

PREMIO FIRENZE UNIVERSITY PRESS  
TESI DI DOTTORATO

- 6 -



Francesco Bemporad

**Folding and aggregation studies  
in the acylphosphatase-like family**

Firenze University Press  
2009

Folding and aggregation studies in the acylphosphatase-like family / Francesco Bemporad. – Firenze : Firenze University Press, 2009.

(Premio FUP. Tesi di dottorato ; 6)

<http://digital.casalini.it/9788884539465>

ISBN 978-88-8453-845-8 (print)

ISBN 978-88-8453-946-5 (online)

Progetto grafico di Alberto Pizarro Fernández

© 2009 Firenze University Press

Università degli Studi di Firenze

Firenze University Press

Borgo Albizi, 28

50122 Firenze, Italy

<http://www.fupress.com/>

*Printed in Italy*

*To my mother*



# Contents

<b>Chapter 1</b>	
<b>Introduction</b>	<b>1</b>
1.1 Different conformational states are populated during life-span of proteins	1
1.2 Protein folding	3
1.2.1 Definition of the protein folding problem	3
1.2.2 The early studies and the characterisation of intermediates	3
1.2.3 $\Phi$ value analysis and folding mechanisms: is there a unifying mechanism?	6
1.2.4 The role of topology in determining folding; the importance of studying structurally related proteins	7
1.3 Protein misfolding	8
1.3.1 Definition and structure of amyloid fibrils	9
1.3.2 Mechanisms of amyloid aggregation	11
1.4 Aim of this thesis	12
1.4.1 The acylphosphatase-like family	12
1.4.2 Folding and aggregation in the acylphosphatase-like family	14
<b>Chapter 2</b>	
<b>Enzymatic activity in non-native Sso AcP</b>	<b>17</b>
2.1 Introduction	17
2.2 Results	20
2.2.1 The partially folded state populated during folding of Sso AcP shows acylphosphatase activity	20
2.2.2 The acylphosphatase activity observed in the Sso AcP partially folded state is highly sensitive to mutations	23
2.2.3 Investigation of the partially folded and transition states of Sso AcP using $\Phi$ value analysis	25
2.3 Discussion	28
2.3.1 Structure of the partially folded state and of the transition state	28
2.3.2 Enzymatic activity in the presence of a highly dynamic catalytic site	29
2.3.3 Biological function in the absence of a three-dimensional fold	31
2.4 Materials and methods	32
2.4.1 Mutagenesis	32
2.4.2 Protein expression and purification	33
2.4.3 Enzymatic activity essay	34
2.4.4 Development of enzymatic activity during folding	34

2.4.5	Equilibrium GdnHCl-induced unfolding curves	35
2.4.6	Folding kinetics	35
2.4.7	Unfolding kinetics	36
2.4.8	$\Phi$ value analysis	36
<b>Chapter 3</b>		
<b>Aggregation studies on Sso AcP</b>		<b>39</b>
3.1	Introduction	39
3.1.1	Aggregation from native states	39
3.1.2	Aggregation of Sso AcP	40
3.2	Results	42
3.2.1	Sso AcP aggregates regardless of the position of the N-terminal segment	42
3.2.2	N-terminal segment does not induce Sso AcP aggregation via a destabilising effect	45
3.2.3	A specific inter-molecular interaction between N-terminal segment and globular Sso AcP leads to the formation of early aggregates	47
3.2.4	Formation of initial aggregates depends on protein concentration	51
3.2.5	Experiments with acrylodan	53
3.3	Conclusions: a possible aggregation mechanism for Sso AcP	58
3.4	Materials and methods	59
3.4.1	Materials	59
3.4.2	Mutagenesis, protein expression and purification	60
3.4.3	Far-UV Circular dichroism	60
3.4.4	Thioflavin T fluorescence	60
3.4.5	Aggregation kinetics followed by static light scattering	61
3.4.6	Congo red staining	61
3.4.7	Enzymatic activity measurements	61
3.4.8	Equilibrium unfolding	62
3.4.9	Dynamic light scattering	62
3.4.10	Cysteine labelling	62
<b>Chapter 4</b>		
<b>Role of <math>\pi</math>-stacking in amyloidoses</b>		<b>65</b>
4.1	Aromatic residues and amyloidoses	65
4.1.1	Introduction	65
4.1.2	A possible role for $\pi$ -stacking in amyloid-like aggregation	66
4.1.3	mt AcP as a model for investigating $\pi$ -stacking	67
4.2	Results	68
4.2.1	Strategy employed	68
4.2.2	Phenylalanine 22	69
4.2.3	Tyrosine 25	70
4.2.4	Tyrosine 91	70
4.2.5	Tyrosine 98	71
4.2.6	Statistical analysis	71
4.3	Discussion	72
4.3.1	Aromatic residues promote amyloid aggregation of mt AcP due to their hydrophobicity and $\beta$ -sheet propensity	72

