millions of years of evolution, is, however, only part of the story. Fully or partially unfolded ensembles on the pathways to their functional states (or generated as a result of stress, chemical modification, or genetic mutation) are particularly vulnerable to aggregation. Peptides and proteins that are natively unfolded, as well as fragments of proteins generated by proteolysis and unable to fold in the absence of the remainder of the polypeptide chain, can also aggregate under some circumstances, for examples, if their concentrations become elevated. Some of the initial amorphous aggregates simply dissociate again, but others may reorganize to form oligomers with the germ of amyloid structure, including the spherical, chain-like, and annular protofibrils observed for many systems. In order to generate long-range order in such structures, a critical number of molecules must be present such that the favourably enthalpic terms associated with their regular stacking can most effectively offset the accompanying loss of configurational entropy.

The structured polypeptide aggregates can then sometimes grow into mature fibrils by further self-association or through the repetitive addition of monomers. Proteins that adopt a folded structure under physiological conditions can also aggregate under some circumstances. This latter type of proteins can either unfold, fully or partially, and aggregate through the mechanism described above or they can oligomerize prior to such a substantial conformational change. In the latter process, a structural reorganization to give amyloid-like assemblies occurs later and may in some cases be promoted by the existence of intermolecular contacts within native-like aggregates (Fig. 3).

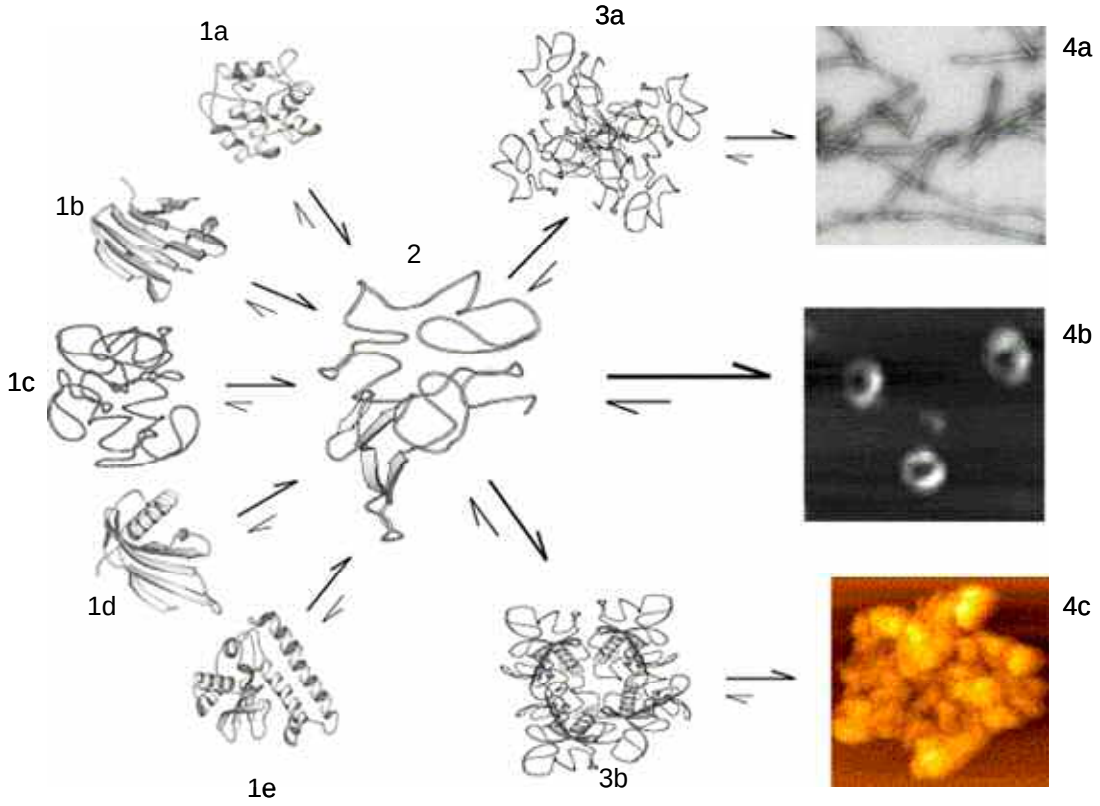


Fig.2 General model of protein aggregation. Aggregation is initiated by a structural transformation in the protein into a partially folded conformation. Proteins with different types of structure (α-helical, β-structural, natively unfolded, α+β or α/β; single-domain or multi-domain; etc., marked as 1a–1e) transform into the partially folded conformation; 2: partially folded molecules can assemble into specific oligomers (nucleus, or protofibrils, 3a and 3b, respectively). Depending on the peculiarities of the amino acid sequence and the environmental conditions, the protein may end up as amyloid fibril, 4a, soluble oligomer (torus-shaped in this case), 4b, or amorphous aggregate, 4c (Adapted from Uversky and Fink, 2004).

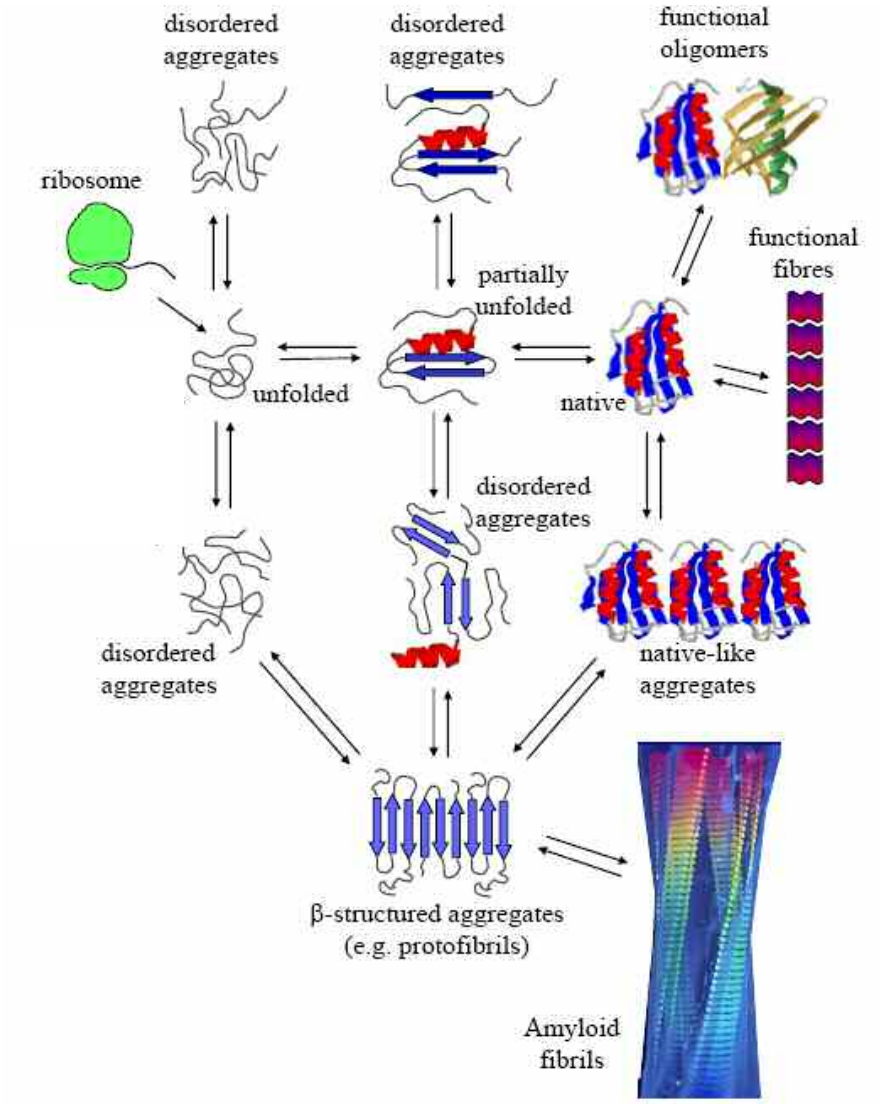


Fig.3 A schematic representation of some of the many conformational states that can be adopted by polypeptide chains and of the ways by which they can be interconverted. The transition from β-structured aggregates to amyloid fibrils can occur by addition of either monomers or protofibrils (depending on protein) to preformed β-aggregates. All of these different conformational states and their interconversions are carefully regulated in the biological environment, by using machinery such as molecular chaperones, degradatory systems, and quality control processes. Many of the various states of proteins are utilized functionally by biology, including unfolded proteins and amyloid fibrils, but conformational diseases will occur when such regulatory systems fail, just as metabolic diseases occur when the regulation of chemical processes becomes impaired.

References

Amaral MD (2004) CFTR and chaperones: processing and degradation. J Mol Neurosci. 23, 41-48.

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

Dobson CM (1999) Protein misfolding, evolution and disease. Trends Biochem. Sci. 24, 329–332.

Dobson CM (2001) Protein folding and its links with human disease. Biochem. Soc. Symp. 1-26.

Dobson CM (2001) The structural basis of protein folding and its links with human disease. Philos. Trans. R. Soc. Lond., B Biol. Sci. 356, 133-145.

Fandrich M (2007). On the structural definition of amyloid fibrils and other polypeptude aggregates. Cell. Mol. Life Sci. 64, 2066-2078.

Fandrich M, Fletcher MA, Dobson CM (2001) Amyloid fibrils from muscle myoglobin. Nature 410, 165-166.

Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold. Des. 3, 9-15.

Goers J, Permyakov SE, Permyakov EA, Uversky VN, Fink AL (2002) Conformational prerequisites for alpha-lactalbumin fibrillation. Biochemistry. 41, 12546-12551.

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

Lashuel HA, Hartley D, Petre BM, Walz T, Lansbury PT Jr. (2002) Neurodegenerative disease: amyloid pores from pathogenic mutations. Nature 418, 291.

Lomas DA and Carrell RW (2002) Serpinopathies and the conformational dementias. Nat Rev Genet. 3, 759-768.

Merlini G and Bellotti V (2003), Molecular mechanisms of amyloidosis, N. Engl. J. Med. 349, 583-596.

Rochet JC and Lansbury PT Jr. (2000) Amyloid fibrillogenesis: themes and variations. Curr. Opin. Struck. Biol. 10, 60-68.

Serpell LC (2000) Alzheimer's amyloid fibrils: structure and assembly. Biochim Biophys Acta. 1502, 16-30.

Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 11, 739-756.

Uversky VN, Fink AL (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim Biophys Acta. 1698, 131-153.

Uversky VN, Li J, Fink AL (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation, J. Biol. Chem. 276, 10737-10744.

Westermark P, Benson MD, Buxbaum JN, Cohen AS, Frangione B, Ikeda S, Masters CL, Merlini G, Saraiva MJ, Sipe JD (2005) Amyloid: toward terminology clarification. Amyloid. 12, 1-4.

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. J. Mol. Biol. 293, 321-331.

CHAPTER 2.
REQUIREMENT FOR PARTIAL UNFOLDING DURING FIBRILLOGENESIS OF GLOBULAR PROTEINS

Data have been reported indicating that the first critical step in protein fibrillogenesis is the partial unfolding of the protein (Kelly, 1998; Rochet and Lansbury, 2000; Dobson, 1999). It is difficult to trap and characterize such partially folded species under physiological conditions because they are only transiently populated on the fibrillation pathway. However, the reality of partial unfolding as an important prerequisite of fibrillation follows from several lines of evidence. Due to structural fluctuations (conformational breathing), the structure of a globular protein under physiological conditions represents a mixture of tightly folded and multiple partially unfolded conformations, with the former greatly predominating. Most mutations associated with accelerate fibrillation and protein deposition diseases have been shown to destabilize the native structure, increasing the steady-state concentration of partially folded conformers. A variety of compounds have been shown to significantly affect the rate of fibrillation in vitro, through a variety of mechanisms. For example, DNA-induced destabilization of the prion protein leads to its enhanced polymerization to amyloid fibrils (Nandi et al., 2002). Conversely, it has been shown that the amyloidogenicity of a protein can be significantly reduced by stabilization of the native structure, for example, via specific binding of ligands (Chiti et al., 2001; Ray et al., 2005; Dumoulin et al., 2003; Johnson et al., 2005).

Moreover, for several proteins it has been shown that destabilization of the native globular structure (e.g. low or high pH, high temperatures, low to moderate concentrations of strong denaturants, organic solvents, etc.) may significantly accelerate the rate of fibrils formation (Hamada et al., 2002).

Generally speaking, it has been concluded that amyloid formation in vitro can be achieved by destabilizing the native state of the protein under conditions in which non covalent interactions still remain favourable (Gujarro et al., 1998; Chiti et al., 1999; Hamada et al., 2002; Ramirez-Alvarado et al., 2000). Denaturation is not necessarily accompanied by the complete unfolding of a protein, but rather results in the appearance of new conformations with properties intermediate between those of the native and the completely unfolded states. It is known that globular proteins may exist in at least four different conformations: native (ordered), molten globule, premolten globule and

unfolded (Uversky and Ptitsyn, 1994, 1996; Uversky, 1997). The structural properties of the molten globule are well known, and have been systematized in a number of reviews (e.g. Ptitsyn 1995). It has been established that the protein molecule in this intermediate state has a globular structure typical of native globular proteins. 2D NMR coupled with hydrogen-deuterium exchange showed that the protein molecule in the molten globule state is characterized not only by the native-like secondary structure content, but also by the native-like folding pattern. A considerable increase in the accessibility of a protein molecule to proteases was noted as a specific property of the molten globule state (Fontana et al., 1993). It was also shown that transformation into this intermediate state is accompanied by a considerable increase in the affinity of a protein molecule to the hydrophobic fluorescence probes (such as 8-anilinonaphtalene-1-sulfonate, ANS) and this behaviour should be considered as a characteristic property of the molten globule state (Semisotnov et al., 1991; Uversky et al., 1996). Finally it was established that the averages value for the increase in the hydrodynamic radius in the molten globule state compared with the native state is no more than 15 %, which corresponds to volume increase of ~ 50 %.

A protein in the pre-molten globule state, instead, is denatured, that is , it has no rigid tertiary structure. It is characterized by a considerable secondary structure, although much less pronounced than that of the native or the molten globule protein (protein in the premolten globule state has ~ 50 % native secondary structure, whereas in the molten globule state the corresponding value is close to 100 %). The protein molecule in the premolten globule state is considerably less compact than in the molten globule or native states, but it is still more compact than the random coil (its hydrodynamic volume in the molten globule, the premolten globule, and the unfolded states, in comparison to that of the native state increases 1.5, ~ 3, and ~ 12 times, respectively). The protein molecule in the premolten globule state can effectively interact with the hydrophobic fluorescent probe ANS, although essentially weaker than in the molten globule state. This means that at least part of the hydrophobic clusters of polypeptide chain accessible to the solvent is already formed in the premolten globule state. Moreover, it has been established that in the premolten globule state the protein molecule has no globular structure. This finding indicates that the premolten globule probably represents a “squeezed” and partially ordered form of a coil. Finally, it has been shown that the premolten globule state is an all-or-none transition, which

represents an intramolecular analog of the first-order phase transition. This means that the molten globule and premolten globule states represent divers thermodynamic states.

In the following paragraphs, examples are reported of two proteins, alpha-lactalbumin and HypF-N, that form amyloidogenic intermediates and for which that is evidence for relatively unfolded amyloidogenic intermediates as critical species in fibrillation.

References

Chiti F, Taddei N, Stefani M, Dobson CM, Ramponi G (2001) Reduction of the amyloidogenicity of a protein by specific binding of ligands to the native conformation, Protein Sci. 10, 879-886.

Dobson CM (1999) Protein misfolding, evolution and disease, Trends Biochem. Sci. 24, 329-332.

Dumoulin M, Last AM, Desmyter A, Decanniere K, Canet D, Larsson G, Spencer A, Archer DB, Sasse J, Muyldermans S, Wyns L, Redfield C, Matagne A, Robinson CV, Dobson CM (2003) A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. Nature. 424, 783-788.

Fontana A, Polverino de Laureto P, De Filippis V (1993) Molecular aspects of proteolysis of globular proteins. In Protein stability and stabilization (eds., W. van den Tweel, A. Harder, and M. Buitelear), 101-110. Elsevier Science, Amsterdam.

Guijarro JI, Sunde M, Jones JA, Campbell ID, Dobson CM (1998) Amyloid fibril formation by an SH3 domain. Proc. Natl. Acad. Sci. U. S. A. 95, 4224-4228.

Hamada D, Dobson CM (2002) A kinetic study of beta-lactoglobulin amyloid fibril formation promoted by urea. Protein Sci. 11, 2417-2426.

Hamada D, Dobson CM (2002) A kinetic study of beta-lactoglobulin amyloid fibril formation promoted by urea. Protein Sci. 11, 2417-2426.

Johnson SM, Wiseman RL, Sekijima Y, Green NS, Adamski-Werner SL, Kelly JW (2005) Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidoses. Acc Chem Res. 38, 911-921.

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

Nandi PK, Leclerc E, Nicole JC, Takahashi M (2002) DNA-induced partial unfolding of prion protein leads to its polymerisation to amyloid. J. Mol. Biol. 322, 153-161.

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

Ramirez-Alvarado M, Merkel JS, Regan L (2000) A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. Proc. Natl. Acad. Sci. U. S. A. 97, 8979-8984.

Ray SS, Nowak RJ, Brown RH Jr, Lansbury PT Jr (2005) Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation. Proc Natl Acad Sci U S A. 102, 3639-3644.

Rochet JC, Lansbury PT Jr. (2000) Amyloid fibrillogenesis: themes and variations. Curr. Opin. Struct. Biol. 10, 60-68.

Semisotnov GV, Rodionova NV, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI (1991) Study of the “molten globule” intermediate state in protein folding by a hydrophobic fluorescent probe. Biopolymers 31, 119-128.

Uversky VN (1997) Diversity of equilibrium compact forms of denatured globular protein. Protein Pept. Lett. 4, 355-367.

Uversky VN, Ptitsyn OB (1994) ‘‘Partly folded’’ state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. Biochemistry 33, 2782-2791.

Uversky VN, Ptitsyn OB (1996) Further evidence on the equilibrium “pre-molten globule state”: Four-state GdmCl-induced unfolding of carbonic anhydrase B at low temperature. J. Mol. Biol. 255, 215-228.

2.1. Alpha-lactalbumin

Bovine α-lactalbumin (LA) is a small (14176.1 Da) calcium-binding milk protein. It is expressed in mammary secretory cells and is one of the two components of lactose synthase, which catalyses the final step in lactose biosynthesis in the lactating mammary gland, by transferring an UDP-galactose residue on glucose (Hill & Brew, 1975).

UDP-gal + glucose → lactose + UDP

The other component of this system is galactosyltransferase (GT), which normal function is to transfer galactosyl groups from UDP-galactose to glycoproteins containing N-acetylglucosamine. Instead, in the lactating mammary gland, the specificity of GT is modulated by interaction with LA, which increases the affinity and specificity of GT only for glucose.

Most of alpha-lactalbumins, including human, guinea pig, bovine, goat, camel equine and rabbit proteins, consist of 123 amino acid residues. LA is homologous in sequence to the lysozyme family, but it differs in the biological activity, since it exhibits cell lytic activity of ~ 10^{-6} of the specific activity of hen egg-white lysozyme. The three-dimensional structure of LA is very similar to that of lysozymes, as judged by X-ray crystallography. The structure of LA is stabilized by four disulfide bridges between Cys 6-120, 28-111, 61-77, and 73-91.

Native LA consists of two domains: a large alpha-helical domain and a small β-sheet domain, connected by a calcium binding loop. The alpha-helical domain is composed of four major alpha-helices A, B, C and D (amino acid residues 5-11, 23-34, 86-98 and 105-111) and two short 3_{10} helices (amino acid residues 18-20 and 115-118) (Peng & Kim, 1994; Wu et al., 1995; Schulman et al., 1995). The small beta-domain is composed by a series of loops, a small three-stranded antiparallel beta-pleated sheet (amino acid residues 41-44, 47-50 and 55-56) and a short 3_{10} helix between residues 77-80. The two domains are divided by a deep "cleft" and are held together by the cysteine bridge between residues 73 and 91, forming the Ca^{2+} binding loop, and by the disulfide bond 61-77.

LA is able to bind several metal cations. It has a single strong calcium-binding site, situated in a loop between the two helices B and C. The calcium binding site is formed by oxygen ligands from the carboxylic groups of three Asp residues (at position 82, 87 and 88) and from the two carbonyl groups of the peptide backbone of amino acid residues Lys79 and Asp84. LA has also several zinc binding sites, one of which is located in the "cleft" region, *i.e.* the region that forms the active site in lysozyme. Besides calcium, LA binds other physiologically significant cations, such as Mg^{2+}, Mn^{2+}, Na^{+} and K^{+}, which can compete with Ca^{2+} for the same binding site (Permyakov et al., 1985). The binding of Ca^{2+} to LA is very important for the maintenance of the native conformation of the protein. In particular, fluorescence and circular dichroism (CD) data showed that Ca^{2+}-binding causes marked changes in tertiary, but not secondary structure. Ca^{2+}-binding is essential also for the biological function of LA, even if its role has not been comprised yet. The binding of other metal ions induces similar, even if smaller, structural changes in LA.

Cation binding to the strong calcium-binding site increases the thermo-stability of LA. From differential scanning calorimetry data, the binding of Ca^{2+} shifts the thermal transition to higher temperatures by ~ 30-40°C. This might be related to some temperature regulation of LA stability and function in the mammary gland. The binding of Mg^{2+}, Na^{+} and K^{+} increases protein stability as well. On the contrary, the binding of Zn^{2+} ions to calcium-loaded LA decreases thermal stability, causes aggregation and increases its susceptibility to protease digestion. The results of this studies showed that LA in the presence of high concentration of Zn^{2+} is in a partly unfolded and partially aggregated state. It has been hypothesized that zinc binding could promote the immobilization and transport of the protein for nutritional purposes.

The binding of metal cations also increases the stability of LA towards the action of denaturing agents like urea or guanidine hydrochloride (Gdn·HCl). Therefore, any denaturation transition in LA (temperature/denaturant-induced) depends upon metal ion concentration, especially that of calcium ion.

LA possesses several fatty acid binding sites. It binds 5-doxylstearic acid (spin-labeled fatty acid analog), stearic acid and palmitic acid. LA interacts also with lipid membranes, like dimyristoylphosphatidylcholine, dipalmitoylphosphatidylcholine and lecithin vesicles. These observations are very interesting for the *in vivo* function and mechanism of action of LA, considering that the protein interacts with galactosyltransferase on membrane surfaces in the Golgi lumen.

A fascinating series of observations have raised the hypothesis that LA may have other important functions beyond its role in lactose biosynthesis. It has recently been reported that some multimeric, human LA derivatives are potent Ca^{2+}- elevating, apoptosis-inducing agent with broad, yet selective, cytotoxic activity on immature transformed, embryonic and lymphoid cells. Moreover, these authors revealed that this human LA folding variant has bactericidal activity against antibiotic-resistant and antibiotic-sensitive strains of *Streptococcus pneumoniae*. The active form of LA was purified from casein by anion exchange and gel-filtration chromatography. Moreover, recombinant native human and whey-derived LA could be converted to the active bactericidal/proapoptotic form by ion exchange chromatography in the presence of a cofactor from human milk casein, namely oleic acid (C18:1 fatty acid). This LA species has been named HAMLET (human α-lactalbumin made lethal to tumor cells). It was found that apoptosis-inducing species are oligomers of LA that have undergone a conformational change toward a MG-like state. Multimeric LA was shown to bind to the cell surface, enter the cytoplasm and accumulate in the cell nuclei. Köler et al. (1999) evidenced that caspases (cysteines-containing aspartate-specific proteases) are activated and involved in apoptosis induced by aggregated LA, through the direct interaction of LA with mitochondria that leads to the release of cytochrome *c*, which may be an important step in the initiation of the caspase cascade.

Furthermore, it has been evidenced that proteolytic digestion of bovine LA by trypsin and chymotrypsin yields three peptides with bactericidal properties, especially against Gram-positive bacteria. This suggests a possible antimicrobial function of LA after its partial digestion by endopeptidases.

In conclusion, binding to cations (like for example zinc) or to some fatty acids (oleic acid) might induce the formation of LA aggregated forms that perform various transport functions (for example with apolar, lipophilic vitamins and metabolites) or possess anticancer activity, by eliminating tumor cells.

2.1.1. Molten globules of bovine α-lactalbumin

The acid state of LA at low pH values (*A-state*) is considered the classical *molten globule* (MG) state (see § 2). The A-state of LA was extensively studied by many researchers who defined it as a compact state with fluctuating tertiary structure, but substantially unchanged content of secondary structure. LA in the acid MG state still retains a globular shape, but is more expanded than the native state, with a radius of gyration of 17.2 Å, compared with 15.7 Å for the native calcium-loaded LA. Hydrogen/deuterium exchange NMR spectroscopy studies evidenced that the most persistent structure is located in the helical domain, while the chain region encompassing the β-sheet is missing the native structural constraints and displays enhanced flexibility. This has led to define the A-state as having a "bipartite structure", where the α-domain is considered a discontinuous domain constituted by the N-terminal segment 1-37 and by the C-terminal segment 86-123, and the β-domain by the region encompassing amino acid residues 38-85.

Partly folded states of LA can also be generated by dissolution of the protein in moderate denaturants, like 2 M Gdn·HCl, by partially reducing the disulfide bonds or by removing the single protein-bound calcium ion by the use of chelating agents (apo-LA).

The conformational features of apo-LA have not been so largely studied as those of the A-state and there are still conflicting proposals about the real structural nature of this state. CD measurements in the far-UV indicate that in apo-LA the secondary structure is almost entirely conserved. On the contrary, near-UV CD spectra show that at temperature above 30°C and low ionic strength the dichroic signal is substantially reduced, suggesting a perturbation of the environment around the aromatic chromophores. Instead at 4.5 °C and at room temperature the CD spectra of native- and apo-LA are similar both in the far- and in the near-UV.

An apo-form of LA was obtained at neutral pH in the presence of EDTA as calcium-chelating agent and its conformational features analyzed by limited proteolysis. It has been reported that in the absence of calcium only the protein segment 34-57 is

flexible enough to be cleaved by different proteases. This region includes the three small β-strands of native LA and has been indicated as the "β-subdomain". Taken together, apo-LA can be considered a folding intermediate of LA having most of the polypeptide chain highly structured and native-like, but a quite disordered β-subdomain. This vision parallels that of a bipartite structure of the A-state, even if in this case the α-domain is considered larger and the β-domain smaller. A peculiarity of apo-LA is that it does not bind to chaperonin GroEL. Since it has been hypothesized that the MG is a sufficiently expanded and flexible conformational state to bind GroEL, the evidence that apo-LA does not bind this chaperonine indicates that its structure is still rather compact, with only limited denatured regions.

The effects of organic solvents, like ethanol, methanol and trifluoroethanol (TFE) on protein conformation have been also studied. It is well known that TFE is able to induce structure in otherwise unstructured peptides and to stabilize helical regions in proteins and peptides. The physical bases of the action of TFE on proteins are quite complex. TFE has at 25°C a lower dielectric constant (26.7) in comparison to water (78.5), making stronger the ionic interactions within the α-helix. Moreover, TFE is more acid than water and this makes weaker the hydrogen bonds between amide bridges and the solvent, while reinforcing intramolecular bonds (Nelson & Kallenbach, 1986). In addition, since TFE is a hydrophobic solvent, it can denature the tertiary and quaternary structure of a protein by interfering with the hydrophobic interactions. Therefore, proteins when dissolved in aqueous TFE usually acquire a stable, partly folded state with a higher content of α-helical conformation in respect to aqueous solution, but lacking the specific tertiary interactions of the native protein.

A TFE-induced, partly folded state of LA at pH 2.0 ("TFE-state") was analyzed by means of circular dichroism, fluorescence and NMR spectroscopy, showing that it possesses a high content of helical secondary structure with native-like characteristics, whereas the single β-sheet of the native protein appears to be largely eliminated. In detail, concentrations of TFE up to 15-20 % have little influence on the conformational properties of LA, whereas clear effects are visible with concentrations ranging between 20-50 %.

In previous studies, both the A-state and TFE-state of LA have been analyzed by the limited proteolysis technique. For analyzing LA in its A-state it was used pepsin, since it is active at acidic pH and has broad substrate specificity, cleaving at the C-

terminus of mainly bulky and hydrophobic amino acid residues. The conformational features of LA dissolved in aqueous TFE were analyzed by thermolysin. This protease appeared to be a most suitable proteolytic probe, because of its noteworthy stability under relatively harsh solvent conditions, including organic solvents and broad substrate specificity, since it cleaves at the amino side of Leu, Ile, Phe, Tyr and other residues as well. It was shown that LA in the two conditions can be cleaved by pepsin and thermolysin, respectively, at very few peptide bonds, localized at the level of the small β-subdomain, leading to quite large protein fragments and/or nicked protein species that are rather resistant to further proteolysis, thus allowing their isolation by reverse-phase HPLC.

These studies point out how some proteins, even such small proteins as LA, can perform several physiological functions depending on their location and most importantly, their conformational state.

2.1.2. Protein dissection enhances the amyloidogenic properties of alpha-lactalbumin

In recent years, numerous studies have been conducted on the low pH molten globule state of LA, nowadays considered a prototype MG in protein folding studies. Moreover, is has been reported that when LA is induced to adopt the MG state at low pH it is also able to form amyloid fibrils. However, fibrils formation at pH 2.0 is much more rapid if the protein, upon partial reduction of its four disulfide bridges, adopts a more open conformation than that of the classical MG in acid solution. This finding correlates with observations made with other systems, such as apomyoglobulin, SH3 and alpha-synuclein, leading to the conclusion that partly folded, but substantially open and dynamic states of proteins are those required for triggering the process of fibrillogenesis.

The work attached hereby, show the results of a deep study on the effects of the dissection of the LA molecule on its conformational features and aggregation processes. Protein dissection has been performed by means of proteolysis and the two proteolyzed species analyzed in the study are a LA species with a nick (Th1-LA) or a chain deletion (desbeta-LA) at the level of the peptide bond 40-41 and of the beta-subdomain of the 123-residue chain of the protein, respectively.

The obtained results gave some important insights into both the LA amyloidogenic behaviour and general rules governing amyloid processes. First, protein

conformations of Th1-LA and desbeta-LA, which are more relaxed and expanded in comparison to intact LA at low pH, are more suitable for triggering the fibrillation process. Importantly, the deletion of the beta-domain in LA does not alter the ability of the protein to assemble into well-ordered fibrils. Indeed, the removal of all the three beta-strands of the protein increases the ability to form amyloid fibrils, implying that the beta-sheet region of LA is not required for fibrillogenesis. In this context, it is interesting to recall that the beta-domain is though to play an important role in the amyloidogenesis of lysozyme, a protein belonging to the same structural superfamily of LA.

Moreover, as shown by FTIR and proteolysis experiments, during the lag phases of the aggregation processes of LA and its cited fragments it is evident that during the early stages of fibrillation every one of these species initially experience a conformational change to a more unfolded/flexible state, and, only later they become more structured.

It has been hence concluded that the inherent flexibility of these LA derivatives allows the large conformational changes required to form the cross-β-structure of the amyloid fibrils. It has been also shown that the initial stages of fibrillation of intact LA at low pH involve protein intermediate(s) characterized by enhanced chain flexibility, emphasizing the belief that the precursor structures of amyloid fibrils require a more unfolded and/or flexible state than that of the molten globule (MG).

At last, this study confirmed the importance of proteolysis not only as a biological way of destabilizing protein structures and triggering amyloidosis, as in Alzheimer's disease, but also as a useful analytical tool to destabilize in a targeted way protein structures and study their relative propensity to form amyloid aggregates.

Published paper (Polverino de Laureto et al., FEBS J. 2005, 272, 2176-2188)

2.2. HypF-N

HypF-N is the acronym for Hydrogenase maturation protein N-terminal domain of E. coli, and it constitutes the N-terminal domain of a maturation factor of the Hydrogenase complex in E. coli.

The N-terminal domain of the bacterial hydrogenase maturation factor HypF (HypF-N) was has been characterized both structurally and with regard to its folding and unfolding behavior (Calloni et al., 2003). This 91-residue protein domain belongs to the acylphosphatase family, though it is devoid of acylphosphatase activity; in fact, it is an α/β globular protein with a central twisted five-stranded ß-sheet faced by two antiparallel a-helices, with strands arranged in the same 4-1-3-2-5 topology as that of any other acylphosphatase (Fig. 4) (Rosano et al., 2002). The similarity to the members of the acylphosphatase family is confirmed by the presence in the sequence of the two conserved residues that are involved in catalysis in all active acylphosphatases, namely Arg-23 and Asn-41, and by ~34% amino acid identity with respect to the eukaryotic acylphosphatases, with a higher occurrence of hydrophobic residues and a lower isoelectric point.

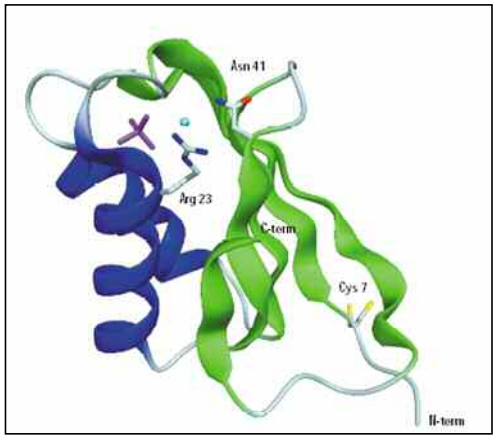
HypF-N has been reported to be able to form amyloid fibrils *in vitro* under conditions that promote its partial unfolding, such as in the presence of 30% (v/v) trifluoroethanol (TFE) at pH 5.5 or following a decrease of pH to pH 3.0 (Chiti et al., 2001; Relini et al., 2004).

The formation of a partially folded state is a key event in the aggregation pathway of this protein in vitro, even under mild destabilizing conditions in which the folded state is by far the predominant species (Marcon et al., 2005). Indeed, under mild denaturing conditions, generated by moderate concentrations of TFE, 6-12 % (v/v), the native state of HypF-N is in rapid equilibrium with a partially folded state , whose population is about 1-2 %, a kinetic analisys has shown that amyloid aggregation under these conditions arises from molecules accessing the amyloidogenic state (Marcon et al., 2005). The aggregation process of HypF-N has also been studied in its physiological environment in E. coli cells (Calloni et al., 2005). In this study it has been shown, by expressing a group of HypF-N variants, that an inverse correlations exists between the free energy change of unfolding of the variants and their tendency to form inclusion bodies in E. coli cells.

2.2.1. Conformational properties of the aggregation precursor state of HypF-N.

The conversion of specific proteins or protein fragments into insoluble, ordered fibrillar aggregates is a fundamental process in protein chemistry, biology, medicine and biotechnology. As this structural conversion seems to be a property shared by many different proteins, understanding the mechanism of this process will be of extreme importance. In this paper we present a structural characterisation of a conformational state populated at low pH by the N-terminal domain of *E. coli* HypF. Combining different biophysical and biochemical techniques, including near- and far-UV circular dichroism, intrinsic and ANS-derived fluorescence, dynamic light scattering and limited proteolysis, we will show that this state is largely unfolded but contains significant secondary structure and hydrophobic clusters. It also appears to be more compact than a random coil-like state but less organised than a molten globule state. In addition, increase of the total ionic strength of the solution induces aggregation of this conformational state into amyloid-like protofibrils, as revealed by AFM, providing a unique opportunity to study both the initial amyloid precursor state and the aggregation process by modulating the ionic strength of the solution.

Submitted paper (Campioni et al., J Mol Biol. 2007)



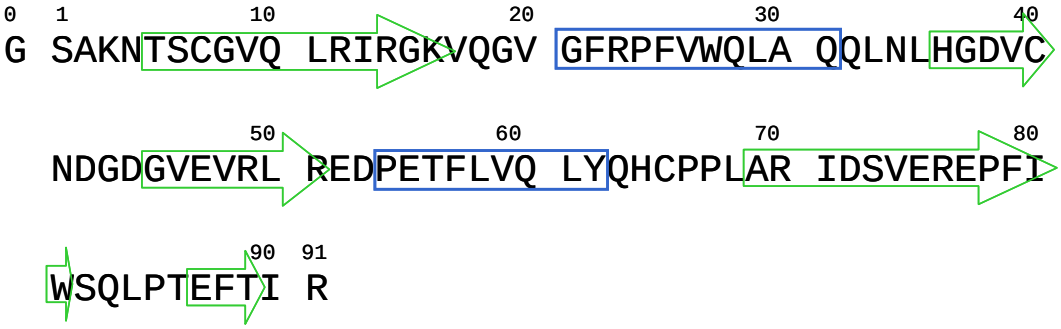


Fig. 4 Secondary structure and sequence of HypF-N. *Top*, a ribbon representation of HyPF-N obtained from the crystallographic data “1GXU” (Rosano et al., 2002). *Bottom*, HypF-N amino acid sequence has been derived fron ExPASy proteomic Server (Identification number P30131). Secondary structure elements are represented as green arrows for beta-strands and blue boxed for alpha-helices.

2.2.2. Identification of the HypF-N region involved in the initial HypF-N region involved in the initial protein-protein interaction leading to amyloid formation.

As it has been just shown (Campioni et al., 2007, submitted), a modulation of the total ionic strength of the solution allows enhancing the apparent rate of aggregation for HypF-N conformational state under acidic conditions. Here, the initial conformational state of HypF-N at acidic pH in presence of high ionic strength has been studied, by using spectroscopic techniques, such as CD and fluorescence, and other chemical methods such as proteolysis and cross-linking. The obtained data evidenced that the initial protein conformation leading very quickly to amyloid aggregates is an oligomer, binding already in the first phases the amyloid specific dye ThT. In these oligomers a particular region, corresponding to residues 57-62, has been found to be protected from proteolysis compared with the monomeric state of HypF-N under acidic conditions and low ionic strength. The mentioned region has the unique properties along the protein sequence to be at the same time highly hydrophobic, prone to form beta-sheet structures and uncharged at pH value in use. These three characteristics make this region the ideal mediator for the interaction between protein molecules and the stabilization of the first aggregation nuclei.

Materials and Methods

Materials

Wild type HypF-N was expressed and purified as previously described (Chiti et al., 2001). Due to this expression method, the protein has the natural N-terminal Met residue substituted by a Gly-Ser dipeptide. Thus, in order to follow the actual amino acid sequence, the N-terminal Gly and Ser residues are attributed position 0 and 1, respectively. The molecular mass of the purified protein (10464.2 Da) was checked with electrospray ionisation mass spectrometry. The protein purity was found by RP-HPLC to be > 95%. The less represented species corresponds to the N-terminally truncated form of the protein, spanning residues 4-91 (10121.5). The large prevalence (more than 96%) of the full length protein makes the sample suitable for the analyses presented here. Before each experiment, in order to avoid the presence of amorphous aggregates or large particles, the stock solution of HypF-N was centrifuged for 10 minutes at 10000 rpm and the supernatant was passed through a filter with a cut off of 20 nm.

All the proteases (porcine pepsin, protease type 13 from Aspergillus satoi, protease type 18 from Rhizopus sp. and proteinase K) and Thioflavin-T (ThT) were purchased from the Sigma Chem. Co. (St. Louis, MO). The reagent for electron microscopy are obtained from TAAB (TAAB-Laboratories Equipment Ltd, Berks, UK). Other chemicals were of analytical reagent grade and were obtained from Sigma or Fluka (Buchs, Switzerland). Filters having a cut off of 20 nm were provided by Whatman (Whatman International Ltd, Maidstone, England).

Methods

Optical spectroscopy

Circular dichroism (CD) spectra were obtained with a Jasco J-710 (Tokyo, Japan) dichrograph, equipped with a thermostated cell holder. The instrument was calibrated with d-(+)10-camphorsulfonic acid. Quartz cells with a 1 mm and 10 mm light pathlength were used for measurements in the far-UV and near-UV region, respectively. The mean residue ellipticity [θ] (deg cm^2 $dmol^{-1}$) was calculated using the formula $[\theta] = (\theta_{obs}/10)(MRW/lc)$, where θ_{obs} is the observed ellipticity in deg, MRW the mean residue molecular weight (113.74 for HypF-N), l the path-length in cm and c the protein concentration in mg/ml. Fluorescence measurements were carried out on a PerkinElmer (Wellesley, MA, USA) model LS-50B fluorimeter in a temperature-controlled cell holder set at 25°C, utilizing a 2 mm x 10 mm path length cuvette. Protein concentrations were evaluated from absorption measurements at 280 nm on a double-beam Lambda-2 spectrophotometer (Perkin Elmer). Extinction coefficient at 280 nm, calculated by the method of Gill & von Hippel, was 1.22 ml · mg^{-1} · cm^{-1} for the produced HypF-N domain.

Limited proteolysis experiments

Limited proteolysis experiments were carried out by incubating HypF-N at acid pH with porcine pepsin (Fruton, 1970), protease type 13 from Aspergillus satoi and protease type 18 from Rhizopus sp. (Cravello et al., 2003), and at neutral pH with proteinase K (Ebeling et al., 1974). Enzymatic digestions were performed in 20 mM TFA, 30 mM or 330 mM NaCl (pH 1.7) at 25°C using enzyme-to-substrate ratios of 1:1500 for pepsin and 1:50 (w/w) for the other acid proteases. Proteolysis of native state HypF-N was performed in 10 mM Tris·HCl, 30mM NaCl, 0.2 mM DTT (pH 7.0) with

Proteinase K as proteolytic probe, using an enzyme-to-substrate ratio of 1:50. Peptide mixtures from the different reactions were fractionated by reverse-phase HPLC on a Phenomenex Jupiter C18 column (4.6 × 150 mm) with a linear gradient of acetonitrile containing 0.1% (v/v) trifluoroacetic acid from 5% to 22% in 5 min and from 22% to 50% in 17 min, at a flow-rate of 0.6 ml/min. The effluent from the column was monitored by measurement of the absorbance at 226 nm. The individual fractions were manually collected and analysed by electrospray ionisation mass spectrometry (ESI MS) with a Q-Tof Micro mass spectrometer (Micromass, Manchester, UK). Samples were injected directly into the ion source by a Harward pump syringe at a flow rate of 20 μl/minute. Data were acquired and processed using Biomultiviewer (Perkin Elmer) software. Phosphoric acid was used to calibrate the instrument. The proteolysis mixtures were analyzed also with SDS-PAGE according to the method of Schägger and von Jagow (1987).

Aggregation processes of HypF-N

HypF-N fibrils were prepared by incubating protein samples (48 μM) at pH 1.7 in presence of 30 mM NaCl or 330 mM NaCl at 25°C for up to 100 days. Aggregation of HypF-N was monitored by ThT fluorescence assay, circular dichroism (CD) and electron microscopy (TEM). For the circular dichroism analyses the aggregation processes were performed directly in the appropriate cuvette, while for the ThT binding assay and the TEM controls, aliquots of each protein solution were taken at regular time-intervals.

The ThT binding assays were performed at regular time-intervals: 60 μl aliquots of each sample were added to 440 μl samples of a freshly prepared solution containing 25 mM ThT and 25 mM sodium phosphate buffer (pH 6.0). The fluorescence of the resulting samples was measured at 25 °C using a 2 x 10 mm path-length cuvette and a Perkin-Elmer model LS-50B fluorimeter (Wellesley, MA, USA). The excitation and emission wavelengths were 440 nm and 485 nm, respectively. All measured fluorescence values are given after subtracting the ThT fluorescence intensity measured in the absence of protein and normalised so that the final fluorescence intensity at the endpoint of the kinetic trace was 100%.

Electron micrographs were acquired using a Tecnai G^2 12 Twin transmission electron microscope (FEI Company, Hillsboro, OR, USA), operating at an excitation voltage of 100 kV. HypF-N (48 μM) was incubated for different periods at pH 1.7, 25°C

in the presence either of 30 mM NaCl or 330 mM NaCl. Aliquots (50 μl) of the sample were placed on a 400-mesh, glow-discharged Butvar and carbon-coated copper grids for 10 seconds and the excess liquid was removed by blotting with filter paper. A drop of 50 μl of 1% (w/v) uranyl acetate was then added, left on the grid for one minute and then blotted off. This latter procedure has been repeated twice.

Glutaraldehyde Cross-linking

Samples of 2.6 M glutaraldehyde were added to 48 μM HypF-N in 20 mM TFA, 30mM or 330 mM NaCl (pH 1.7) or in 10mM sodium phosphate (pH 7.0) using a protein-to-crosslinker ratio of 1:100 (w/w). The reaction was performed at 25°C and was quenched after 10 minutes adding an excess of Tris·HCl. The samples were then subjected to SDS-PAGE in 15% gel (Laemmli, 1970) and Coomassie staining.

Results

Conformational characterization of HypF-N at low pH.

The conformational features of HypF-N in solution were examined by circular dichroism (CD) and fluorescence spectroscopy. All the analyses (Fig. 5 A) were performed at room temperature (25°C), at neutral pH (10mM Tris·HCl, 30mM NaCl, pH 7.0) (dotted line) and at pH 1.7 in the presence of low (30 mM NaCl; dashed line) and high 330 mM NaCl; solid line) ionic strength. The protein concentration was reduced to 0.1 mg/ml (9.6 μM) in order to avoid the rapid formation of aggregates during the experiment time/period. In Figure 1 are shown the CD spectra of HypF-N recorded in the far-UV (A) and near-UV (B) regions at pH 7.0 and pH 1.7 with 30mM NaCl or 330mM NaCl. While the far-UV CD spectrum at pH 7.0 shows the form of an alpha-beta structured protein, at pH 1.7 and low ionic strength an ellipticity minimum at ~ 200 nm is predominant, which could be indicative of a little structured protein. Increasing the ionic strength, however, the shape of the far-UV CD spectrum changes as if the protein gained some structure or became more rigid.