Contents	IX
4.3.2 Aromatic residues are frequent in the cross- $\beta$ core of fibrils but are not necessarily required	73
4.4 Conclusions	75
4.5 Materials and methods	75
4.5.1 Mutagenesis, protein expression and purification	75
4.5.2 Aggregation kinetics with Thioflavin T fluorescence	75
4.5.3 Data calculation	76
<b>Chapter 5</b>	
<b>Final remarks</b>	<b>77</b>
5.1 Conformational states distinct from the fully folded structure can present enzymatic activity	77
5.2 A comparison between the transition state ensembles of mt AcP and Sso AcP	78
5.3 Different regions of the sequence are involved in protein folding and amyloid-like aggregation	80
5.4 Conclusions	82
<b>Acknowledgements</b>	<b>83</b>
<b>Appendix A Equations and formulas</b>	<b>103</b>
<b>Appendix B Models for Sso AcP aggregation</b>	<b>111</b>
<b>Bibliography</b>	<b>119</b>



# Chapter 1

## Introduction

### 1.1 Different conformational states are populated during life-span of proteins

Proteins are among the most important molecules in living organisms as they carry out a variety of different and fundamental functions. The function of a given protein is directly related to the unique features of its functional state, which determines partners, ligands or substrates. Nevertheless, during their life-span proteins can adopt many different conformational states (Chiti and Dobson 2006).

With reference to figure 1.1, proteins are synthesized on ribosomes from the genetic information encoded in the cellular DNA. In some cases the protein is biologically active directly after translation, even though it populates an unfolded state (Dunker et al. 2001). In many other cases proteins need to achieve a folded structure to be functional. Folding *in vivo* is in some cases co-translational, implying that it is initiated before protein synthesis is completed, when the nascent chain is still attached to the ribosome (Dobson 2003). Other proteins, however, undergo folding after release from the ribosome, whereas others fold in specific compartments, such as mitochondria or the endoplasmic reticulum (ER), after trafficking and translocation through membranes (Dobson 2003). In the majority of cases, folding in cytoplasm, ER and mitochondria is chaperon assisted (Bukau et al. 2006). The native state of the protein is attained through formation of partially folded conformations. Native proteins can then interact with partners to form functional oligomers or polymers. Regulation of these processes is crucial for cells as errors in folding can give rise to potentially dangerous folds (Jahn and Radford 2005; Chiti and Dobson 2006). In physiological conditions misfolded peptides are either refolded or degraded but, if this protein quality control is impaired, they can expose sticky regions and form toxic oligomers. These species can give rise to  $\beta$ -strand rich protofibrils and fibrils (see section 1.3). In the next sections the processes of folding and misfolding, that represents topics of the present thesis, are discussed in detail. The problems that will be discussed in the next chapters will also be introduced.

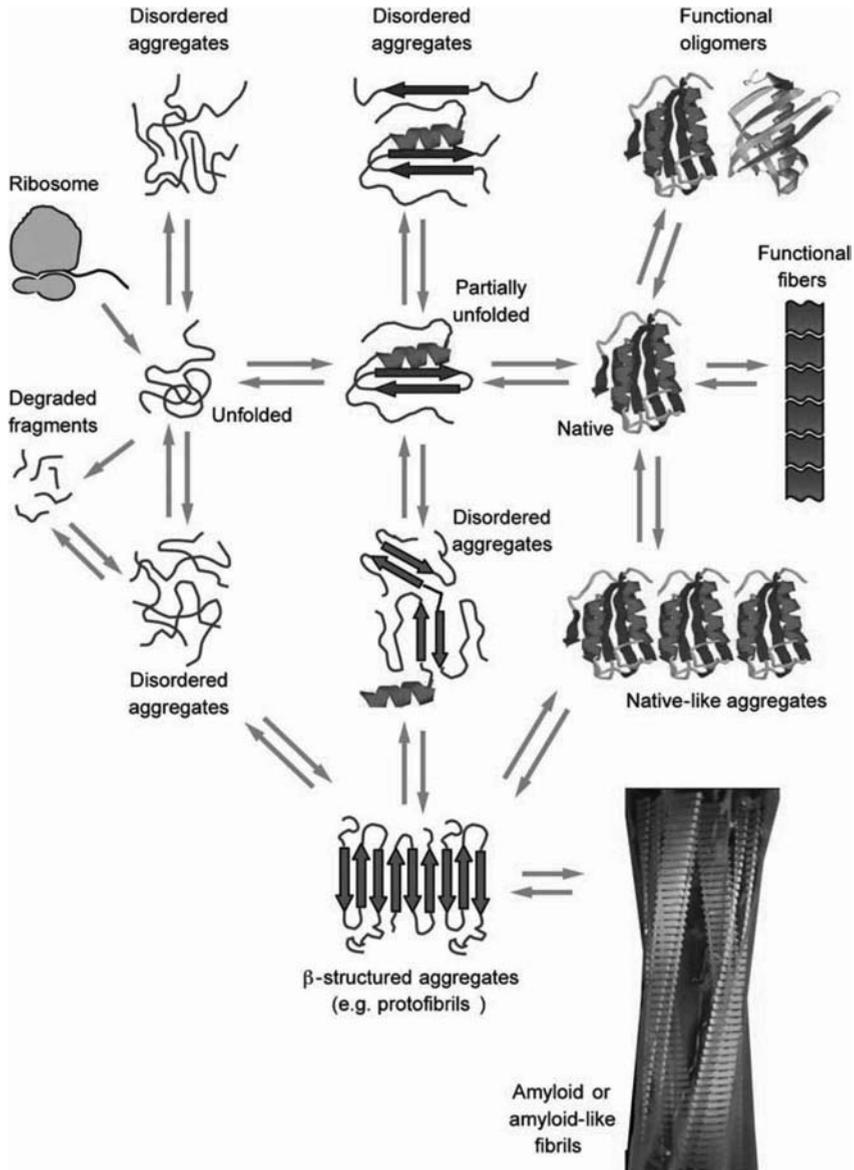


Figure 1.1: Different conformational states can be populated by a polypeptide chain. A nascent protein folds to reach a native and biologically active state. During this process partially folded conformations are transiently populated. Native states can form oligomers or polymers. At the end of its life-span a protein is degraded. The equilibrium between these species is crucial as folded, partially folded and unfolded states can aggregate. These aggregates can be either off-pathway (top of the figure) or on-pathway (bottom of the figure) and initially maintain the structural features of the precursor conformational states. They can later reorganise to form  $\beta$ -sheet containing aggregates and then amyloid. Reprinted from (Chiti and Dobson 2006).

## 1.2 Protein folding

### 1.2.1 Definition of the protein folding problem

Protein folding is the physical process by which a polypeptide chain changes its conformation to reach a biologically active three-dimensional structure. After decades of studies, protein folding has been recently named by *Science* among the 100 biggest unsolved problems in science (Editorial 2005). Traditionally, the protein folding problem is made by three distinct problems (Dill et al. 2007):