The near-UV CD spectrum at pH 7.0 of HypF-N, which contains Trp27, Trp81 and Tyr62, is dominated by two sharp positive bands between 270 nm and 290 nm (Figure 5 B), suggesting that at neutral pH HypF-N has a well defined tertiary structure. By contrast, the near-UV CD spectrum of the protein at pH 1.7 and low ionic strength is characteristic of a fully random polypeptide. Interestingly, the increase of the ionic

strength to 330 mM NaCl produced the appearance of two positive bands between 270 nm and 290 nm in correspondence of those in the neutral conditions, although the intensity of the signal appears to be somewhat reduced.

Also fluorescence spectra (Fig. 5 C) registered in the same three conditions showed that while at pH 7.0 the wavelength of maximum emission is 346 nm, due to the fact that Trp81 is fully buried and Trp27 is for the 25% of its surface in contact with the solvent (Marcon et al., 2005), at pH 1.7 and in presence of low ionic strength there is a red shift of the wavelength of maximum emission toward 350 nm, revealing the loss of the native-like close packing around one or both of the Trp residues, as already seen by Chiti and coworkers (Marcon et al., 2005). Of note, the fluorescence spectrum of HypF-N at pH 1.7 in presence of 330 mM NaCl is characterized by a blue shift of the wavelength of maximum emission toward 344 nm, underlining a substantial structuration around the Trp residues.

Limited proteolysis

In order to map the polypeptide chain flexibility and accessibility (Fontana et al., 2004) limited proteolysis experiments were performed on HypF-N at a concentration of 0.1 mg/ml (9.6 μM), at room temperature, at pH 1.7 in the presence of low and high ionic strength and compared with the proteolysis experiments conduced at neutral pH (10mM Tris·HCl, 30mM NaCl, pH 7.0) on the native protein. At acid pH as proteases were used porcine pepsin, protease type 13 from Aspergillus satoi and protease type 18 fron Rhizopus sp. (Cravello et al., 2003; Del Mar et al., 2005), while at neutral pH proteinase K. For each protease many trials have been conducted to determine the right enzyme-to-substrate ratio, in order to obtain a non-exaustive, but limited, proteolysis and be able to identify the first cleavage sites on the polypeptide chain. For this reason porcine pepsin was used with an enzyme-to-substrate ratio of 1:1500 (w/w), while for the other three proteases the enzyme-to-substrate ratio value chosen was 1:50(w/w). The reaction time was fixed on 10 minutes, and then the reaction was blocked with pepstatin for acid proteases and with TFA 4% with proteinase K. Every proteolysis reaction was analyzed by SDS-PAGE, RP-HPLC and all the proteic material eluted from the column was identify by mass spectrometry.

The proteolysis of the native state (pH 7.0) with proteinase K (Fig. 6) produced two major bands in SDS-PAGE and two peaks in reverse-phase HPLC, corresponding to the full length protein (0-91) and the N-terminal truncated species 6-91. These data

confirmed the substantial stability and rigidity of the native state of HypF-N, where only the N-terminal region is sensible to proteolysis as proved also by the B-factor pattern obtained from the X-ray structure "1GXU" (Rosano et al., 2002).

At acid pH (Fig. 7), the proteolysis results evidenced a more complex situation. Summing up, at acid pH and low ionic strength the primary proteolysis sites were at the level of the peptide bonds Phe57-Leu58, Leu58-Val59, Leu61-Tyr62, E87-Phe88 and Phe88-Thr89 for pepsin (Fig. 7, A), at the level of the peptide bonds Q60-Leu61, Asp72-Ser73 and E87-Phe88 for protease type 13 from Aspergillus satoi (Fig. 7, C) and at the level of the peptide bonds Phe57-Leu58, Q60-Leu61, Asp72-Ser73, E87-Phe88 and Phe88-Thr89 for protease type 18 from Rhizopus sp. (Fig. 7, E). In the presence of 330 mM NaCl, instead, the results are quite different. The primary proteolysis sites were at the level of the peptide bonds Q10-Leu11, Leu11-Arg12, E87-Phe88 and Phe88-Thr89 for pepsin (Fig. 7, B), while the protein is completely resistant to the other proteases (Fig. 7, D, F). None of the reaction was exhaustive, and a certain amount of HypF-N molecules did not undergo proteolysis: in Fig. 6 and 7 only the primary proteolysis sites, that are the only ones that generate complementary fragments along the polypeptide chain, have been reported.

Aggregation of HypF-N at acid pH

Aggregation of HypF-N was monitored at acid pH (20 mM TFA, pH 1.7) in presence of high ionic strength, 330 mM NaCl, and low ionic strength, 30mM NaCl. In both cases, the protein was incubated at a concentration of 0.5 mg/ml (48 μM), at 25°C without stirring. In presence of high ionic strength the formation of aggregates was detectable after a period of 2-4 hours. The amyloid nature of the formed aggregates was confirmed by ThT binding assay (Fig. 8 A): the fluorescence of ThT, in fact, increases little by little following the growing amyloid structure. Interestingly, this process does not present a lag phase. The initial ThT fluorescence value, indeed, corresponds to the 20 % of the final ThT fluorescence value, evidencing the existence in solution of some structures capable of binding ThT. Far-UV CD spectra (Fig. 8 B) recorded during the aggregation

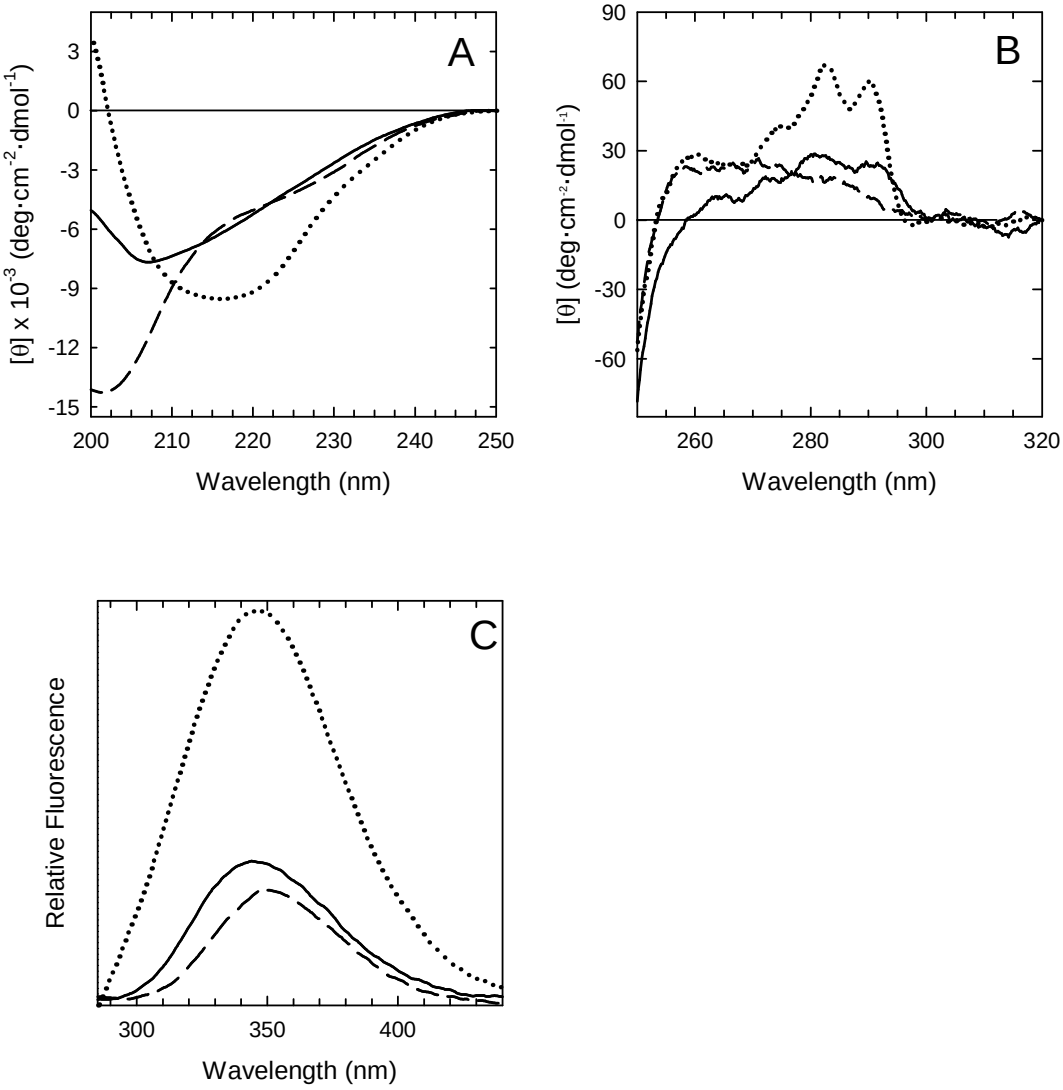


Fig. 5 Conformational characterization of HypF-N using far-UV CD (A), near-UV CD (B) and fluorescence (C) spectroscopy. Spectra were taken at 25 °C at neutral pH (10mM Tris·HCl, 30mM NaCl, pH 7.0, dotted line) and at pH 1.7 in the presence of 30 mM (dashed line) or 330 mM (solid line) NaCl.

process revealed the conformational changes involved in fibrillogenesis: the initial (t = 0) partially structured conformation of HypF-N, the subsequent (30min-3hours) shift of the minimum of signal to 217-218 nm corresponding to the acquisition of beta-structure and, after 4 hours, the disappearance of the signal due to the precipitation of the aggregates. Previous AFM measurements (Campioni et al., 2007, submitted) demonstrated that HypF-N at acidic pH in presence of high ionic strength evolves into amyloid structures. Here TEM has been applied to characterized the morphology of the initial stages of the aggregation process in the cited conditions. The electron micrographies taken at once after the protein dissolution in the aggregation buffer (Fig 8 C) shows that the starting point of the aggregation is characterized by the presence of globular aggregates having a diameter ranging from 15 to 18 nm.

Cross-linking

To confirm the presence of protein aggregates a sample of the initial stage of aggregation of HypF-N at low pH and high ionic strength was cross-linked by glutaraldehyde and compared with the same analyses performed on HypF-N at low pH but low ionic strength. HypF-N was cross-linked both at neutral pH and at acid pH in the presence of low and high ionic strength, and the cross-linking reactions products were visualized by SDS-PAGE (Fig. 9). At pH 7.0 (line 4) and at pH 1.7 and low ionic strength (line 3) the only cross-linking product appears to be a dimer (dashed arrow), in agreement with the fact that the cross-linking of a monomeric protein solution gives cross-linked products due to the casual and statistic interaction between monomers (Bitan et al., 2001). Surprisingly, at pH 1.7 and high ionic strength (line 2) the gel shows the presence of many cross-linked products, from dimer (in an extent much greater than the other two reaction conditions) to esamer (dashed and continue arrows). Even analyzing these reactions by SDS-PAGE in 10% gel or performing other cross-linking reactions with a greater amount of glutaraldehyde (data not shown), the aggregates pattern was always the same and no oligomers bigger than esamers seemed to be produced.

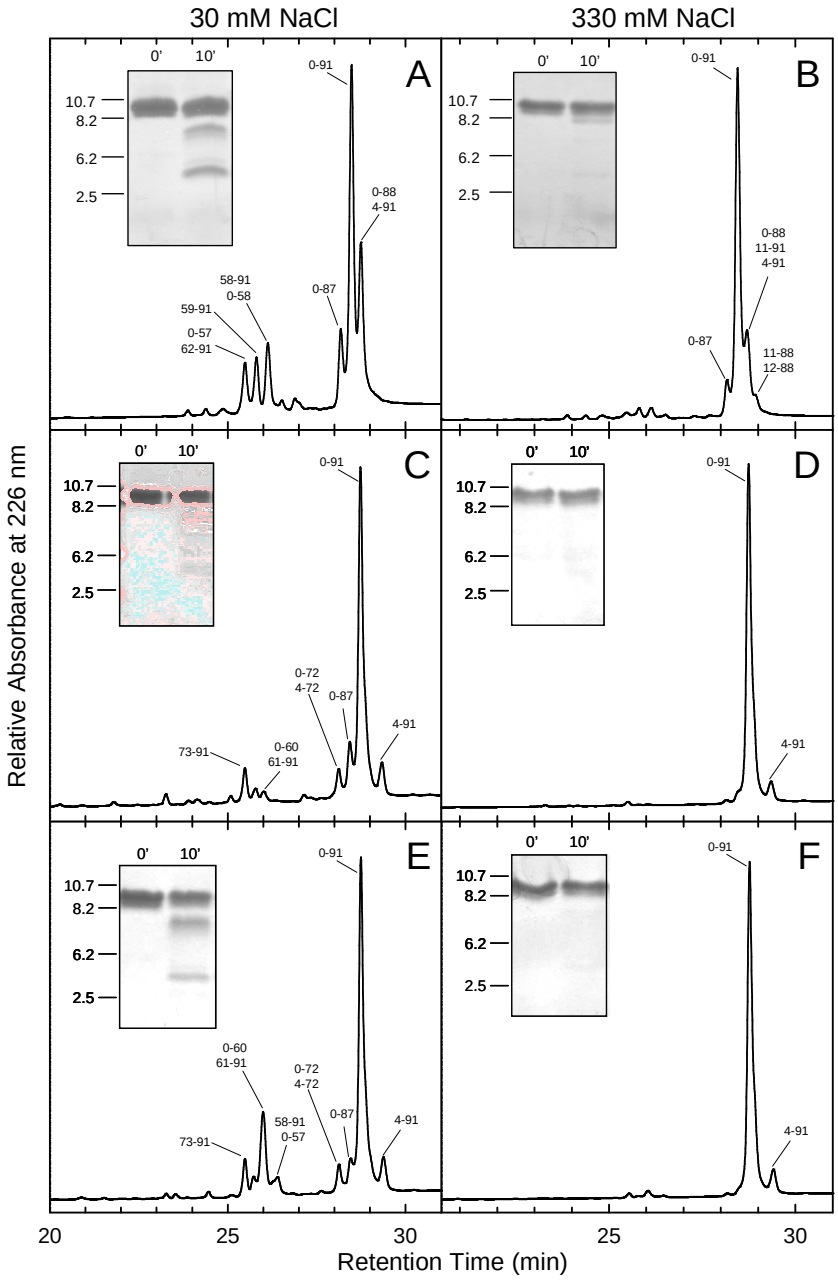


Fig. 6 Limited proteolysis of HypF-N at acid pH by porcine pepsin (A, B), protease type 13 from Aspergillus satoi (C, D) and protease type 18 from Rhizopus sp. (E, F) followed by RP-HPLC and SDS-PAGE (inserts). Proteolyses were performed at 25 °C in 20 mM TFA pH1.7, in the presence of 30 mM (A, C, E) or 330 mM NaCl (B, D, F) using an E : S ratio of 1 : 1500 (by weight) for pepsin and 1 : 50 for protease type 13 from Aspergillus satoi and protease type 18 from Rhizopus sp., at a protein concentration 0.1 mg/ml. Aliquots were taken from the proteolysis mixture after 10 minutes and analyzed by RP-HPLC (see Material and Methods). The identity of the various protein species was determined by mass spectrometry (see the text and Table I).

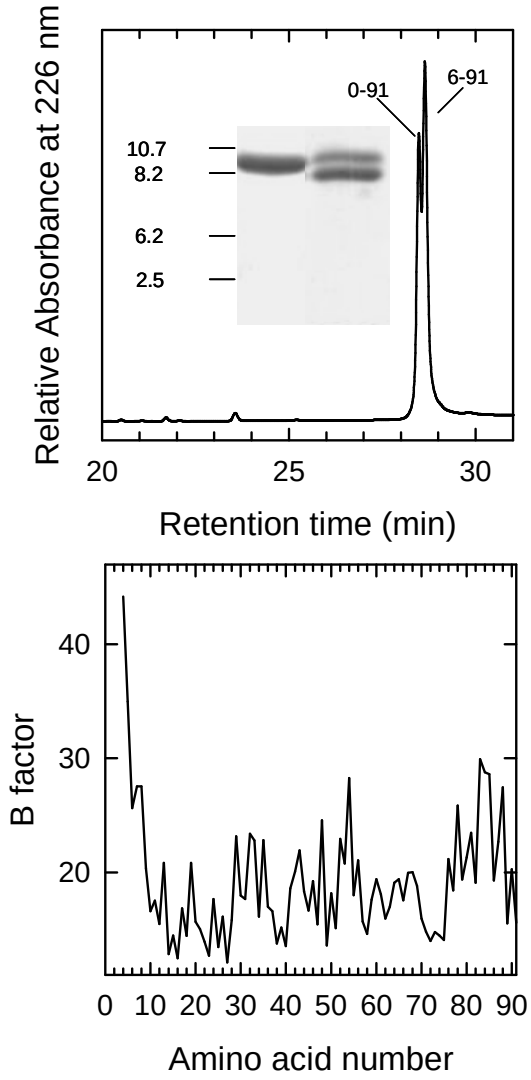


Fig. 7 Proteolysis of HypF-N at neutral pH by proteinase K followed by RP-HPLC (*Top*) and SDS-PAGE (*Top, insert*). Proteolysis was performed at 25 °C in 10 mM Tris-HCl pH 7.0, in the presence of 30 mM and 0.2 mM DTT, using an E : S ratio of 1 : 50 (by weight) at a protein concentration of 0.1 mg/ml. An aliquot was taken from the proteolysis mixture after 10 minutes, and analyzed by RP-HPLC (see Materials and Methods). The identity of the various protein species was determined by mass spectrometry (see the text and Table I). *Bottom*, B-factor pattern form "1GXU" crystallographic data.

Table II Analytical characterization of the peptide fragments obtained by degradation, acid hydrolysis and proteolysis of HypF-N.

Mass		Fragment
Found [a]	Calculated [b]	
10464.2	10464.8	0-91
10121.5	10121.5	4-91
10094.8	10094.4	0-88
9946.9	9947.3	0-87
9906.6	9906.3	6-91
9431.2	9431.8	11-91
9061.7	9061.3	11-88
8948.0	8948.2	12-88
8147.6	8147.3	0-72
7803.6	7803.9	4-72
6739.2	6739.6	0-60
6512.4	6512.4	0-58
6399.3	6399.2	0-57
5924.2	5924.7	0-53
5612.9	5612.4	45-91
4870.1	4870.5	0-44
4558.6	4558.2	54-91
4083.9	4083.7	58-91
3970.1	3970.6	59-91
3743.7	3743.3	61-91
3629.9	3630.1	62-91
2335.2	2335.6	73-91

[a] Determined by ESI-MS.

[b] Molecular masses calculated from the amino acid sequence of the expressed HypF-N.

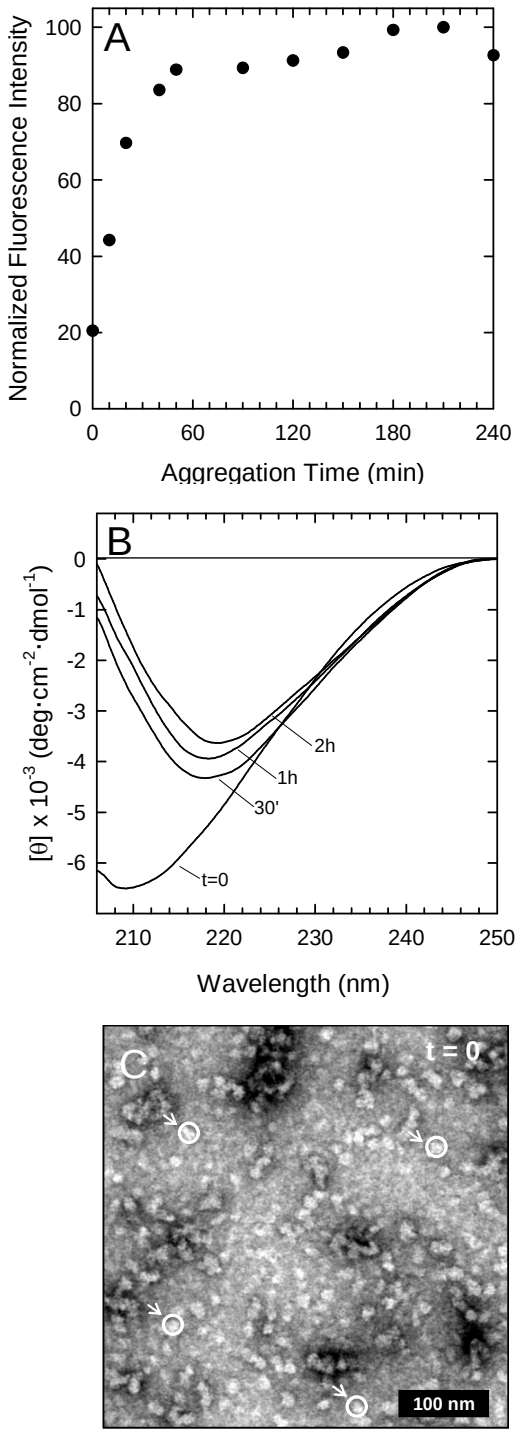


Fig. 8 HypF-N aggregation at pH 1.7 in presence of 330 mM NaCl monitored by ThT binding, circular dichroism and electron microscopy. (A) ThT binding curve obtained by adding an aliquot of the protein solution, incubated for the period of time indicated, to a solution of ThT. The excitation wavelength was fixed at 440 nm and the emission was collected at 485 nm. (B) Time-dependence of far-UV CD spectra of HypF-N during aggregation at pH 1.7, 330 mM NaCl, 25°C. (C) Electron micrographs of starting 0.5 mg/ml HypF-N solution at pH 1.7, 330 mM NaCl, 25°C.

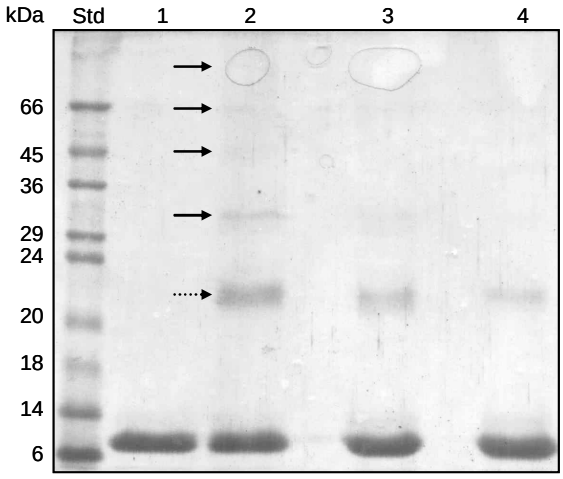


Fig. 9 SDS-PAGE analysis of glutaraldehyde cross-linking (GCL) of HypF-N. Samples of 2.6 M glutaraldehyde were added to 48 μM HypF-N in 20 mM TFA pH 1.7 in presence of 330 mM NaCl (line 2) or 30mM (line 3) or at pH 7.0 (line 4) using a protein-to-crosslinker ratio of 1:100 (w/w). The reaction was performed at 25°C and was quenched after 10 minutes adding an excess of Tris·HCl. Line 1 represents the HypF-N monomer not treated with glutaraldehyde.