1. The thermodynamic question of how a native structure results from the interatomic forces acting on amino acid sequence (the folding code). This problem first arose when Christian B. Anfinsen showed that the three-dimensional structure of a protein is encoded by its amino acid sequence (Anfinsen et al. 1961; Anfinsen 1973). Although much work must be done to address this issue, it is now clear that the folding code is distributed both locally and not locally in sequence, that its dominant component is the hydrophobic interaction and that secondary structure is more a consequence than a cause of folding (Dill 1999). Moreover, novel proteins are now being designed as variants of existing proteins (Dwyer et al. 2004; Kaplan and DeGrado 2004).
2. The computational problem of how to predict three-dimensional structures solely on the basis of the primary sequence of polypeptides. To address this issue, two different approaches have been proposed to date: (1) the development of algorithms that use amino acid sequences as an input and produce, by homology modelling, structures as an output. A major milestone in this field is CASP (Critical Assessment of Techniques for Structure Prediction), a community-wide blind test to predict unknown structures (Moult 2006). Currently, structures of small globular proteins (i.e. about 90 residue long peptides) can be predicted within RMSD of 2-6 Å (Bradley et al. 2005; Zhang and Arakaki 2005). (2) The development of physics-only methods aimed to understand the final structure without database-derived knowledge. Although these methods are limited by huge computational requirements, using a distributed computing system (Folding@Home) Pande et al. folded villin to a distance RMSD of 1.7 Å (Zagrovic et al. 2002).
3. The kinetic question of how, and with which mechanism, can a protein fold so quickly. In 1968 Cyrus Levinthal first pointed out that, if a given protein is to attain its correctly folded configuration by sequentially sampling all the possible conformations, it would require a time longer than the age of the universe to arrive at its correct native conformation (Levinthal 1969). Solving Levinthal's paradox implies that folding proceeds in a step-wise manner and that several intermediates with increasing native-like structure are populated along the folding coordinate. In the past two decades major advances have been done in folding experiments and possible mechanisms of protein folding have been proposed (sections 1.2.2 and 1.2.3).

### 1.2.2 The early studies and the characterisation of intermediates

As a consequence of the Levinthal's paradox, the idea emerged that the characterization of partially folded states transiently populated during folding would give important

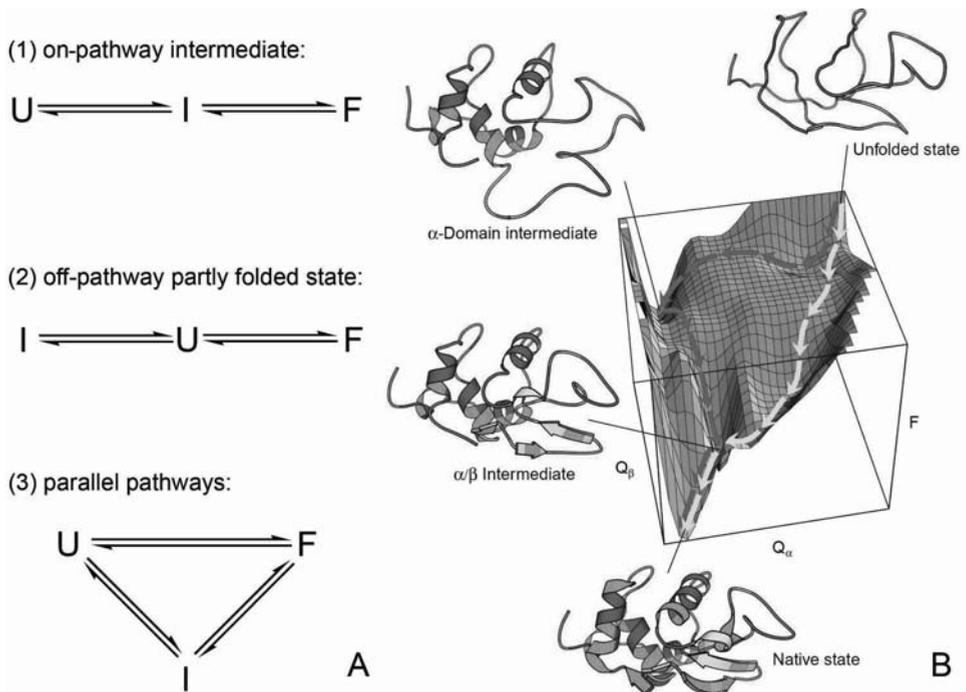


Figure 1.2: (A) Partially folded states (I) can play different roles in folding. They can be (1) on-pathway intermediates, (2) off-pathway partially folded states, (3) local minima transiently populated in one of many parallel pathways that lead to the formation of the native state (F). U refers to the unfolded state. (B) An example of energy landscape for folding of hen egg lysozyme (reprinted from (Dinner et al. 2000)). In this graph free energy is reported versus the number of native contacts in  $\alpha$  ( $Q_\alpha$ ) and  $\beta$  ( $Q_\beta$ ) domains. Two parallel pathways are shown. A fast pathway (yellow line) directly leads to the formation of the transition state and thus to attaining the native state. A slow pathway (red line) leads to the formation of a conformation that is in a local minimum, corresponding to a partially folded state in which only  $\alpha$  domain is structured.

insight into folding mechanism. At the end of 80s, the use of hydrogen exchange pulse labelling coupled to NMR (Roder et al. 1988) and of protein engineering methods (Matouschek and Fersht 1991) allowed structural characterization of partially folded states. More recently, the development of new instrumentation, such as ultra-rapid mixing devices (Shastry et al. 1998) and temperature jump relaxation techniques (Mayor et al. 2003) allowed the measurements of events within dead-time of normal stopped flow experiments. Finally, the destabilization of native state has allowed to increase the equilibrium population of partially folded states (Religa et al. 2005). This has allowed solution NMR methods to be applied to solve the structure of folding intermediates of small proteins (Religa et al. 2005). Generally, it has been shown that in the intermediate states the overall topology resembles the native structure and that some regions are highly structured while other regions are more denatured-like (Matouschek et al. 1989b; Salvatella et al. 2005).

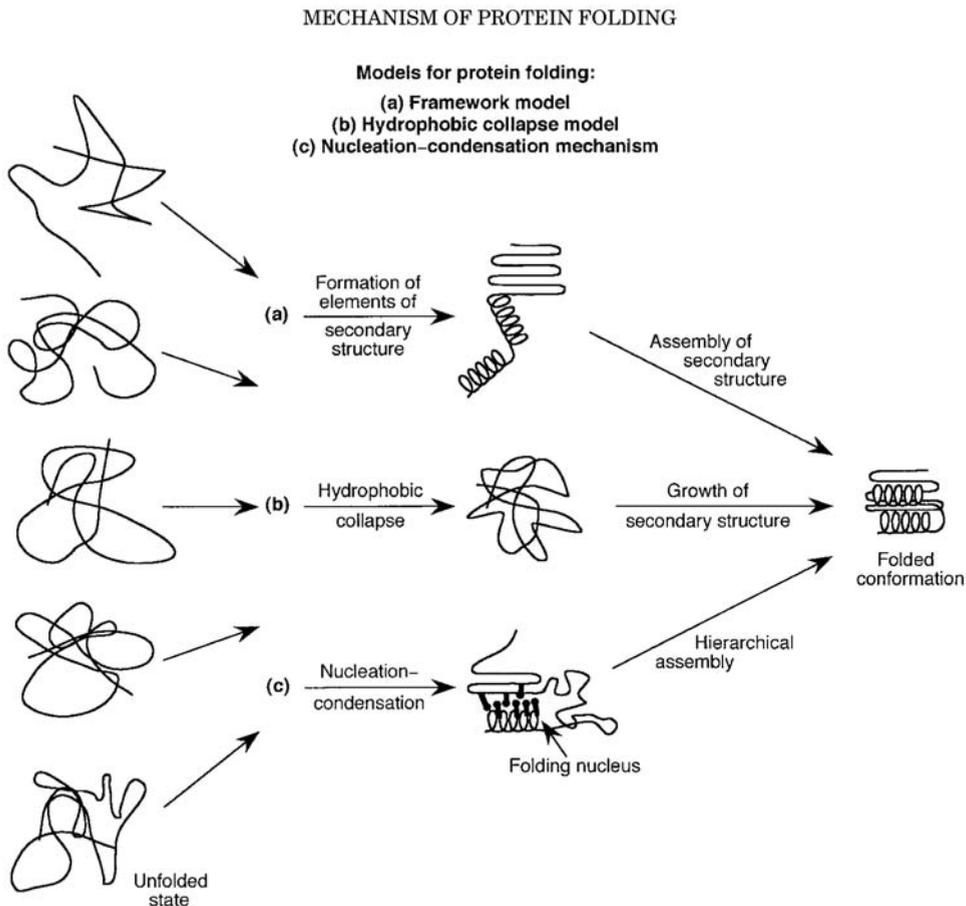


Figure 1.3: Models for protein folding (reprinted from (Nolting and Andert 2000)). Framework model, hydrophobic collapse model and nucleation-condensation model are shown. These model are described in the text.