Discussion

Here it is proposed that aggregation of HypF-N at acidic pH in presence of high ionic strength is triggered by the initial formation of globular aggregates. The region of the protein sequence that results to be exposed and flexible at acidic pH and low ionic strength where the protein is monomeric, but protected from proteolytic attack in these aggregates corresponds to residues 57-62 (Fig. 10 A, B).

HypF-N is readily able to form amyloid aggregates at acidic pH and in presence of high ionic strength, conditions in which the molecule is in a conformation very prone to aggregation as demonstrated by CD spectroscopy, ThT binding assay and TEM. ThT binding assay has evidenced that HypF-N binds the dye already in the early phase of aggregation, since ThT fluorescence emission is 20% of the plateau at the initial time (Fig. 8 A), and that there is no lag phase. Under these conditions HypF-N has been shown to be more structured than in acidic pH in presence of low ionic strength. Moreover, it is also resistant to acidic hydrolysis and proteolytic attack. Indeed, as previously reported (Campioni et al., 2007 submitted), at acidic pH and in presence of high ionic strength HypF-N is stable against acidic cleavage at the level of peptide bond Asp53-Pro54, the peptide bond most sensitive to acid hydrolysis (Li et al., 2001). In presence of low ionic strength, on the counterpart, this peptide bond is quickly hydrolyzed.

Proteolysis experiments performed on HypF-N have shown that the protein is easily cleaved by using a set of acidic proteases in presence of low ionic strength, while increasing the amount of salts the protein become more resistant to proteolytic attack. In particular, the region protected from proteolysis spans from residue 57 to 62. Notably, these two protein cleavage mechanisms, acidic hydrolysis and enzymatic proteolysis, depend of course on the sequence of the polypeptide chain, but also on the extent of exposition of the cleavable peptide bonds to the cleavage agent, both acidic solvent or protease.

Finally, TEM and crosslinking experiments have demonstrated the presence of aggregates at the starting point of aggregation.

Taken together these data prove that at the initial point of aggregation hypF-n collapse/exist into globular aggregates partially structured, quickly triggering the aggregation phenomenon/process.

It is well known (Dubay et al, 2004; Pawar et al., 2005, Calamai et al., 2003) that protein regions flexible, hydrophobic and highly prone to form beta-sheet possess

all the ideal characteristic to promote fibrils formation, because they allow the interaction between protein molecules and the stabilization of the first aggregation nuclei. Hence, we have compared the beta-sheet propensity and hydropathy profiles of HypF-N using the Street and Mayo scale (Street and Mayo,1999) and the Roseman scale (Roseman, 1988) respectively (Fig. 10 C, D). It clearly appears that the region 57-62 shows both a high hydrophobicity and a high propensity to form beta-sheet structure. The proteolysis sites hidden and protected at high ionic strength, turn out to be the most suitable to induce intermolecular interactions, by means of their flexibility and their exposure (Fontana et al., 1999). This central region of HypF-N could thus mediate the formation of the first soluble oligomers characterized by weak interactions between molecules and that need further conformational reorganization to gain the typical beta sheet structure of amyloid fibrils (Monti et al., 2004; Chiti and Dobson, 206). The high ionic strength used in this study accelerated the aggregation process helping the formation of these kinds of aggregates. In this scenery, the increase of ionic strength acts as an enhancer of the interactions between this critical region of different molecules, through a “salting out” mechanism. In presence of low ionic strength, in fact, the same region is exposed to the solvent, but there is no other driving force toward formation of the first aggregates.

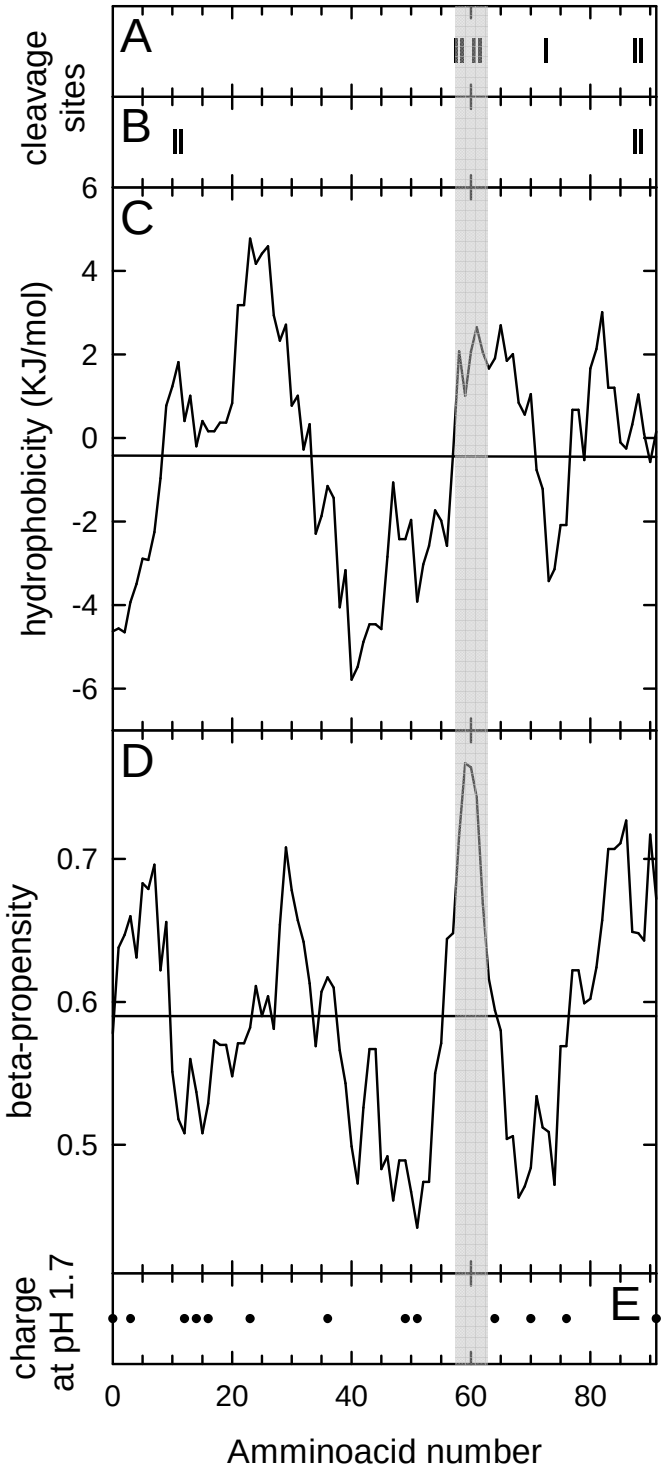


Fig. 10 Correlation between proteolysis patterns of HypF-N at pH 1.7 in presence of 30 mM (A) and 330 mM (B) NaCl and hydropathy (C) and beta-sheet propensity (D) profiles of HypF-N polypeptide chain sequence, determined using the Roseman scale and the Street and Mayo scale respectively. In C, circles represent the aminoacid positions of positively charged residues at pH 1.7.

References

Bitan G, Lomakin A, Teplow DB (2001). Amyloid beta-protein oligomerization: prenucleation interactions revealed by photo-induced cross-linking of unmodified proteins. J. Biol. Chem. 276, 35176-35184.

Calamai M, Taddei N, Stefani M, Ramponi G, Chiti F. (2003). Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. Biochemistry 42, 15078-15083.

Calloni G, Taddei N, Plaxco KW, Ramponi G, Stefani M, Chiti F. (2003) Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. J Mol Biol. 330, 577-591.

Calloni G, Zoffoli S, Stefani M, Dobson CM, Chiti F. (2005) Investigating the effects of mutations on protein aggregation in the cell. J Biol Chem. 280, 10607-10613.

Chiti F and Dobson CM. (2006). Protein misfolding, functional amyloid, and human disease. Ann. Rev. Biochem. 75, 333-366.

Chiti F, Bucciantini M, Capanni C, Taddei N, Dobson CM, Stefani M. (2001). Solution conditions can promote formation of either amyloid protofilaments or mature fibrils from the HypF-N-terminal domain. Protein Science 10, 2541-2547.

Cravello L, Lascoux D, Forest E. (2003). Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. Rapid Commun Mass Spectrom. 17, 2387-2393.

DuBay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M. (2004). Prediction of the absolute aggregation rates of amyloidogenic polypepide chains. J. Mol. Biol. 341, 1317–1326.

Ebeling W, Hennrich N, Klockow M, Metz H, Orth HD, Lang H. (1974) Proteinase K from Tritirachium album Limber. Eur J Biochem. 47, 91-97.

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, E. E. & Stöcker, W., eds), pp. 257–284, Springer Verlag, Heidelberg.

Fontana A, Polverino de Laureto P, Spolaore B, Frare E, Picotti P, Zambonin M (2004) Probing protein structure by limited proteolysis. Acta Biochim. Pol. 51, 299-321.

Fruton JS (1970). The specificity and mechanism of pepsin action. Advan. Enzymol. Relat. Areas Mol. Biol. 33, 401-443.

Gill SG and von Hippel PH (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182, 319-326.

Li A, Sowder RC, Henderson LE, Moore SP, Garfinkel DJ, Fisher RJ. (2001) Chemical cleavage at aspartyl residues for protein identification. Anal Chem. 73, 5395-5402.

Marcon G, Plakoutsi G, Canale C, Relini A, Taddei N, Dobson CM, Ramponi G, Chiti F. (2005) Amyloid formation from HypF-N under conditions in which the protein is initially in its native state. J Mol Biol. 347, 323-335.

Monti M, Garolla di Bard BL, Calloni G, Chiti F, Amoresano A, Ramponi G, Pucci P. (2004). The regions of the sequence most exposed to the solvent within the amyloidogenic state of a protein initiate the aggregation process. J Mol Biol. 336, 253-262.

Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM. (2005). Prediction of “aggregation-prone” and “aggregation-susceptible” regions in proteins associated with neurodegenerative diseases. J. Mol. Biol. 350, 379-392.

Relini A, Torrassa S, Rolandi R, Gliozzi A, Rosano C, Canale C, Bolognesi M, Plakoutsi G, Bucciantini M, Chiti F, Stefani M. (2004) Monitoring the process of HypF fibrillization and liposome permeabilization by protofibrils. J Mol Biol. 338, 943-957.

Rosano C, Zuccotti S, Bucciantini M, Stefani M, Ramponi G, Bolognesi M (2002). Crystal structure and anion binding in the prokaryotic hydrogenase maturation factor HypF acylphosphatase-like domain. J. Mol. Biol. 321, 785-796.

Roseman MA. (1988). Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J Mol Biol. 200, 513-522.

Schägger H, von Jagow G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem. 166, 368-379.

Street AG, Mayo SL. (1999). Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. Proc Natl Acad Sci U S A. 96, 9074–9076.

CHAPTER 3.
THE STRUCTURES OF AMYLOID FIBRILS.

Interest in amyloid fibrils within the biochemical and biophysical research communities arises from fundamental questions regarding the nature of the interactions that make amyloid fibrils a stable structural state for polypeptide chains. A defining characteristic of amyloid fibrils is the presence of the cross-beta structural motif, originally revealed by X-ray fiber diffraction (Serpell, 2000), in which ribbon-like beta-sheets, extending over the length of the fibril, are formed by beta-strands that run nearly perpendicular to the long axis of the fibril, with backbone hydrogen bonds that run parallel to the long axis (see § 1). Apart from the existence of the cross-beta motif, little was known until recently about the molecular-level structures of amyloid fibrils. Many questions are still difficult to answer (Tycko, 2004). Does the cross-beta motif contain a well-ordered pattern of intermolecular hydrogen bonds? Do peptides and proteins have well-ordered conformations in amyloid fibrils? Which peptide segments participate in the cross-beta motif? Do amyloid fibrils contain non-beta secondary structures? To what extent do amyloid fibrils formed by different peptides share a common molecular structure?

Information about the molecular structures of amyloid fibrils will lead to a better understanding of the intermolecular interactions that stabilize these structures is likely to provide important clues about the mechanisms of fibril formation and may facilitate the development of therapeutic strategies for amyloid diseases.

Over the past several years, considerable progress has been made towards the elucidation of the molecular structures of amyloid fibrils. Solid-state NMR methods have proven to be particularly valuable as direct structural probes of amyloid fibrils, because these methods can provide constraints on interatomic distances and torsion angles at a site-specific level in noncrystalline materials with complex chemical structures. Important recent contributions have also been made by electron paramagnetic resonance (EPR) spectroscopy, EM, X-ray and neutron scattering and biochemical methods, like hydrogen/deuterium exchange and proteolysis.

The observation by solid-state NMR of both parallel and antiparallel beta-sheets in amyloid fibrils means that there is no absolutely universal molecular structure for amyloid fibrils. To date, antiparallel beta-sheets have only been demonstrated in amyloid fibrils composed of relatively short peptides. However, the possibility of

antiparallel beta-sheets in fibrils formed by longer peptides or proteins certainly exists. Topological constraints imposed by intrachain and interchain disulphide bridges, present in cases such as insulin fibrils but absent in other cases, further suggest that significant variations in molecular structure must exist. Although amyloid fibrils do not share a single common molecular structure (apart from the defining characteristic that a cross-beta motif be present), it may eventually prove possible to group amyloid fibrils into classes based on common structural and morphological features.

Substantial progress has been made towards the elucidation of the molecular structures of amyloid fibrils through the development and application of novel and diverse experimental methods. Although the high-resolution molecular structure of an amyloid fibril (including both the molecular conformation and the supramolecular organization) has not yet been completely determined from experimental data, it appears likely that complete structure determination will be achieved in the near future, based on some combination of solid-state NMR, EM, EPR, X-ray and neutron scattering, and biochemical measurements. Further studies will probably lead to an improved understanding of the interactions that stabilize amyloid structures, the mechanisms by which amyloid fibrils form, and the potential technological utility of amyloid structures. Further structural studies will also most likely yield new insights into the aetiology and treatment of amyloid diseases.

Comparison of the information about the structural properties of various fibrillar systems allows to draw a number of conclusions about their similarities and differences. Different fibrils clearly have many properties in common, including the canonical cross-beta structure and the frequent presence of repetitive hydrophobic or polar interactions along the fibrillar axis. The ubiquitous presence of beta-structure strongly supports the view that the physicochemical properties of the polypeptide chain are the major determinants of the fibrillar structure in each case. Moreover, several of the proposed structures, despite very different sequences of their component polypeptides, suggest that the core region is composed of two to four sheets that interact closely to each other. An interesting feature of these sheets is that they appear to be much less twisted than expected from the analysis of the short arrays of beta-strands that form beta-sheets in globular protein structures. This feature was first proposed from cryo-EM and has been supported by Fourier transform infrared (FTIR) analyses (Jimenez et al., 1999; Zandomeneghi et al., 2004).

Nevertheless, it is clear that there are significant differences in detail attributable to the influence of the side chains on the structures adopted by the various systems. These appear to include the lengths of the beta-strands and whether they are arranged in a parallel or antiparallel arrangement within each sheet; the lengths and the conformational properties of the loops, turns, and other regions that are not included within the core structure; and the number of beta-sheets in the protofilament. It is clear that the fraction of the residues of a polypeptide chain that are incorporated in the core structure can vary substantially (e.g., from all the residues of the 7-mer peptide to only about 13 % of the residues in the full-length HET-s) and that the exact spacing between the beta-sheets varies with factors such as steric bulk of the side chains that are packed together in the core (Fandrich and Dobson, 2002). In addition the presence of disulfide bonds in proteins such as insulin may perturb the way in which the sheets can stack together (Jimenez et al., 2002).

The structure that will normally be adopted in the fibrils will be the lowest in free energy and/or the most kinetically accessible. What is clear, therefore, is that the interactions of the various side chains with each other and with solvent are crucial in determining the variations in the fibrillar architecture even though the main-chain interactions determine the overall framework within which these variations can occur. In other words, the interactions and conditions involving the side chains in a given sequence can tip the balance between the alternative “variations on a common theme” arrangements of a polypeptide “polymer” chain in its fibrillar structure. Such a situation contrasts with that pertaining to the native structures of the highly selected protein molecules, which are able to fold to unique structures that are significantly more stable for a given sequence than any alternative.

Even before the molecular structures of amyloid fibrils began to emerge, it was clear that significant morphological variations can exist between different fibrils formed from the same peptide or protein (Jimenez et al., 1999; Goldsbury et al., 1997). Evidence is now accumulating that such variations in morphology is linked to heterogeneity in molecular structure, i.e., in the structural positioning of the polypeptide chains within the fibrils. Conformational polymorphism has been studied in amyloid systems involving glucagon (Pedersen et al., 2005), Abeta 1-40 (Petkova et al., 2005), the yeast prion protein Ure2p (Baxa et al., 2005). All the results show that each protein sequence can form a spectrum of structurally distinct fibrillar aggregates and that kinetic factors can dictate which of these alternatives is dominant under given circumstances.

Of the many possible conformations that could be present in the amyloid core for a given protein, the specific ones that play this role will depend simply on the thermodynamic and, in many cases, the kinetic factors that are dominant under those circumstances. By contrast, natural globular proteins have been selected by evolution to fold into one specific three-dimensional structure, and the complex free-energy landscape associated with their sequences have a single and well-defined minimum, under physiological conditions, corresponding to the native state.

References

Baxa U, Cheng N, Winkler DC, Chiu TK, Davies DR, Sharma D, Inouye H, Kirschner DA, Wickner RB, Steven AC. (2005) Filaments of the Ure2p prion protein have a cross-beta core structure. J Struct Biol. 150, 170-179.

Fändrich M, Dobson CM. (2002) The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. EMBO J. 21, 5682-5690.

Goldsbury CS, Cooper GJ, Goldie KN, Müller SA, Saafi EL, Gruijters WT, Misur MP, Engel A, Aebi U, Kistler J. (1997) Polymorphic fibrillar assembly of human amylin. J Struct Biol. 119, 17-27.

Jiménez JL, Guijarro JI, Orlova E, Zurdo J, Dobson CM, Sunde M, Saibil HR. (1999) Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. EMBO J. 18, 815-821.

Jiménez JL, Nettleton EJ, Bouchard M, Robinson CV, Dobson CM, Saibil HR. (2002) The protofilament structure of insulin amyloid fibrils. Proc Natl Acad Sci U S A. 99, 9196-9201.

Kajava AV, Aebi U, Steven AC. (2005) The parallel superpleated beta-structure as a model for amyloid fibrils of human amylin. J Mol Biol. 348, 247-252.

Krishnan R, Lindquist SL. (2005) Structural insights into a yeast prion illuminate nucleation and strain diversity. Nature. 435, 765-772.

Nelson R, Sawaya MR, Balbirnie M, Madsen AØ, Riekel C, Grothe R, Eisenberg D. (2005) Structure of the cross-beta spine of amyloid-like fibrils. Nature. 435, 773-778.

Pedersen JS, Dikov D, Flink JL, Hjuler HA, Christiansen G, Otzen DE. (2006) The changing face of glucagon fibrillation: structural polymorphism and conformational imprinting. J Mol Biol. 355, 501-523.

Petkova AT, Leapman RD, Guo Z, Yau WM, Mattson MP, Tycko R. (2005) Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils. Science. 307, 262-265.

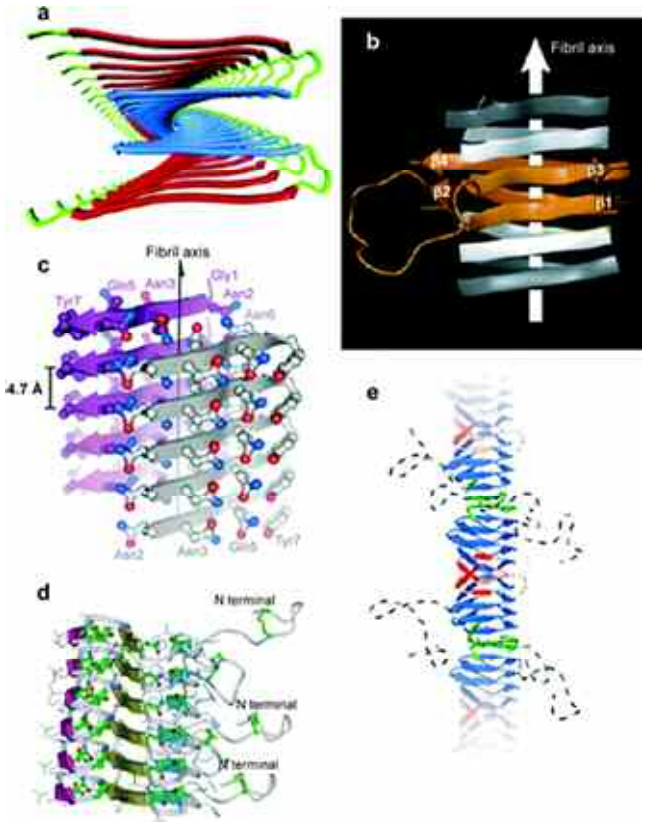
Ritter C, Maddelein ML, Siemer AB, Lührs T, Ernst M, Meier BH, Saupe SJ, Riek R (2005) Correlation of structural elements and infectivity of the HET-s prion. Nature. 435, 844-848.

Serpell LC. (2000) Alzheimer's amyloid fibrils: Structure and assembly, Biochim. Biophys. Acta 1502, 16-30.

Tycko R. (2004) Progress towards a molecular-level structural understanding of amyloid fibrils. Curr Opin Struct Biol. 14, 96-103.

Zandomeneghi G, Krebs MR, McCammon MG, Fändrich M. (2004) FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. Protein Sci. 13, 3314-3321.

BOX 3 Recent three-dimensional structural models of fibrillar aggregates from different sources (Chiti and Dobson, 2006)



(*a*) The protofilament of Aβ viewed down the long axis of the fibril (Tycko, 2004). The segments 12–24 (*red*) and 30–40 (*blue*) are shown.
(*b*) The fibril from the C-terminal domain 218–289 of the fungal prion protein HET-s [reproduced with permission (Ritter et al., 2005). The ribbon diagram shows the four β-strands (*orange*) (residues 226–234, 237–245, 262–270, and 273–282) and the long loop between β2 and β3 from one molecule. Flanking molecules along the fibril axis (*gray*) are shown.
(*c*) Atomic structure of the microcrystals assembled from the GNNQQNY peptide [reproduced with permission (Nelson et al., 2005). Each β-strand is a peptide molecule.
(*d*) The protofilament from amylin (Kajava et al., 2005). Green, yellow, and pink β-strands indicate residues 12–17, 22–27, and 31–37, respectively. The unstructured N-terminal tail is shown on the right of the panel along with the disulfide bridge between Cys2 and Cys7.
(*e*) The fibril from the NM region of Sup35p (Krishnan and Lindquist, 2005). The colored ribbons indicate residues 25–38 (*red*), 39–90 (*blue*), and 91–106 (*green*). The unstructured regions 1–20 (*red dashed lines*) and 158–250 (*black dashed lines*) are shown

3.1. Lysozyme

Lysozymes (1,4-β-N-acetylmuramidases) are a family of globular proteins with bactericidal activity, constituted by a single polypeptide chain of 127-130 amino acid residues (MW ~ 15 kDa), widely distributed in animal and plant tissues. Lysozyme performs its activity by hydrolyzing mucopolysaccharides that constitute the cell wall of Gram + bacteria. In particular, lysozyme hydrolyzes glycosidic bonds between N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) of peptidoglycans, provoking lyses of the bacterial membrane, cellular swelling and loss of cellular content.

Lysozymes from various species are highly homologous in the amino acid sequence. In particular, the position of the four disulfide bridges between Cys6-127, 30-115, 64-80 and 76-94 is conserved, since they are essential for the correct conformation of the protein and for the performance of the biological function.

The secondary structure of lysozyme from many different sources has been investigated. Generally, it possesses a high content of helical structure, with four α-helices and two little 3_{10} helices (Radford et al., 1992) and a smaller region with β-sheet structure, formed by antiparallel β-strands. In Fig. 11 the sequence and the structure of human lysozyme are reported, as, during my PhD course, many attention has been focused on human lysozyme, because of the large availability of information on its structure, dynamics and folding and of its implication in an amyloid pathological disease (Merlini and Bellotti, 2005).

In fact, in 1993 it has been demonstrated (Pepys et al., 1993) that a familial non-neurophatic amyloidosis was induced by a variant of human lysozyme. In particular, human lysozyme variants form amyloid fibrils in individuals suffering from a familial non-neuropathic systemic amyloidosis. The single-point mutants of lysozyme I56T, F57I, W64R and D67H (Pepys et al., 1993; Valleix et al., 2002; Yazaki et al., 2003) and the two double mutants F57I/T70N and W112R/T70N (Yazaki et al., 2003; Röcken et al., 2006) have been found to form amyloid plaques in individuals suffering from non-neuropathic hereditary amyloidosis. Another natural variant, T70N, has been found in 5% of the British population and 12% of the Caucasian Canadian population but has not so far been associated with disease (Yazaki et al., 2003; Booth et al., 2000). It has been shown that the I56T, F57I, W64R and D67H variants are less stable than the wild-type protein (Booth et al., 1997; Dumoulin et al., 2005; Kumita et al., 2006).

Despite its greater stability, however, wild-type human lysozyme can form fibrils *in vitro* that are very similar to those of the pathological variants; for example, when incubated at low pH and elevated temperature (Morozova-Roche et al., 2000) or following application of high hydrostatic pressure (De Felice et al., 2004). Lysozyme is an excellent system with which to study protein aggregation processes, as its structure and folding mechanism are known in great detail (Redfield and Dobson, 1990; Hooke et al., 1994). Moreover, the mechanism of aggregation of both variant and wild-type lysozymes has been investigated in depth (Dumoulin et al., 2006). It has been shown that the region encompassing the β-sheet and helix C of the amyloidogenic variants I56T and D67H undergoes a locally cooperative unfolding event under physiologically relevant conditions (Dumoulin et al., 2005; Canet et al., 2002). A proteolytic fragment encompassing approximately the same sequence region in the highly homologous hen egg-white protein has been found to be the most aggregation-prone segment of the protein, at least at acid pH (Frare et al., 2004). Moreover, synthetic fragments corresponding to the β-sheet region of hen lysozyme have been shown to form amyloid structures very readily (Krebs et al., 2000). Overall, this evidence has led to the hypothesis that lysozyme amyloid fibril formation may be initiated by the partial unfolding of the central region of the protein encompassing the β-domain and the C-helix located in the alpha-domain, to which it is linked by a disulfide bond (Booth et al., 1997; Dumoulin et al., 2005).

While many molecular features of the process of lysozyme aggregation are known from these previous studies, a detailed molecular description of the protein aggregates, including the mature fibrils, is still an important objective.

References

Booth DR, Pepys MB, Hawkins PN (2000) A novel variant of human lysozyme (T70N) is common in the normal population. Hum. Mutat. 16, 180.

Booth DR, Sunde M, Bellotti V, Robinson CV, Hutchinson WL, Fraser PE et al. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. Nature 27, 787-793

Canet D, Last AM, Tito P, Sunde M, Spencer Al, Archer DB et al. (2002) Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. Nature Struct. Biol. 9, 308-315.

De Felice FG, Vieira MN, Meirelles MN, Morozova-Roche LA, Dobson CM, Ferriera ST (2004) Formation of amyloid aggregates from human lysozyme and its disease-associated variants using hydrostatic pressure. FASEB J. 18, 1099-1101.

Dumoulin M, Canet D, Last AM, Pardon E, Archer DB, Muyldermans S et al. (2005) Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. J. Mol. Biol. 25, 773-788.

Frare E, Polverino de Laureto P, Zurdo J, Dobson CM and Fontana A (2004) A highly amyloidogenic region of hen lysozyme. J. Mol. Biol. 23, 1153–1165.

Hooke SD, Radford SE, Dobson CM (1994) The refolding of human lysozyme: A comparison with the structurally homologous hen lysozyme. Biochemistry 33, 5867-5876.

Krebs MR, Wilkins DK, Chung EW, Pitkeathly MC, Chamberlain AK, Zurdo J et al. (2000) Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the beta-domain. J. Mol. Biol. 14, 541-549.

Kumita JR, Johnson RJ, Alcocer MJ, Dumoulin M, Holmqvist F, McCammon MG et al. (2006) Impact of the native-state stability of human lysozyme variants on protein secretion by Pichia pastoris. FEBS J. 273, 711-720.

Merlini G and Bellotti V (2003) Molecular mechanisms of amyloidosis. N. Engl. J. Med. 349, 583-596.

Morozova-Roche LA, Zurdo J, Spencer A, Noppe W, Receveur V, Archer DB et al. (2000) Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants. J. Struct. Biol. 130, 339-351.

Pepys MB, Hawkins PN, Booth DR, Vigushin DM, Tennent GA and Soutar AK et al., (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis. Nature 362, 553-557

Radford SE, Dobson CM, Evans PA. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358, 302-307.

Redfield C and Dobson CM (1990) ^{1}H NMR studies of human lysozyme: Spectral assignment and comparison with hen lysozyme. Biochemistry 29, 7201-7214.

Röcken C, Becker K, Fändrich M, Schroeckh V, Stix B, Rath T et al. (2006) A Lys amyloidosis caused by compound heterozygosity in exon 2 (Thr70Asn) and exon 4 (Trp112Arg) of the lysozyme gene. Hum. Mutat. 27, 119-120.

Valleix S, Drunat S, Philit JB, Adoue D, Piette JC, Droz D et al. (2002) Hereditary renal amyloidosis caused by a new variant lysozyme W64R in a French family. Kidney Int. 61, 907-912.

Yazaki M, Farrell SA, Benson MD (2003) A novel lysozyme mutation Phe57Ile associated with hereditary renal amyloidosis. Kidney Int. 63, 1652-1657.

3.1.1. Identification of the core sequence of human lysozyme fibrils by proteolysis

In this study, fibrils of wild-type human lysozyme formed at low pH have been analyzed by a combination of limited proteolysis and Fourier-transform infrared (FTIR) spectroscopy, in order to map conformational features of the 130 residue chain of lysozyme when embedded in the amyloid aggregates. After digestion with pepsin at low pH, the lysozyme fibrils were found to be composed primarily of N and C-terminally truncated protein species encompassing residues 26–123 and 32–108, although a

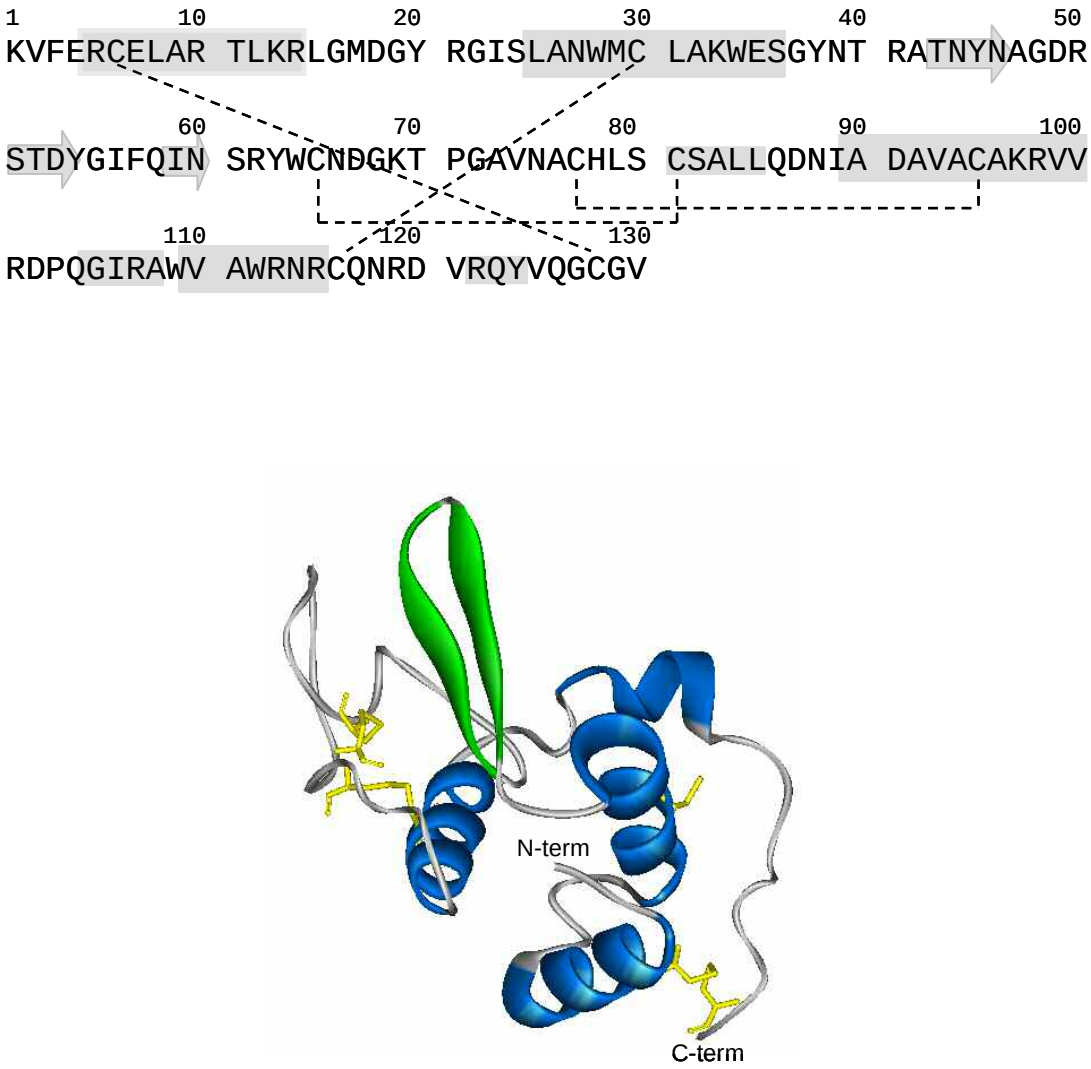


Fig. 11 Amino acid sequence and secondary structure wild type human lysozyme (HuL). *Top* Human lysozyme sequence has been derived from the ExPASy proteomic server (Identification number P61626). The dashed lines represent the four disulfide bridges connecting the eight cystein residues. Secondary structure elements are evidenced in grey shapes: big boxes correspond to alpha-helices, small boxes to 3_{10}-helices and arrows to beta-strands. *Bottom* Representation of the 3D structure of HuL. The diagram was drawn using the PDB file "1LZ1" using the program WebLab Viewer Pro 4.0 (Molecular Simulations Inc.). The chain segments corresponding to alpha-helices are colored in blue and beta strands in green. The four disulfide bonds are represented by yellow sticks.

significant minority of molecules was found to be completely resistant to proteolysis under these conditions. FTIR spectra provide evidence that lysozyme fibrils contain extensive β-sheet structure and a substantial element of non β-sheet or random structure that is reduced significantly in the fibrils after digestion. The sequence 32–108 includes the β-sheet and helix C of the native protein, previously found to be prone to unfold locally in human lysozyme and its pathogenic variants. Moreover, this core structure of the lysozyme fibrils encompasses the highly aggregation-prone region of the sequence recently identified in hen lysozyme. The present proteolytic data indicate that the region of the lysozyme molecule that unfolds and aggregates most readily corresponds to the most highly protease-resistant and thus highly structured region of the majority of mature amyloid fibrils. Overall, the data show that amyloid formation does not require the participation of the entire lysozyme chain. The majority of amyloid fibrils formed from lysozyme under the conditions used here contain a core structure involving some 50% of the polypeptide chain that is flanked by proteolytically accessible N and C-terminal regions.

The pepsin-resistant segment 32–108 encompasses the β-sheet and helix C of native lysozyme and corresponds to the region, approximately residues 31–104 that readily undergoes local unfolding in the amyloidogenic variants I56T and D67H of human lysozyme (Dumoulin et al., 2005; Canet et al., 2002) and in the wild-type protein under more extreme conditions (Johnson et al., 2005). Moreover, this region also encompasses fragment 57–107 that was found to be highly amyloidogenic in hen lysozyme (Frare et al., 2004). In addition, a camel antibody fragment that binds to 14 residues located in the loop between helices A and B, in the long loop within the β-domain and in the C-helix was found to be able to restore the global cooperativity of mutant lysozymes I56T and D67H, therefore preventing the aggregation process of these mutants (Dumoulin et al., 2003, 2005). These data suggest that the region involved in the cooperative local unfolding of the molecule, and which is also the most amyloidogenic part of the sequence, is a key factor in determining the specific structure of the lysozyme fibrils. Interestingly, proteolysis experiments conducted at 45 °C on monomeric human lysozyme at pD 1.5; on the conformational state of the protein that is the precursor to the fibrils, show that the first peptide bonds that are cleaved by pepsin are Phe57-Gln58 and Ala108-Trp109, located in the β-domain and in the 3_{10} helix between helices C and D, respectively (not shown). These results support the conclusion that unfolding of these segments of the chain triggers the formation of the amyloid

structure by human lysozyme. Indeed, it is very likely that partial unfolding of a globular protein will often be a key factor in initiating the aggregation process. Protein destabilization, by the addition of a denaturant, low pH, high temperature or amino acid substitution, can lead to an increased population of partially unfolded monomeric protein molecules that are key species in the initiation of fibril formation. Such partially unfolded conformations presumably enable the specific intermolecular interactions that are necessary for protein oligomerization and fibrillation (Booth et al., 1997; Uversky and Fink, 2004; Dobson, 2003).

The incomplete digestion of lysozyme fibrils by pepsin can be explained on the basis that the non-β-sheet segments of specific regions of the molecules in the fibrils are so tightly packed in the structure that they are not accessible to the protease. There is also evidence that the fibrils exhibit a significant degree of polymorphism, as a population of protein molecules (approximately 40%) appears to be completely resistant to proteolysis and, therefore, could in principle have their entire polypeptide chain involved in the highly organized core structure of the fibrils. This hypothesis appears to be unlikely, however, as the process of aggregation of the protein was conducted whilst stirring the solution, and electron microscopy suggests that the fibrils are relatively homogeneous, at least in terms of their morphology. In addition, the FTIR spectrum of the proteolyzed fibrils indicates the presence of relatively unstructured regions that have not been cleaved by the protease. This finding suggests instead that pepsin may not be able to gain access to all the disorganized or loop regions within the fibrils, perhaps because they are buried in the rigid structure of the fibril, by the packing of the constituent protofilaments. Interestingly, the inaccessibility of approximately 50% of flexible regions of proteins in amyloid fibrils has been seen elsewhere (Kheterpal et al., 2000; Baldwin et al., 2006), a fact that could be related to the idea that a double sheet could be a common building block of the amyloid structure (Nelson et al., 2005; Dobson, 2005).

Moreover, this study emphasizes further the value of limited proteolysis as a technique that can lead to useful structural information even for high molecular mass protein complexes, such as amyloid aggregates, and in this specific case provides important information about the core structure of the lysozyme fibrils.

Published paper (Frare et al., J Mol Biol. 2006, 361, 551-561)

3.1.2. Conformational polymorphism of lysozyme fibrils depending on solution conditions.

As mentioned before (§ 3.1) some of the natural variants of human lysozyme are responsible for the formation of amyloid plaques in vivo, in a familial non-neuropathic systemic amyloidosis. In order to study the process of amyloid aggregation by these human lysozyme variants, many studies have been conducted on wild type human lysozyme (HuL), exploiting the available wealth of information on structure, dynamics and folding of lysozyme (Matagne et al., 2000). In order to give insights into this pathological process, many experimental conditions have been developed in order to make also wild type human lysozyme form fibrils quite similar to the pathological ones, knowing that amyloid formation begins from the population of a partially unfolded state of a protein. Most of these studies, however, have been carried out under acidic conditions, since it is well known that the protein conformation that is prone to aggregate is a partially unfolded state, an easily accessible condition for HuL at pH 2.0 and high temperature. HuL variants, however, aggregate in a physiological environment, roughly at pH 7-7.5 at 37 °C, because of their innate instability (Dumoulin et al., 2005). It is possible also to force HuL to aggregate in conditions more similar to the pathological ones, presumably neutral pH and 37 °C, and observe if the same HuL amino acid chain can be able to form amyloid fibrils in such different conditions, and try to compare the structure and the stability of the produced fibrils. In the following study, fibrils made of wild type human lysozyme have been formed in two very different conditions, in each case destabilizing and promoting the interactions between lysozyme molecules. Lysozyme fibrils were produced at acidic and at neutral pH, leading both to the formation of fibrils having the three hallmarks of amyloid material, that are cross-beta structure, binding of ThT and an overall amyloid fiber morphology, but that, in a more detailed analysis, revealed a completely different structure. Fibrils have been studied by means of ANS binding, FTIR and X-ray fiber diffraction in order to characterize the differences in the structure. Guanidinium-induced fibrils dissociation, instead, has been applied in order to test the chemical stability of the two kinds of fibrils. The results clearly indicate that the solution conditions used for lysozyme aggregation promote the formation of amyloid fibrils with different structure and stability, due to the different rearrangements of the lysozyme polypeptide chain into the fibril structure.

Many studies have been conducted on different amyloid aggregates formed under different conditions with peptides, such as Aβ (Petkova et al., 2005). Studying these processes with a full-length protein increases the complexity of the problem, since the dense packing reachable in amyloid fibrils made of peptides (10-40 residues) could not be accomplished in all the amino acid residues of a full-length protein, except in the core regions (Chatani and Goto, 2005). Studying different amyloid fibrils conformations from HuL is more challenging because it is a 130 amino acid chain with the structural constrains given by the four disulfide bridges present in the lysozyme molecule. This study can also give some insights into the complex problem of strains diversity, such as for prion diseases, helping the clarification of the structural principles of amyloid fibrils which can produce multiple and distinct amyloid conformations from one protein sequence.

Materials and methods

Materials

Wild-type human lysozyme (HuL) was expressed as described (...). Proteinase K, ANS and thioflavin-T (ThT) were purchased from the Sigma Chem. Co. (St. Louis, MO). All other chemicals were of analytical reagent grade and were obtained from Sigma or Fluka (Buchs, Switzerland).

Methods

Formation of human lysozyme amyloid fibrils at pH 2.0 and pH 7.5.

Fibrils at pH 7.5 were prepared dissolving HuL in 50 mM Na_2HPO_4 pH 7.5 at a concentration of 0.7 mM and stirring this solution at 60°C for 1 day. Fibrils at pH 2.0 were prepared by seeding. An aliquot of HuL fibrils previously formed at pH 2.0 was added to a HuL solution (1 mM) in 10 mM HCl pH 2.0 in ratio of 2 % (w/w). The suspension was then left at 47°C under stirring for up to 6 days.

Optical spectroscopy

Protein concentrations were evaluated from absorption measurements at 280 nm on a single-beam Cary 400 Scan spectrophotometer (Varian, Palo Alto, CA, USA). The extinction coefficient of full-length human lysozyme at 280 nm, calculated with the method of Gill & von Hippel, was 2.5 ml mg^{-1} cm^{-1}. Fluorescence measurements were

carried out on a Varian (Palo Alto, CA, USA) model Cary Eclipse spectrofluorimeter in a temperature-controlled cell holder, utilizing a 2 mm x 10 mm path length cuvette. For each measurement a protein concentration of 2.4 μM (0.013 mg/ml) was used. The ThT binding was monitored by exciting the sample at 440 nm and recording the emission fluorescence spectrum from 450 to 600 nm. For every measurement, 25 μl of a ThT stock solution (2.5 mM ThT in 10 mM Phosphate buffer pH 7.0 containing 150 mM NaCl) were added to a volume of fibrils corresponding to 60 μgr and a volume of 1.5 ml was reached with ThT buffer. The ANS binding was monitored by exciting the sample at 350 nm and recording the emission fluorescence spectrum from 360 to 680 nm, using a protein to ANS molar ratio of 1:30 at pH 2.0 and of 1:150 for pH 7.0 (Semisotnov et al., 1991). ANS was prepared in water and the concentration was evaluated from absorption measurement at 350 nm, using an extinction coefficient at 350 nm of 4950 cm^{-1} M^{-1}. Thermal unfolding was followed both by Trp fluorescence, by exciting the protein solution at 280 nm and recording the fluorescence emission intensity at 360 nm rising the temperature from 10 to 90 °C, by ANS binding, by exciting the protein solution containing ANS at 350 nm and monitoring the ANS fluorescence emission at 475 nm and by CD, following the dichroic signal at 222 nm. Samples were analyzed in a Bruker BioATRCell II using a Bruker Equinox 55 Fourier transform infrared spectroscopy (FTIR) spectrometer (Bruker Optics Limited, UK) equipped with a liquid nitrogen cooled mercury cadmium telluride (MCT) detector and a silicon internal reflection element (IRE). For each spectrum 256 interferograms were coadded at 2 cm^{-1} resolution, and the water background was independently measured and subtracted from each protein spectrum.