The characterisation of partially folded states raised the question as to whether these states are productive species en-route to the native state (true intermediates, see model (1) in figure 1.2A) or kinetic traps that slow down the process (off-pathway partially folded conformations, see model (2) in figure 1.2A) (Bai 1999; Gianni et al. 2007b). If many proteins were shown to form on-pathway partially folded states (Bai 1999; Capaldi et al. 2001; Travaglini-Allocatelli et al. 2003; Jemth et al. 2004), recent observation that non-native interactions may be observed for productive on-pathway intermediates suggests that partial protein misfolding may be an obligatory step preceding native state consolidation (Capaldi et al. 2002; Religa et al. 2005). Moreover, in many cases evidence emerged that parallel pathways can lead to the formation of native states (Matagne and Dobson 1998) (see model (3) in figure 1.2A). This led to the description of energy landscapes for protein folding (figure 1.2B) (Dinner et al. 2000; Dobson 2003; Vendruscolo and Dobson 2005). This “new view” uses the idea of an en-

ergy surface to describe the conformational ensemble accessible during folding. Unfolded molecules, structurally different, follow different pathways and populate different conformational states characterised by weak interactions (Matagne and Dobson 1998). Faster paths and local minima exist in the landscape, but Brownian motions allow each molecule to escape kinetic traps and continue the search for the native state. If the trajectories can be numerous, the transition state is unique and all pathways lead to the formation of native state as native interactions are the most stable (Dinner et al. 2000). Thus, energy landscapes are not in contrast with the concept of folding pathway; they are instead able to join different pathways that can be detected with different experimental approaches.

### 1.2.3 $\Phi$ value analysis and folding mechanisms: is there a unifying mechanism?

Several methods have been developed in the two past decades to study folding (Zarrine-Afsar and Davidson 2004; Dill et al. 2007). Single molecule measurements can now be performed and FRET methods can watch directly the formation of particular contacts (Schuler et al. 2002; Magg et al. 2006). A major milestone in the analysis of protein folding mechanisms was the introduction, at the end of the 80s, of protein engineering methods to perform  $\Phi$  value analysis (Matouschek et al. 1989b). In this method a single point mutation is introduced to remove a specific contact from the native state of a given protein. The change in conformational stability upon mutation for the native state and for the investigated state is calculated. The ratio between these two quantities is a number ( $\Phi$  value) that varies from 0 to 1. A  $\Phi$  value equal to 0 suggests that the contact removed by mutagenesis is not formed in the investigated state. A  $\Phi$  value equal to 1 implies a native-like interaction in the investigated state for the removed group. In this way the characterisation of conformational states populated along the folding coordinate is possible. Although the structural information contained in this energetic parameter has been long discussed, protein engineering methods have been applied to analyse protein folding transition states at atomic level, to find amino acids that control folding speed (Matouschek et al. 1989b) and to investigate partially folded states (Matouschek et al. 1989a; Matouschek et al. 1992). In this thesis  $\Phi$  value analysis shall be used to investigate a partially folded state transiently populated during folding of the acylphosphatase from *Sulfolobus solfataricus* (see table 2.3).

Historically, the use of  $\Phi$  value analysis, coupled to other techniques (Nolting and Andert 2000; Daggett and Fersht 2003), allowed three mechanisms to be proposed for protein folding (Nolting and Andert 2000):

- **hydrophobic collapse:** In this model (figure 1.3) the initial event of the reaction is thought to be a relatively uniform collapse of the protein molecule, mainly driven by the hydrophobic effect (Baldwin 1989; Dill 1990). Stable secondary structure starts to grow only in the collapsed state, which narrows in a confined volume the conformational search to the native state. Although this model was initially supported by the observation that the hydrophobic driving force provided by the expulsion of water from the burial of non-polar surfaces is substantial, the hydrophobic collapse presents a problem because an excess of non-native interactions will hinder reorganisation of both the polypeptide chain and side chains (Daggett and Fersht 2003).