Transmission Electron Microscopy

Samples were applied to Formvar-coated nicked grids, stained with 2% (w/v) uranyl acetate solution and viewed in a Phillips CEM100 transmission electron microscope operating at 80 kV.

X-ray fibre diffraction

Amyloid fibril samples were prepared for X-ray diffraction analysis using a modification of the stretch-frame method (Serpell et al., 2000). In this modification, the distance between capillary tubes is not changed during the production of the semi-aligned, dry stalk. A 10μl droplet of fibril suspension was placed between two

horizontal wax-plugged glass capillary tubes of 1mm diameter held approximately 2mm apart. Upon drying of the droplet to approximately a quarter of its initial volume, a further 5µl of fibril suspension was added and the process was repeated a further three times. Subsequently, the droplet was left to dry completely in air. All scattering patterns were obtained on a crystallography beamline of the Biochemistry Department at Cambridge University. X-rays with a wavelength of 1.54Å were produced by a rotating copper anode, and collimated and focused by Osmic Max-flux optics. Images were acquired on a Marr image plate. The sample–detector distance was 300mm, and data acquisition times were typically 15min. 'Images were analysed and radially integrated to generate 1D scattering patterns using Fit2D (http://www.esrf.fr/computing/scientific/FIT2D) to obtain accurate reflection positions

Denaturation of fibrils by Guanidine salts

The denaturation of HuL fibrils, that is the degree of depolymerization, was measured both at pH 2.0 and pH 7.5 at 25 °C. The process was studied by diluting aliquots of the fibrils into buffered solutions containing increasing concentrations of GdnHCl or GdnSCN, from 0 to 7 M. After 48 h at 25 °C, the samples were ultracentrifuged at 90000 rpm, 20°C for 45 min: the protein concentration in the supernatants was measured recording the absorbance at 280 nm in presence of GdnHCl, and using the Bradford assay in presence of GdnSCN. The necessity of using the Bradford assay was due to the absorbance in the UV region of GdnSCN. For the Bradford assay, the protein solutions were diluted in the Bradford solution, left for 30min at room temperature, and the measured absorbance at 595 nm was compared with a calibration curve previously built with human lysozyme solutions. The protein concentration values were plotted against the denaturant concentration.

Results

Formation of full-length HuL fibrils at pH 2.0 and at pH 7.5

Thermal stability experiments performed on HuL followed by Trp fluorescence showed that the melting temperature of HuL is 57°C at pH 2.0 and 70°C at pH 7.5, implying a fairly higher stability under neutral conditions. Since it has been previously shown (Frare et al., 2006) that even below the melting temperature amyloid aggregation can be induced by stirring the protein solution, the same rationale has been applied to fibrils formation at neutral pH. The aim was to avoid too high temperatures responsible

in many cases of protein degradation processes such as peptide bonds breakage, deamidation and scrambled disulfide bridges (Bischof and He, 2005). Comparing the population of HuL intermediate species binding the hydrophobic fluorophore ANS (Fig. 13) during thermal unfolding at pH 7.5 (triangles) and pH 2.0 (circles), it was evident the larger population of HuL partially folded states is present at 57°C at pH 2.0 and at 71°C at pH 7.5. Interestingly, at pH 7.5 and 60°C the amount of HuL molecules binding ANS is about one third of the maximum level reached at 71°C and pH 7.5.

In order to produce HuL fibrils under neutral conditions, a 0.7 mM HuL solution at pH 7.5 was stirred at 60°C. Aggregation at neutral pH was monitored by following the increase in ThT fluorescence emission of aliquots removed from the reaction mixture at different time-points. The experimental data points fit to a sigmoidal curve, indicating the presence of a lag phase of about 5 hours, followed by a rapid growth phase that reaches a plateau after 24 hours (Fig. 14A). After 1 day, the solution appeared white and the ThT assay confirmed the presence of amyloid fibrils. After ultracentrifugation of the suspension and measurement of the protein concentration of the supernatant, it was found that the fibrils correspond to the 99 % of the protein solution. SDS-PAGE analysis of the fibrils sample confirmed that they are made of full-length HuL. Full-length HuL fibrils were formed at pH 2.0 by adding a sample of preformed fibrils to a 1 mM HuL solution at pH 2.0, corresponding to the 2 % (w/w) of the total amount of protein in the solution. The protein sample, then, was left at 47°C under stirring for 6 days, after which a dense suspension was clearly formed. The kinetics of aggregation was monitored by following the increase in ThT fluorescence emission of aliquots removed from the reaction mixture at different time-points. The experimental data points evidence the lack of a lag phase, due to the seeding effect (Fig. 14B). To verify the integrity of fibrillar HuL, an aliquot of fibrils isolated by ultracentrifugation was analyzed by SDS-PAGE, confirming that even if the seeding was composed by some nicked HuL, the newly obtained fibril are made of full-length HuL.

The amyloidogenic nature of fibrils produced under the two described conditions was confirmed also by TEM, showing in both cases unbranched fibrils characterized at pH 7.5 (Fig. 14C) by a short single-fibre morphology and a thinner diameter (3-5 nm) and at pH 2.0 by an inter-twisted arrangement and a diameter of 8-10 nm (Fig. 14D).

Conformational characterization of fibrils

ANS binding assay (Fig. 15), used to recognize exposed hydrophobic surfaces on proteins, revealed that, while the monomer at both the pH values does not bind ANS at all, both kinds of fibrils bind ANS strongly, giving rise to a blue shift of the maximum fluorescence wavelength and an increase in the fluorescence intensity. The fact that the ANS fluorescence intensity value in the presence of fibrils formed at pH 2.0 is higher with respect to fibrils at pH 7.5, can arise from a larger exposed hydrophobic surface or from the fact that ANS has an intrinsic greater fluorescence in the presence of hydrophobic surfaces at acidic conditions rather than neutral pH conditions (Semisotnov et al., 1991). These data confirm that both the fibrils possess hydrophobic surfaces exposed to the solvent. The same experiments were performed also on insulin fibrils and on fibrils formed by the TTR fragment 105-115 both formed at pH 2.0, and the results evidenced that while insulin fibrils bind ANS, TTR fibrils do not bind it (Fig. 15C). Since we know from the literature that insulin fibrils contain non beta regions (Jimenez et al., 2002) while TTR 105-115 are fully arranged in the beta-sheet structure (Jaroniec et al., 2004), it can be concluded that ANS can be a useful tool to detect hydrophobic patches exposed on the fibrils surface and not included in the cross-beta core of fibrils, and that both kinds of HuL fibrils have hydrophobic regions exposed to the solvent.

Analysis of fibrils structure by ATR-FTIR (Fig. 16) showed substantial differences between the secondary structure components of the two kinds of fibrils. The spectra obtained for the two kinds of fibrils were both subjected to curve-fitting in order to determine the secondary structure content. In Fig. 16, the dashed lines represent the spectra obtained from the protein samples. Black lines show the bands corresponding to secondary structure elements, as derived from the curve-fitting of the spectra. Grey lines represent the signals arising from the side chains in the spectra. For the evaluation of the secondary structure content of fibrils, the bands corresponding to side chains absorption have not been considered, and the sum of the areas of the other bands, corresponding to secondary structure elements, have been assigned the value of 100%. IR spectrum of fibrils formed at neutral pH (Fig.16, top) is characterized by three bands corresponding to the aggregated beta-sheet structure, centred on 1620, 1636 and 1692 cm^{-1}, giving a percentage of fibrillar structure of 32%. Bands corresponding to random and alpha-helix structure (1646 cm^{-1}) and turns and bends (1662 cm^{-1}) are present at a lower extent, respectively the 23% and 26% of the total area. Additional bands corresponding to

antiparallel beta-sheet are found to be centred on 1636 cm^{-1} and 1678 cm^{-1}, (see Table III). On the counterpart, fibrils formed at acidic pH have a spectrum typical of amyloid fibrils with a major band assignable to the cross-beta structure centred on 1622 cm^{-1} (Fig. 16, bottom), corresponding alone to the 65% of the total area. Other bands relative to aggregated beta-sheet are centered on 18684 and 1693 cm^{-1}, giving an overall percentage of aggregated beta-sheet of about 70%. Bands at 1636 and 1671 cm^{-1} correspond to antiparallel beta-structure (14%). Other bands, corresponding to random or alpha-helix (1652 cm^{-1}) and turns and bends (1671 cm^{-1}), represent the 15% and the 1 % of the total area respectively.

The great difference in the beta-sheet arrangement between the two kinds of fibrils is confirmed also by the X-ray diffraction patterns. Indeed, while the meridional reflection is at 4.8 Å for both kinds of fibrils, the equatorial reflection is at 10.4 Å for fibrils at pH 2.0 while fibrils at pH 7.5 present a double signal, at 9.4 and 11.0 Å (Fig. 17). The meridional reflection, in fact, corresponds to the hydrogen bonding distance between the beta-strands that constitute the beta-sheet and, depending on the invariant geometry of the polypeptide backbone, differs only marginally for different amyloid fibrils. The equatorial reflection, instead, corresponds to the spacing between the beta-sheets, and hence can vary depending on the amino acid residues and on the polypeptide sequence involved in amyloid fibril formation.

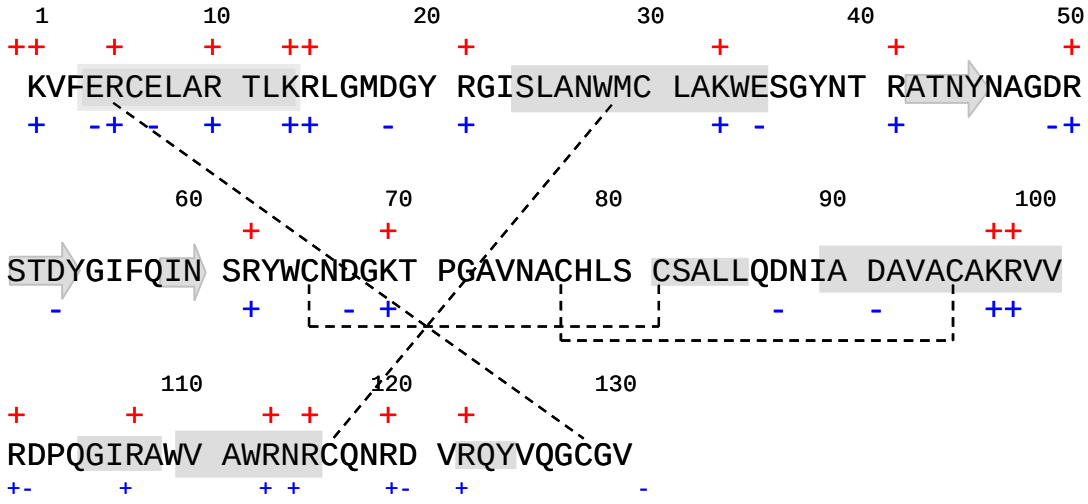


Fig. 12. Amino acid sequence and charge distribution of wild type human lysozyme (HuL) at pH 7.5 and pH 2.0. Human lysozyme sequence has been derived from the ExPASy proteomic server (Identification number P61626). The dashed lines represent the four disulfide bridges connecting the eight cystein residues. Secondary structure elements are evidenced in grey shapes: big boxes correspond to α-helices, small boxes to 3_{10}-helices and arrows to β-strands. The symbols + and – indicate the charges present on HuL molecule at pH 2.0 (red) and at pH 7.5 (blue).

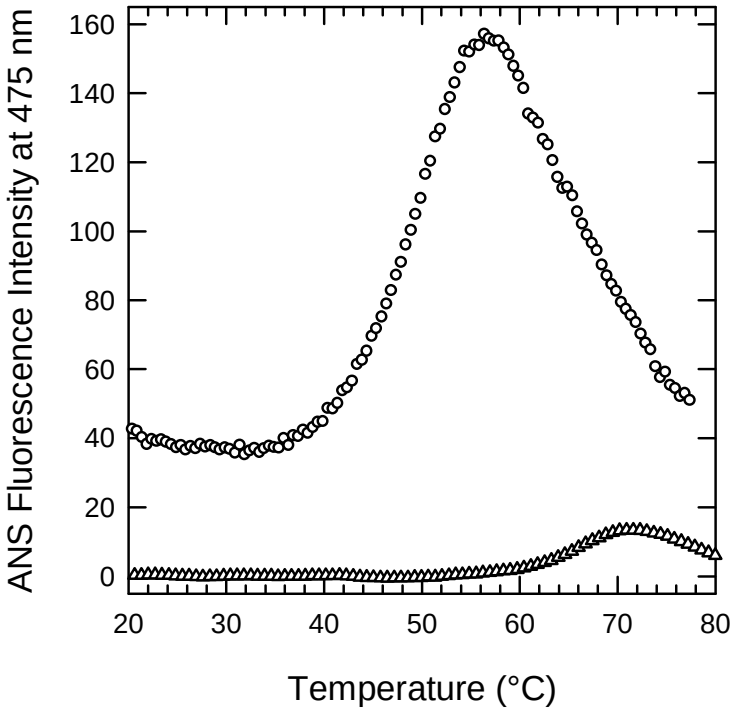


Fig. 13. Thermal denaturation of HuL at pH 7.5 (triangles) and pH 2.0 (circles) in the presence of ANS. ANS fluorescence emission during thermal denaturation of HuL at pH 7.5 (50mM Na_2HPO_4) and pH 2.0 (10mM HCl). All samples were performedwith 2.4 μM protein and 360 μM and 72 μM ANS at pH 7.5 and pH 2.0, respectively.

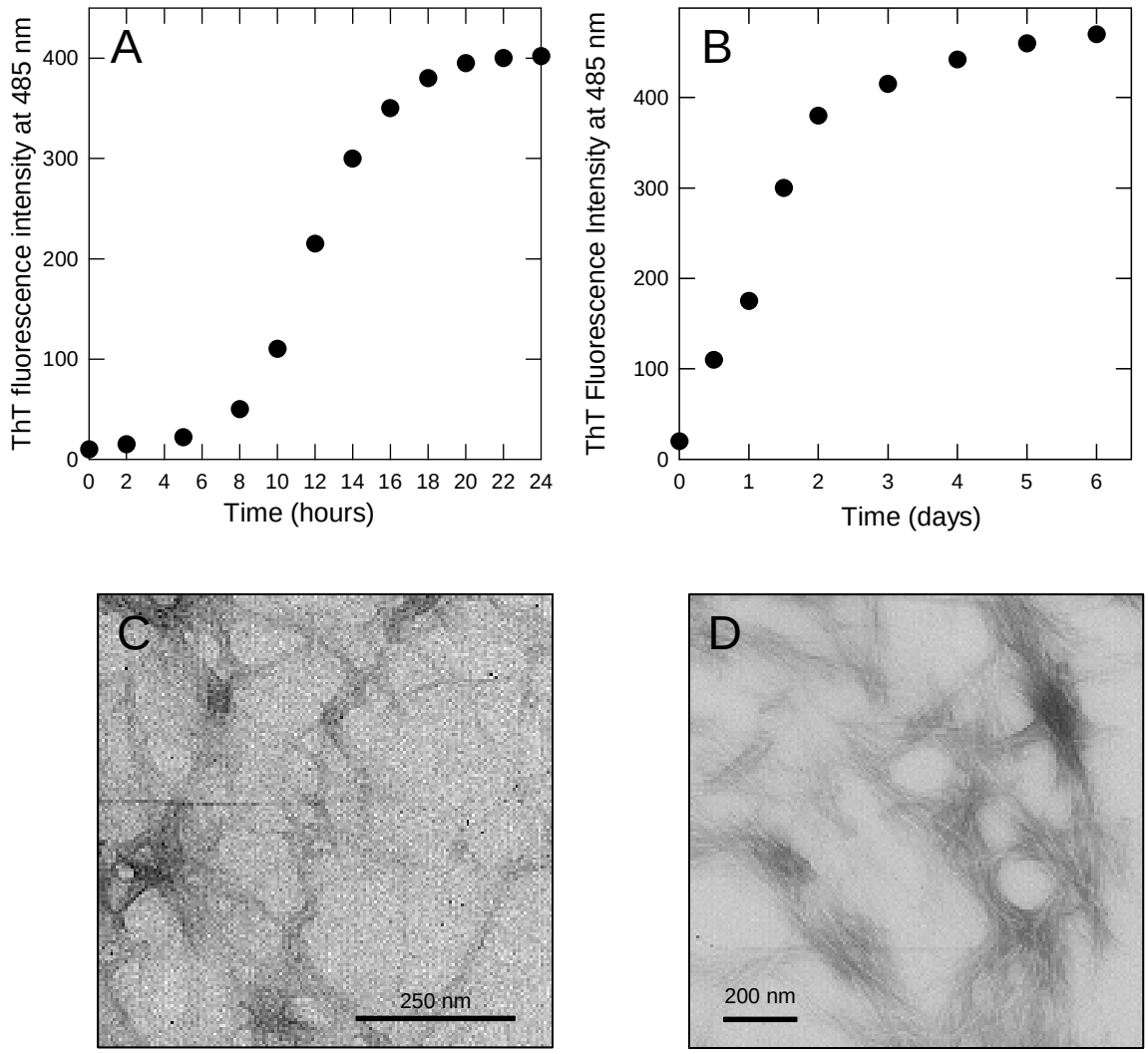


Fig. 14. Kinetics of human lysozyme amyloid aggregation at pH 2.0 and pH 7.5. HuL aggregation processes at pH 7.5 (A) and pH 2.0 (B) were followed by ThT binding assay. Electron micrographs of the fibrils formed at pH 7.5 (C) and pH 2.0 (D) were taken on samples after 24 hours and 6 days, respectively.

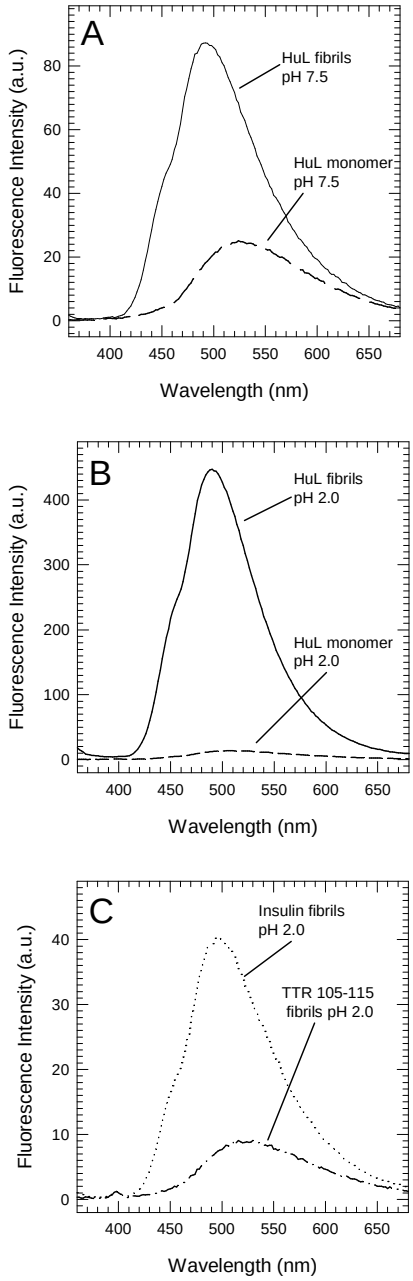


Fig. 15 Fluorescence emission spectra of ANS in the presence of HuL fibrils and monomer at pH 7.5, respectively in A solid and dashed line, at pH 2.0, respectively in B solid and dashed line, and of insulin fibrils and TTR 105-115 fibrils formed both at pH 2.0, respectively in C dotted and dashed line. All the spectra were acquired at 20°C. The excitation wavelength was 360 nm, and the excitation and emission slit widths were both set to 5 nm. For all the measurements protein concentration was 2.4 mM and ANS was added at a ANS:protein ratio of 1:30 at pH 2 and 1:150 at pH 7.5.

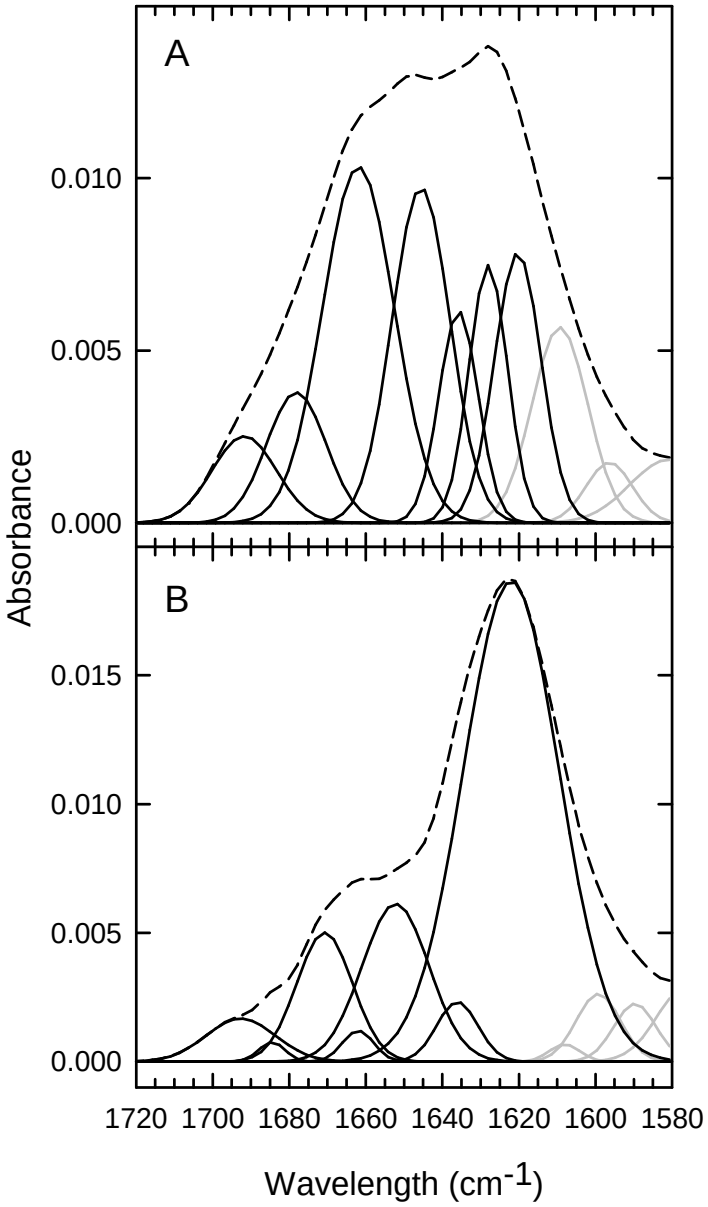


Fig. 16. ATR-FTIR analysis of lysozyme fibrils formed at pH 7.5 (A) and at pH 7.5 (B). The spectra are reported as dashed lines, while solid lines corresponds to the components arising from curve fitting. Dark solid lines represent bands corresponding to secondary structure elements, while grey lines are relative to bands corresponding to side chains signals. These latter have been excluded from the calculation of the areas of the spectra and the subsequent measurement of the relative percentage of secondary structure components.

Table III Secondary structure content of HuL fibrils as determined by curve fitting of ATR-FTIR spectra shown in Fig. 16.

HuL fibrils pH 7.5			
Wavenumber[a] (cm^{-1})	*%[b]*		*Structural assignment*
1620	15	26	Aggregated β-sheet
1628	11		
1636	10		Antiparallel β-sheet
1646	23		Random/α-helix
1662	26		Turns/bends
1678	9		Antiparallel β-sheet
1692	6		Aggregated β-sheet

HuL fibrils pH 2.0			
Wavenumber[a] (cm^{-1})	*%[b]*		*Structural assignment*
1622	65		Aggregated β-sheet
1636	4		Antiparallel β-sheet
1652	15		Random/α-helix
1662	1		Turns/bends
1671	10		Antiparallel β-sheet
1684	1	5	Aggregated β-sheet
1693	4		

[a] Peak position of the amide I band components, as deduced by the second derivative spectra.

[b] Percentage area of the amide I band components, as obtained by integrating the area under each deconvoluted band. The areas corresponding to side chain contributions located at 1580-1610 cm^{-1} have not been considered.

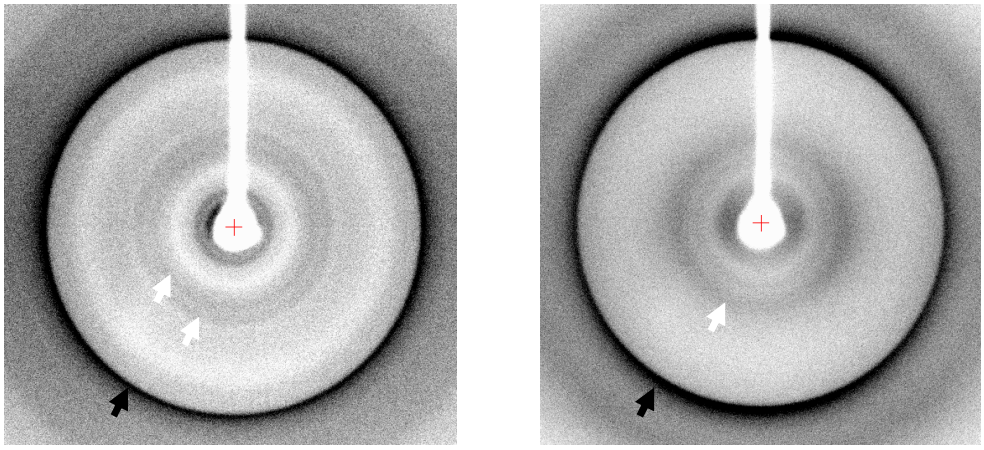


Fig. 17. X-ray fibre diffraction pattern of the fibrils formed at pH 7.5 (left) and at pH 2.0 (right). Both the kinds of fibrils show a prominent meridional reflection at 4.8 Å (black arrows), typical of inter-beta strand distance in the fibrils. The equatorial reflections of the two patterns (white arrows), corresponding to the inter-beta sheet distance, instead, evidence different values: at 10.4 Å for fibrils formed at pH 2.0, and at 9.45 Å and at 11.0 Å for fibrils formed at pH 7.5.

Chemical denaturation of fibrils

The dissolution of HuL fibrils was followed both at pH 2.0 and pH 7.5 at 25 °C. The process was studied using as denaturants both GdnHCl and GdnSCN. The latest was chosen because of its greater denaturing ability, since it was immediately evident that while fibrils formed at pH 7.5 are completely dissolved in 7 M GdnHCl (Fig. 6A), fibrils formed at pH 2.0 need stronger conditions to be fully dissociated (Fig. 6B). Using GdnSCN as denaturing agent, instead, fibrils formed at pH 7.5 were considerably dissociated also with low concentration of denaturant, and no pre-transition region was found (Fig. 6C). On the other hand, GdnSCN-induced dissociation of fibrils formed at pH 2.0 perfectly worked (Fig. 6D). The midpoints for dissociation curves of HuL fibrils formed at pH 2.0 and pH 7.5 are 2.4 M GdnHCl and 2.9 M GdnSCN respectively.

Main conclusions

Typical conditions exploited to produce HuL fibrils were acidic environment and high temperature, in order to destabilize sufficiently HuL structure (Morozova-Roche et al., 2000). In the present work, it has been shown that HuL fibrils can be obtained also at neutral pH. In the aggregation process the solution stirring is a crucial factor, since it is known that agitation promotes a more efficient interaction between molecules in solution (Chatani and Goto, 2005). The most relevant effect of agitation during fibrillogenesis involves the interaction between partially unfolded monomers producing nuclei from which fibrils grow.

HuL fibrils formed at neutral pH apparently share the same properties of those obtained from HuL at acidic pH, since they bind ThT, they have cross-beta structure and fibrillar morphology. However, a substantial difference in the beta-sheet content and rearrangement at the level of the fibril core has been demonstrated. Indeed, X-ray fiber diffraction evidenced that in fibrils formed at neutral pH the inter-sheet distance is more heterogeneous compared to fibrils formed at acidic pH. ATR-FTIR, moreover, confirmed that fibrils formed at pH 7.5 have about the 32 % of cross-beta structure against the 70 % of fibrils formed at pH 2.0. Besides, fibrils formed at neutral pH contain two major populations of beta-sheets in similar percentage, as visible by FTIR, while fibrils formed at acidic pH have only one main component corresponding to beta-structure.

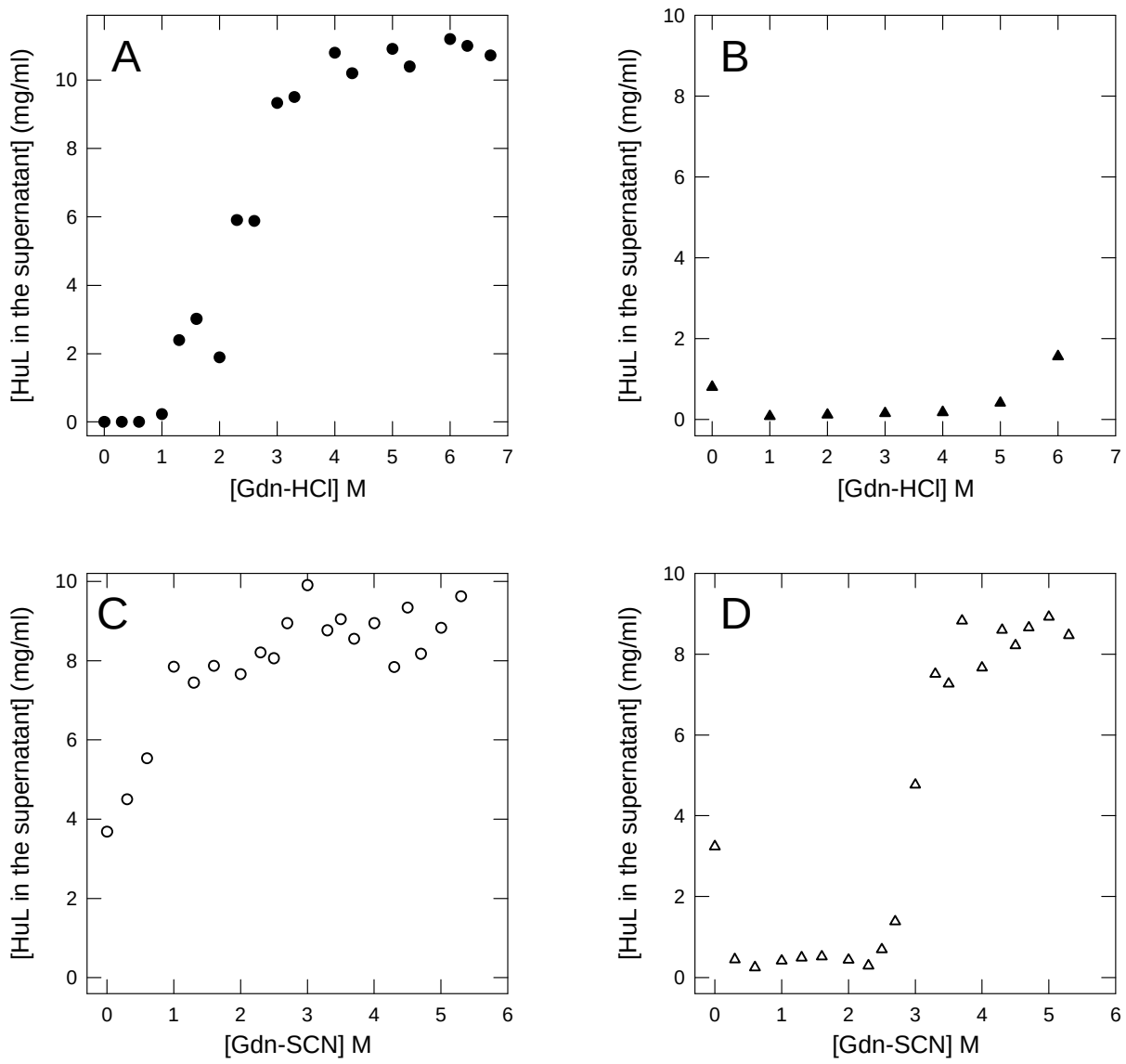


Fig. 18. Guanidinium-induced dissociation of HuL fibrils. A and B describe the GdnHCl denaturation of HuL fibrils formed respectively at pH 7.5 and pH 2.0 (see Methods). The same processes in GdnSCN are described in C and D, respectively for HuL fibrils formed at pH 7.5 and pH 2.0.

These structural differences have effects also on fibrils stability, since fibrils having less beta-structure content are also more sensitive to dissociation by denaturants, and hence less chemically stable.

This conformational polymorphism can be explained by two kinds of phenomena. It could be proposed that the segment known to be involved in the cross-beta core of HuL fibrils formed at pH 2.0 is the same involved in the cross-beta structure of fibrils formed at neutral pH, with the only difference of being rearranged in different ways, concerning the lengths of the beta-strands, whether they are arranged in a parallel or antiparallel arrangement within each sheet and the number of beta-sheet in each protofilament. A second explanation of this polymorphism is that different segments of the HuL sequence form the cross-beta structure at neutral pH compared to acidic pH: in this case, the variations can be attributable to the parameters cited above and to other factors such as the lengths and the conformational properties of the loops, turns and other regions that are not included within the core structure and the fraction of residues incorporated in the core structure. Another factor to be considered is that the exact spacing between the beta-sheets varies with factors such as steric bulk of the side chains that are packed together in the core.

The study of the fibrils dissociation induced by chemicals is indeed a good method to have a direct evaluation of stability of fibrils. Interestingly, *ex vivo* amyloid fibrils made of Asp67His HuL variant have been reported to be solubilized by denaturation in 6 M GdnHCl (Booth et al., 1997), conditions in which HuL fibrils formed at acidic pH are completely stable. The differences between these two kinds of fibrils are more than just “variation on a common theme” (Chiti and Dobson, 2006), but must define different families of amyloid structures.

References

Bischof JC, He X. (2005) Thermal stability of proteins. Ann N Y Acad Sci. 1066, 12-33.

Chatani E, Goto Y. (2005) Structural stability of amyloid fibrils of beta(2)-microglobulin in comparison with its native fold. Biochim Biophys Acta. 1753, 64-75.

Chiti F and Dobson CM. (2006). Protein misfolding, functional amyloid, and human disease. Ann. Rev. Biochem. 75, 333-366.

Dumoulin M, Canet D, Last AM, Pardon E, Archer DB, Muyldermans S et al. (2005) Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations, J. Mol. Biol. 25, 773-788.

Frare E, Mossuto MF, Polverino de Laureto P, Dumoulin M, Dobson CM, Fontana A. (2006) Identification of the core structure of lysozyme amyloid fibrils by proteolysis. J Mol Biol. 361, 551-561.

Jaroniec CP, MacPhee CE, Bajaj VS, McMahon MT, Dobson CM, Griffin RG. (2004) High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. Proc Natl Acad Sci U S A. 101, 711-716.

Jiménez JL, Nettleton EJ, Bouchard M, Robinson CV, Dobson CM, Saibil HR. (2002) The protofilament structure of insulin amyloid fibrils. Proc Natl Acad Sci U S A. 99, 9196-9201.

Matagne A, Jamin M, Chung EW, Robinson CV, Radford SE, Dobson CM. (2000) Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme. J Mol Biol. 297, 193-210.

Morozova-Roche LA, Zurdo J, Spencer A, Noppe W, Receveur V, Archer DB et al. (2000) Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants, J. Struct. Biol. 130, 339-351.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas' AF, Gilmanshin RI. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. Biopolymers. 31, 119-128.

Serpell LC. (2000) Alzheimer's amyloid fibrils: structure and assembly. Biochim Biophys Acta. 1502, 16-30.

CONCLUSIONS

During my PhD thesis work many aspects regarding protein amyloidogenesis have been studied. First, insights into the amyloid prone conformations of two proteins, LA and HypF-N, have been produced. The experiments evidenced the relevance of the unfolding of a globular protein triggering fibrils formation. The unfolding process, in order to be most effective, must lead to the population of pre-molten globule-like conformations, relaxed and flexible enough to allow the intra- and inter-molecular rearrangements necessary to the amyloid cross-beta structure formation.

The structure of amyloid fibrils has been the second object of the Thesis, trying to answer two main questions. Is the entire sequence of a protein involved in the cross-beta structure of fibrils? And, can a polypeptide chain rearrange in different ways into the cross-beta core of fibrils? The results shown in this Thesis clearly answer these questions, as far as human lysozyme (HuL) fibrils are concerned. In HuL fibrils produced at acidic pH only the central region of the polypeptide chain is involved in the cross-beta structure of the fibrils, while the rest of the protein assumes a random-like conformation. Moreover, it has been shown that HuL fibrils formed in conditions different from acidic pH have the overall properties of amyloid fibrils, that are fibrillar morphology, cross-beta structure and the ability to bind ThT. Despite these homologies, a more detailed analysis evidenced structural and stability differences. This polymorphism arises from the conformational heterogeneity in the beta-sheet rearrangement between the various kinds of fibrils.

APPENDIX: MAIN ANALYTICAL TECHNIQUES TO INVESTIGATE AMYLOID FORMATION

I. Circular Dichroism

Circular dichroism (CD) is the election spectroscopic technique used to determine the presence and the type of secondary structure in peptides and proteins, to evaluate their tertiary structure and to monitor structural transitions during unfolding and folding processes (Woody, 1995; Kelly et al., 2005). CD refers to the differential adsorption of the left and right circularly polarized components of plane-polarised radiation. This effect will occur when a chromophore is chiral (optically active)either (a) intrinsically by reason of its structure, or (b) by being covalently linked to a chiral centre, or (c) by being placed in an asymmetric environment. In practice the plane polarized radiation is split into its two circularly polarized components by passage through a modulator subjected to an alternating (50 kHz) electric field. The modulator usually consists of a piezoelectric quartz crystal and a thin plate of isotropic material (e.g. quartz) tighly coupled to the crystal. The alternating electric field induced structural changes in the quartz crystal which make the plate transmit circularly polarized light at the extremes of the field. If, after the passage through the sample, the left and right circularly polarized components are not absorbed (or are absorbed to the same extent), combination of the components would regenerate radiation polarized in the original plane. However, if one of the components is absorbed by the sample to a greater extent than the other, the resultant radiation (combined components) would now be elliptically polarized, i.e. the resultant would trace out an ellipse. In practice, the CD instrument (spectropolarimeter) does not recombine the components, but detects the two components separately; it will then display the dichroism at a given wavelength of radiation expressed as either difference in absorbance of the two components ($\Delta A = A_L - A_R$) or as the ellipticity in degrees (θ) ($\theta = \tan^{-1}(b/a)$, where b and a are the minor and major axes of the resultant ellipse). There is a simple numerical relationship between DA and q (in degrees), i.e. $\theta = 33 (A_L - A_R)$.

A CD spectrum is obtained when the dichroism, i.e. the variation of θ expressed in mdeg, is measured as a function of wavelength. Once that the spectrum has been acquired, the measurements must be normalized in order to make them independent from the protein concentration and the cuvette pathlength. This step is reached

expressing the ellipticity as mean residue ellipticity, $[\theta]_{MRW}$, calculated dividing the ellipticity for the molar concentration (using as molecular weight the mean molecular weight for residue) and for the pathlength:

$$[\theta] = \theta \cdot MRW / 10 \cdot c \cdot d$$

Where θ is espressed in mdeg, the protein concentration c in mg/ml, the pathlength d in cm and MRW is the mean residue molecular weight, obtained dividing the protein molecular weight for the number of its amino acid residues. The value $[\theta]$ is expressed in $deg \cdot cm^2 \cdot dmol^{-1}$.

Every chiral chromophore or belonging to a chiral molecule presents a characteristic activity of circular dichroism, i.e. it is active at certain wavelength values. As far as protein are concerned, interesting information can be obtained evaluating the region between 250 and 180 nm (far-UV CD). In this region the chromophore responsible for the dichroic signal is the amidic bond that presents a different ellipticity pattern depending on the conformation of the adjacent bond angles , i.e. depending on the kind of secondary structure in which it is embedded (alpha-helix, beta-sheet or random coil). Each kind of secondary structure possesses a typical spectrum, characterized by specific signals. Alpha-helix is characterized by two intense negative bands at 222 and 208 nm and one positive band at 192 nm. Beta-sheet structure presents a weak negative band at 218 nm and a positive band at 198 nm, while random coil structure shows a weak positive band at 218 nm and an intense negative one at ~ 198 nm.

In the UV region between 350 and 260 nm (near-UV) the dichroic signal is generated by disulfide bridges (intrinsically chiral), by aromatic groups of the side chains of residues such as Trp, Tyr and Phe that are in an asymmetric and rigid environment, like for instance in secondary structure segments, and by prosthetic groups. The CD spectrum in this region allows to obtain information on tertiary structure of a protein. The shape of the spectrum cannot be predicted: it is characteristic of each protein and defines a sort of fingerprint of the topology of aromatic residues and disulfide bridges in a protein molecule. Moreover, the spectrum shape can be influenced by many factors such as the protein rigidity, the interactions with surrounding elements, the number of aromatic residues, the solvent polarity and the temperature. In particular, if the amino acid side chains are mobile, the intensity of the CD band decreases. However, every chromophore has characteristic bands, useful for the identification: Phe

residues have an intense band between 260 and 270 nm, Tyr residues present a signal between 272 and 282 nm, and Trp residues are responsible for a band located between 288 and 293 nm. Disulfide bridges present a positive band around 250 nm.

II. Fluorescence

Fluorescence emission is observed when an excited electron returns from the first excited state back to the ground state. As some energy is always lost by non-radiative processes, such as vibrational transitions, the energyof the emitted light is always less than that of the absorbed light. Hence the fluorescence emission is shifted to longer wavelengths compared with the absorption of the respective chromophore.

The fluorescence of proteins originated from phenylalanine, tyrosine and tryptophan residues. The three aromatic amino acids absorbs in different extent, and also the fluorescence properties differ from each of them, as described in the Table below.

	Absorbance		*Fluorescence*		
Compound	λ_{max} *(nm)*	ε_{max} $(M^{-1} cm^{-1})$	λ_{max} *(nm)*	*Fluorescence quantum yield* (Φ_F)	*Sensitivity* ε_{max} Φ_F
Tryptophan	280	5600	348	0.20	1100
Tyrosine	274	1400	303	0.14	200
Phenylalanine	257	200	282	0.04	8

In proteins that contain all three aromatic amino acids, fluorescence is usually dominated by the contribution of the Trp residues, because both their absorbance at the wavelength of excitation and their quantum yield of emission are considerably greater than the respective values for Tyr and Phe.This is expressed by the “sensitivity” parameter (Table up), which is 1100 fro Trp and 200 for Tyr. Phe fluorescence is not observed in native proteins because its sensitivity of 8 is very low. The other factor in fluorescence is transfer of energy between residues. For example, Phe fluorescence also is barely observed becauseits emission is effiviently quenched by energy transfer to the other two aromatic amino acids. Tyr and Trp absorb strongly around 280 nm, where Phe emits fluorescence. In proteins that contain both Tyr and Trp, Tyr fluorescence is barely

detectable because Trp emission is strong, because in folded proteins Trp emission is frequently shifted to shorter wavelengths towards Tyr, and because non-radiative energy transfer can occur from Tyr residues to Trp residues in compact native state.

Changes in protein conformation, such as unfolding, very often lead to large changes in the fluorescence emission. In proteins that contain Trp, both shifts in wavelength and changes in intensity are generally observed upon unfolding. The emission maximum is usually shifted from shorter wavelengths to about 350 nm, which corresponds to the fluorescence maximum of Tro in aqueous solution. The exact location of this maximum depends to some extent on the nature and concentration of the buffer. In a hydrophobic environment, such as the interior of a folded protein, Trp emission occurs at shorter wavelengths.

Emission of both Tyr and Trp residues is observed when protein fluorescence is excited near the absorbance maximum around 280 nm. Trp fluorescence can be investigated selectively by excitation at wavelengths greater than 295 nm. Because of the red shift and the increased intensity of the absorbance spectrum of Trp when compared with Tyr, protein absorbance above 295 nm originates almost exclusively form Trp residues. A comparison of the emission spectra observed after excitation at 280 nm and at 295 nm gives information about the contribution of the Tyr residues to the observed emission. The fluorescence of the exposed aromatic amino acids of a protein, moreover, depends on the solvent conditions even in the absence of conformational changes. This effect can be used to probe the solvent accessibility of the aromatic residues of a native protein by fluorescence quenching techniques. It is also important for the evaluation of fluorescence difference spectra between native and unfolded proteins, the most commonly used unfolding agents are temperature, GdnHCl and urea. The fluorescence intensity generally decreases with increasing temperature: this decrease is substantial, about 1 % per degree in temperature for Tyr, and even more pronounced for Trp (about 2% per degree in temperature). Both GdnHCl and urea, moreover, exert a significant influence on the fluorescence of Tyr and Trp. Their fluorescence emission intensity slowly increases with increasing amount of the denaturing agents.

In order to study protein misfolding and aggregation two fluorescent probes have been used: ANS (8-anilino-1-naphthalene sulfonic acid) and thioflavin T (ThT). ANS is an hydrophobic able to bind exposed hydrophobic regions in proteins (Semisotnov et al., 1991): upon binding to a protein, its fluorescence intensity increases several times while the wavelength of maximum fluorescence emission shifts to shorter values, from 520 to

about 460-480 nm. ThT, instead, is a fluorescent dye that upon binding to the cross-beta structure of amyloid fibrils changes its fluorescence properties (LeVine, 1993). When it is free it absorbs at 342 nm and emits at 430 nm: when it is bound to amyloid structures it absorbs specifically at 442 nm and emits at 482 nm. The mechanism by which ThT works is still not well understood.

III. Fourier Transform InfraRed (FTIR) Spectroscopy

Infrared (IR) spectroscopy has proved to be a powerful tool for studying biological molecules and the applications of this technique to biological problems id continually expanding, particularly with the advent of Fourier-transform infrared (FT-IR) spectroscopy in recent decades. This technique is based on the vibrations of the atoms of a molecule. An infrared spectrum is obtained by passing infrared radiation through a sample and determining what fraction of the incident radiation is absorbed at a particularly energy. The energy at which any peak in an absorption spectrum appears corresponds to the frequency of a vibration of a part of the sample molecule.

Infrared (IR) spectroscopy is a chemical analytical technique, which measures the infrared intensity versus wavelength (wavenumber) of light. Based upon the wavenumber, infrared light can be categorized as far infrared (4-400cm^{-1}), mid infrared (400-4,000cm^{-1}) and near infrared (4,000-14,000cm^{-1}). Infrared spectroscopy detects the vibration characteristics of chemical functional groups in a sample. When an infrared light interacts with the matter, chemical bonds will stretch, contract and bend. As a result, a chemical functional group tends to adsorb infrared radiation in a specific wavenumber range regardless of the structure of the rest of the molecule. For example, the C=O stretch of a carbonyl group appears at 1700cm^{-1} in a variety of molecules. Hence, the correlation of the band wavenumber position with the chemical structure is used to identify a functional group in a sample. The wavenumber positions where functional groups adsorb are consistent, despite the effect of temperature, pressure, sampling, or change in the molecule structure in other parts of the molecules. Thus the presence of specific functional groups can be monitored by these types of infrared bands, which are called group wavenumbers.

The early-stage IR instrument is of the dispersive type, which uses a prism or a grating monochromator. The dispersive instrument is characteristic of a slow scanning. A Fourier Transform Infrared (FTIR) spectrometer obtains infrared spectra by first collecting an interferogram of a sample signal with an interferometer, which measures

all of infrared frequencies simultaneously. An FTIR spectrometer acquires and digitizes the interferogram, performs the FT function, and outputs the spectrum. An interferometer utilizes a beamsplitter to split the incoming infrared beam into two optical beams. One beam reflects off of a flat mirror which is fixed in place. Another beam reflects off of a flat mirror which travels a very short distance (typically a few millimeters) away from the beamsplitter. The two beams reflect off of their respective mirrors and are recombined when they meet together at the beamsplitter. The re-combined signal results from the "interfering" with each other. Consequently, the resulting signal is called interferogram, which has every infrared frequency "encoded" into it. When the interferogram signal is transmitted through or reflected off of the sample surface, the specific frequencies of energy are adsorbed by the sample due to the excited vibration of function groups in molecules. The infrared signal after interaction with the sample is uniquely characteristic of the sample. The beam finally arrives at the detector and is measure by the detector. The detected interferogram can not be directly interpreted. It has to be "decoded" with a well-known mathematical technique in term of Fourier Transformation. The computer can perform the Fourier transformation calculation and present an infrared spectrum, which plots adsorbance (or transmittance) versus wavenumber.

When an interferogram is Fourier transformed, a single beam spectrum is generated. A single beam spectrum is a plot of raw detector response versus wavenumber. A single beam spectrum obtained without a sample is called a background spectrum, which is induced by the instrument and the environments. Characteristic bands around 3500 cm^{-1} and 1630 cm^{-1} are ascribed to atmospheric water vapour, and the bands at 2350 cm^{-1} and 667 cm^{-1} are attributed to carbon dioxide. A background spectrum must always be run when analyzing samples by FTIR. When an interferogram is measured with a sample and Fourier transformed, a sample single beam spectrum is obtained. It looks similar to the background spectrum except that the sample peaks are superimposed upon the instrumental and atmospheric contributions to the spectrum. To eliminate these contributions, the sample single beam spectrum must be normalized against the background spectrum.

The infrared spectra of proteins exhibit absorption bands associated with their characteristic amide group, the structural unit common to all molecules of this type (Arrondo et al., 1993). The characteristic bands of the amide groups of protein chains are similar to the absorption bands exhibited by secondary amides in general, and are

labelled as *amide* bands. These bands are nine, named amide A, amide B and amide I-VII, in order of decreasing frequency. An isolated planar amide group gives rise to five *in-plane* modes (C=O stretching, C-N stretching, N-H stretching, OCN bending and CNH bending) and one *out-of-plane* mode (C-N torsion). The amidic bands I and II are the two principal bands of a proteins IR spectrum.

Amidic Band I, from 1600 to 1700 cm^{-1}, is principally associated to the stretching vibration of C=O bond, and in minor part to the stretching vibration of the C-N bond. This band is the most intense in the spectrum and its exact position is directly connected to the conformation of the polypeptide chain and to the pattern of hydrogen bonds. The absorbances of the band seem to be independent from the protein amino acid composition or from its hydrophobic, hydrophilic and charge properties.

Amidic band II is positioned between 1510 and 1580 cm^{-1} and is more complex than band I. It derives for the 40-60 % from the bending of N-H bonds, the 18-40 % from the stretching of C-N bonds and the 10 % from the stretching of C-C bonds. Also this band is affected by the protein conformation.

Bands A and B (at 3500 and 3100 cm^{-1} respectively), arise from the stretching vibrations of N-H bonds. Bands III and IV (at 1300 and 625 cm^{-1} respectively) are very complex and depend on many factors, such as the nature of side chains and of hydrogen bonds. Bands V, VI and VII are located at 725, 600 and 200 cm^{-1}, respectively, and hence are less useful in studying protein conformations. Also side chains give IR bands, in the region comprised between 1800 and 1400 cm^{-1} (bands I and II). Among the 20 amino acids only 9 (Asp, Asn, Glu, Gln, Lys, Arg, Tyr, Phr and His) show significant absorbances, depending also on the solution pH and hence the charge state of the amino acid.

Many synthetic polypeptides with known structure have been used to develop methods able to characterize the IR spectrum, and so the secondary structure, of unknown proteins. For example, polylysine is known to adopt random structure, beta-sheet structure or alpha-helix structure depending on temperature and solution pH (Susi et al., 1967). From this kind of studies it is possible to associate to an IR signal one secondary structure element. Beta-sheet structure absorbs between 1615 and 1640 cm^{-1} as well as between 1671 and 1679 cm^{-1}. Signals corresponding to alpha-helices are positioned between 1651 and 1657 cm^{-1}. Beta-turn structures are represented in a band between 1658 and 1671 cm^{-1} as well as between 1681 and 1696cm^{-1}. Bands corresponding to random coil structure are between 1641 and 1647cm^{-1}.

Proteins generally possess a variety of domains containing polypeptide fragments in different conformations. As a consequence, the observed amide I band is usually a complex composite, consisting of a number of overlapping component bands representing helices, beta-structures, turns and random structures. The intrinsic wideness of the single bands characteristic of the secondary structure is often greater than the separation between the bands. Hence, the single bands overlay and they cannot be separated in the whole spectrum. Having spectra with a high signal/noise ratio and very innovative softwares, it has been possible to develop methods to determine the secondary structure components of a protein from an IR spectrum. The best method to use for the estimation of protein secondary structure involves band-fitting the amide I band. The fractional areas of the fitted component bands are directly proportional to the relative amounts of structure that they represent. The percentages of helices, beta-structures and turns are estimated by addition of the areas of all of the component bands assigned to each of these structures and then expressing the sum as a fraction of the amide I area.

During FTIR experiments H_2O is a great problem: in fact, the band corresponding to the O-H bending mode is at 1644 cm^{-1}, obscuring the amide I band. If the protein solvent is H_2O, the O-H bonds absorb so greatly that it is impossible to subtract the H_2O signal. This problem limits the length of the pathlength to 10 μm or less, and as counterpart the protein concentration must be so high that the experiment become very protein-consuming. To solve this problem, standard FTIR experiments are usually performed in deuterated buffers and with deuterated proteins. The wavenumber of O-D bonds stretching, indeed, is lowered of 400 cm^{-1}, to be about 1244 cm^{-1}, totally compatible with the Amide I band. In this way, it is possible to use pathlength of 50 μm and protein concentrations between 0.1 and 10 mg/ml, instead of 20-50 mg/ml.

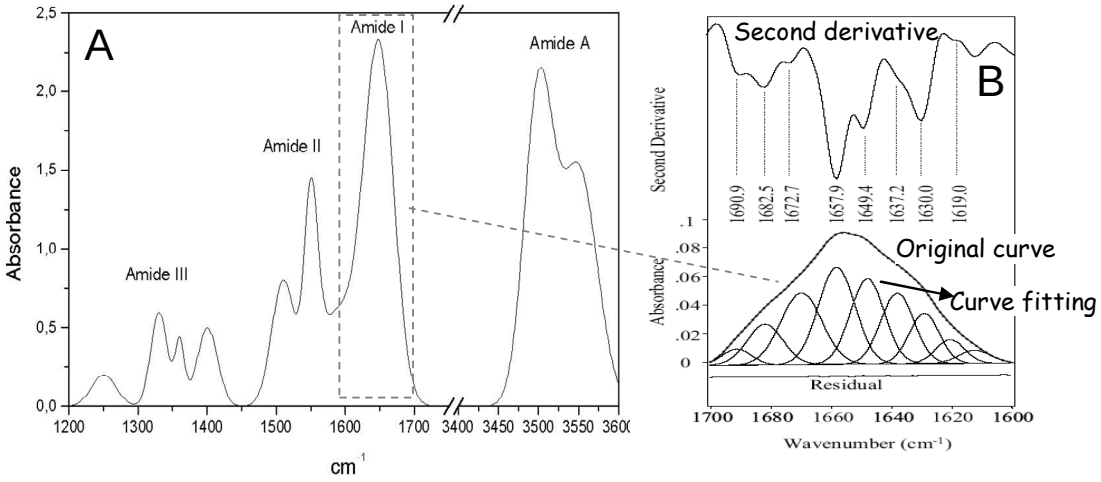
Some of the FTIR experiments reported in this Thesis work have been performed using *Attenuated total reflectance* (ATR) spectroscopy, a technique that utilizes the phenomenon of total internal reflection. A beam of radiation entering a crystal will undergo total internal reflection when the angle of incidence at the interface between the ample and the crystal is greater than the critical angle. The critical angle is a function of the refractive indices of the two surfaces. The beam penetrates a fraction of a wavelength beyond the reflecting surface and when a material which selectively absorbs radiation is in close contact with the reflecting surface, the beam loses energy at the wavelength where the material absorbs. The resultant attenuated radiation is

measured and plotted as a function of the wavelength by the spectrometer and gives rise to the absorption spectral characteristics of the sample. In IR spectroscopy, attenuated total reflectance technique is relatively sensitive and practical, as far as routine-type methods are concerned. ATR spectra with reasonable signal-to-noise ratios can be easily obtained. The technique is rapid, simple and required very little sample preparation. Also, one of the major advantages of the ATR technique is that the spectrum obtained is independent of the sample thickness. Typically, the reflected radiation penetrates the sample to a depth of only a few microns.

IV. Proteolysis

Proteolysis of a protein substrate can occur only if the polypeptide chain can bind and adapt to specific stereochemistry of the proteases active site (Fontana et al., 2004). However, since the active site 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. Therefore, the native rigid structure of a globular protein cannot act as a substrate for a protease, as documented by the fact that folded proteins under physiological conditions are rather resistant to proteolysis. 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 Fig. 20, 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 crystallographic



Amide I components (cm-1)	*Assignment*
1621-1627	beta-structure
1628-1634	beta-structure
1635-1640	beta-structure
1641-1647	random coil
1651-1657	alpha-helix
1658-1666	Turns and bends
1668-1671	Turns and bends
1671-1679	beta-structure
1681-1685	Turns and bends
1687-1690	Turns and bends
1692-1696	Turns and bends

C

Fig. 19 Typical IR spectrum of a protein. Amide I and II are the two principal bands in a protein IR spectrum (A). The analysis of Amide I band give information about the secondary structure content of a protein, both by calculating the second derivative and by the curve fitting. The minima position in the second derivative curve shows the wavenumbers of the main components of the spectrum (B, top). In the curve fitting, instead, the spectrum is deconvoluted into a number of curves whose sum is the starting spectrum. Every curve represents a secondary structure element (B, bottom). (C) Characteristic amide I band frequencies of protein secondary structures (Arrondo et al., 1993).

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 Cα-backbone. The main chain B-factor is a measure of average displacements of a 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 requires 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. 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.

The results obtained with thermolysin are in line with those derived from limited proteolysis experiments conducted on a variety of other proteins of known (3D) structure. 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 globular 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. (1998) 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. 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. Therefore, it can be proposed that proteolysis of rigid elements of secondary structure is thermodynamically very disadvantageous.

The limited proteolysis approach fro 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 (Fontana *et al*., 1999; Hubbard, 1998). To this aim, the most suitable proteasea are those displaying broad substrate specificità, such as subtilysin, thermolysin, proteinase K and pepsin (Bond, 1990). These endopeptidases display a moderate preference for 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. 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 peptidi fission in a globular protein, requiring seconds or days for the limited proteolysis event. 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 protein 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 (Fig. 20).

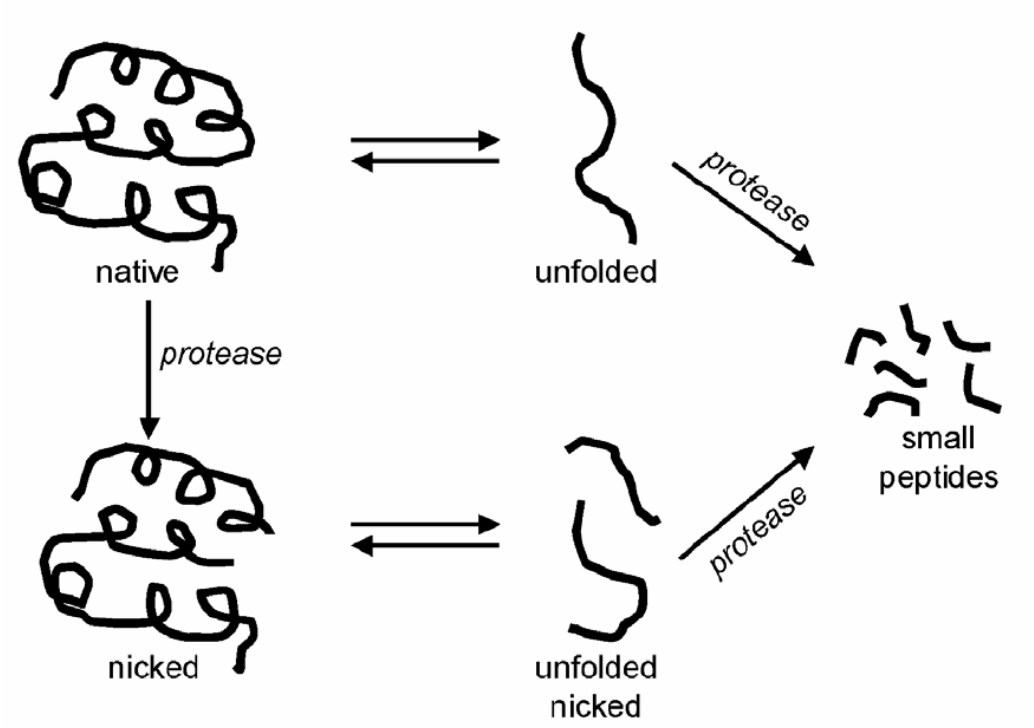


Fig. 20 Schematic view of the mechanism of proteolysis of a globular protein. A dual mechanism of protein degradation is shown, the one involving as substrate for proteolysis the (fully) unfolded protein and the other the native form of the protein. In this last case proteolysis is limited and occurs at flexible site(s), leading to a nicked protein species that can unfold and then be degraded to small peptides.

V. Mass Spectrometry

Thanks to recent technological progresses, mass spectrometry has become one of the most powerful tool in protein analysis (Aebersold and Mann, 2003). A further incentive to the development os suche techniques has come also from the huge amount of data deriving from the sequencing of intire genomes. Indeed, mass spectrometry (MS), allowing the identification of thousands of proteins starting from complex mixtures, has given significant insights into biological and medical matters. Ms can be exploited in many fields, ranging from chemistry and biology to pharmaceutical analysis. It can be used to measure the accurate molecular weight od peptides or proteins, both chemically sintetized or produced by recombinant methods. Other applications are the evaluation of the purity of a protein sample, the identification of post-translational modifications, the monitoring of chemical/enzymatic reactions, the sequencing of proteins or oligonucleotides. Lately, MS has been used also to study the folding and unfolding processes of proteins, or the formation of sovramolecular complexes and the determination of macromolecular structures.

Analysis of peptides and proteins using MS has been allowed by the development of ionization techniques not inducing protein fragmentation. The instrument used in the work presented in this Thesis is an ESI-Q-TOF (*Electrospray-Quadrupole-Time of Flight*)from Micromass (Manchester, UK), an instrument having an electrospray sorgent and an analyzer system composed of two quadrupoles and a time of flight, in series. Between the two analyzers ther is a collision chamber, used, when needed, for ions fragmentation.

In a generic MS experiment, a peptide/protein sample, dissolved in a solvent mixture containing H_2O/Acetonitrile/formic acid, is injected in the electrospray (ESI) ion source through a very small, charged and usually metal capillary. In electrospray ionization, a liquid is pushed though the capillary, with the analyte, dissolved in a large amount of solvent, which is usually much more volatile than the analyte. Volatile acids, bases or buffers are often added to this solution too. The analyte exists as an ion in solution either in its anion or cation form. Because like charges repel, the liquid pushes itself out of the capillary and forms an aerosol, a mist of small droplets about 10 μm across. An uncharged carrier gas such as nitrogen is sometimes used to help nebulize the liquid and to help evaporate the neutral solvent in the droplets. As the solvent evaporates, the analyte molecules are forced closer together, repel each other and break up the droplets. This process is called Coulombic fission because it is driven by

repulsive Coulombic forces between charged molecules. When the analyte is an ion free of solvent, it moves to the mass analyzer.

In the following phase the ion produced is analyzed, in the instrument, by a Q-Tof Micro (Micromass, Manchester), that connects an ESI source with two combined analyzers quadrupole Tof (*time of flight*). Mass analyzers separate ions according to their mass-to-charge ratio, following the dynamic properties of charged particles in electric and magnetic fields in vacuum. The quadrupole mass analyzer uses oscillating electrical fields to selectively stabilize or destabilize ions passing through a radio frequency (RF) quadrupole field, acting as a mass selective filter. The time-of-flight (Tof) analyzer uses an electric field to accelerate the ions through the same potential, and then measures the time they take to reach the detector. If the particles all have the same charge, then their kinetic energies will be identical, and their speed will depend only on their masses. Lighter ions will reach the detector first.

The data produced are represented in a mass chromatogram of total ion current (TIC), measured in the ion source. The instrument acquires a mass spectrum of the injected peptide/protein, based on an intensity vs. *m/z* (mass-to-charge ratio) plot. Afterwards, the instrument can determine the sequence of the peptides though a Tandem mass spectrometry analysis.

VI. Electron microscopy

Negative staining has been a useful specimen preparation technique for biological and medical electron microscopists for almost 50 years, following its introduction as an established procedure by Robert (Bob) Horne (Brenner et al., 1959).

Preparation of carbon supports

A solution of 0.5% w/v butvar was prepared in a 9% v/v glycerol–91% v/v chloroform emulsion. This emulsion is stable at room temperature and needs only to be vigorously shaken before use. Clean glass microscope slides were inserted into the glycerol–chloroform emulsion to approx. two thirds of their length and withdrawn vertically. Excess emulsion was removed by touching the end onto a tissue paper and one side of the slide/mica was wiped to remove the surface film. The slide was then positioned horizontally and the fluid film on the upper glass surface allowed to dry. The small glycerol–water droplets penetrate the drying film of butvar, thereby creating small holes. After wiping the edges of the slide with a tissue, the perforated butvar film was

floated onto a clean water surface. An evenly opaque appearance indicates the presence of a suitable array of small holes. EM 400 mesh copper grids (shiny side up) were then placed on the floating film, and a piece of white paper over-layered without moving the grids. The paper slowly became completely wet and was then removed with forceps, along with the attached grids and perforated butvar film, and dried in a dust-free environment. The sheets of paper+grids and perforated butvar film were then carbon-coated. Before use, the butvar was dissolved by spraying the grids with chloroform and grids were briefly glow-discharge treated (20 s) immediately before use. The hole size ranges from *ca*.1 to 10 μm.

Preparation of negatively stained specimens on holey carbon supports

To prepare the specimens using carbon support films we used the single-droplet Parafilm procedure. Protein or lipid sample concentrations were generally in the range 0.5–2.0 mg/ml. In order to avoid the precipitation of the sample and the formation of acetate salts crystals, the dilution to the desired concentration was obtained with the acid buffer for acid samples, and with mQ water for neutral samples. Specimens were individually negatively stained with 2 % w/v uranyl acetate.

Briefly, 25 μl sample droplets and 20 μl droplets of negative stain were placed in rows on a clean Parafilm surface. The sample was applied to the holey carbon support film by touching a grid to the droplet surface and most of the fluid removed by touching to the edge of a filter paper wedge. Then, depending upon the salt concentration of the sample solution, the negative stain solution was applied and removed as a single or multiple droplets (e.g. ×2 or ×3), with the intention of sequentially washing away the buffer and other salts or solutes that may interfere with the production of an amorphous stain-sugar film.

At the final stage, maximal removal of negative stain was necessary, so the filter paper was held in contact with the grid edge for ca. 10–20 s, thereby leaving only a thin film of aqueous negative stain+sample. If this final precaution is not adhered to, the stain film will often tend to be too thick. Specimen grids, still held by fine forceps, were then positioned horizontally and the sample+stain allowed to air-dry at room temperature (22°C).

References

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature. 422, 198-207.

Arrondo JL, Muga A, Castresana J, Goni FM (1993) Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy, Prog. Biophys. Mol. Biol. 59, 23-56.

Brenner S, Horne RW (1959) A negative staining method for high resolution electron microscopy of viruses. Biochim Biophys Acta. 34, 103-110.

Creighton TE (1997) Protein Structure: A Practical Approach (Practical Approach Series)

Fontana A, Polverino de Laureto P, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. Acta Biochim Pol. 2004;51(2):299-321.

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, Stocker W, eds, 257-284. Springer Verlag, Heidelberg.

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

Kelly SM, Jess TJ, Price NC. (2005) How to study proteins by circular dichroism. Biochim Biophys Acta. 1751, 119-139.

LeVine H (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution, Protein Sci. 2, 404-410.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas' AF, Gilmanshin RI. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe.Biopolymers. 31, 119-128.

Woody RW. (1995) Circular dichroism. Methods Enzymol. 246, 34-71.