- **framework model:** According to this model (figure 1.3) folding starts with the formation of elements of secondary structure independently of tertiary structure, or at least before tertiary structure is locked in place. These elements then assemble into the tightly packed native tertiary structure either by diffusion and collision (Karplus and Weaver 1994) or by propagation of structure in a stepwise manner (Wetlaufer 1973). The validity of this model has been mainly shown for small helical proteins (Mayor et al. 2000; Myers and Oas 2001).
- **nucleation-condensation:** In the early 90s the previous model was challenged by two observations: (1) some proteins fold in a two-state process without accumulation of secondary-structure intermediates (Jackson and Fersht 1991); (2) the use of  $\Phi$  value analysis (Matouschek et al. 1989b) showed that in the folding transition state secondary and tertiary structure form in parallel (Otzen et al. 1994). This led some authors to propose a model (figure 1.3) in which early formation of a folding nucleus catalyses further folding (Fersht 1995; 1997). The nucleus primarily consists of a few adjacent residues that have some correct secondary and tertiary structure interactions but is stable only in the presence of further approximately correct interactions. The presence of the folding nucleus allows the transition state to bear a native-like topology (Lindorff-Larsen et al. 2005b) and the number of contacts that must be sampled dramatically decreases. Many small  $\alpha$  and  $\alpha/\beta$  proteins were shown to fold similarly (Clarke et al. 1997; Chiti et al. 1999b).

The three models mentioned above are in apparent contrast. Nevertheless, evidence is now emerging that framework and nucleation-condensation represent extreme manifestations of an underlying common mechanism. It was shown that when the helical propensity is increased, folding turns from nucleation condensation behaviour to diffusion-collision behaviour, in which helical elements are fully preformed (Gianni et al. 2003; White et al. 2005). In agreement with this observation, it was proposed that significant secondary structure is present in the denatured state if such structure is sufficiently stable. Thus, the rate-limiting step involves docking of these elements (Daggett and Fersht 2003). If instead the secondary structure is not stable in the unfolded state, a nucleation event is required to favour collapse of the structure (Daggett and Fersht 2003). Finally, a PDZ domain was recently shown to recapitulate nucleation condensation and diffusion-collision models in three steps: (1) the early formation of a weak nucleus formed by few residues with fractional  $\Phi$  values that determine the native-like topology of a large portion of the structure, (2) a global collapse of the entire polypeptide chain, and (3) the consolidation of the remaining partially structured regions to achieve the native state conformation (Gianni et al. 2007a).

#### 1.2.4 The role of topology in determining folding; the importance of studying structurally related proteins

If the fold is important in determining folding mechanism and speed, one can conclude a major role for topology in protein folding. Perhaps the most dramatic evidence for such a conclusion is the observation of a remarkable correlation between the experimental folding rates of a wide range of small proteins and the complexity of their folds, measured by the contact order (Plaxco et al. 1998). The latter is the average separation in the sequence between residues that are in contact with each other in the native structure. A correlation was also found between contact order and the position of the transi-

tion state along the folding coordinate (Plaxco et al. 1998). Many studies are now investigating the folding mechanism of different proteins sharing the same fold. The folding of several cytochrome c proteins involves the formation of a partially structured intermediate and some essential structural features of the intermediate and transition states are highly conserved across this protein family (Travaglini-Allocatelli et al. 2004). Five different PDZ domains have been shown to fold via an on-pathway intermediate and two transition states whose position is conserved along the folding coordinate (Chi et al. 2007). A comparison between  $\Phi$  values of structurally related proteins with divergent sequence composition suggested that protein families with conserved transition states are confined to a single folding trajectory whereas protein families with variable transition states have access to multiple pathways (Zarrine-Afsar et al. 2005). As an addition to this simple rule, it has been proposed that a two strand-helix (i.e. two  $\beta$ -strands docking against a single  $\alpha$ -helix) motif is the minimal folding nucleus, called foldon, and that pathway multiplicity is linked to the multiplicity of foldons within the protein structure (Lindberg and Oliveberg 2007).

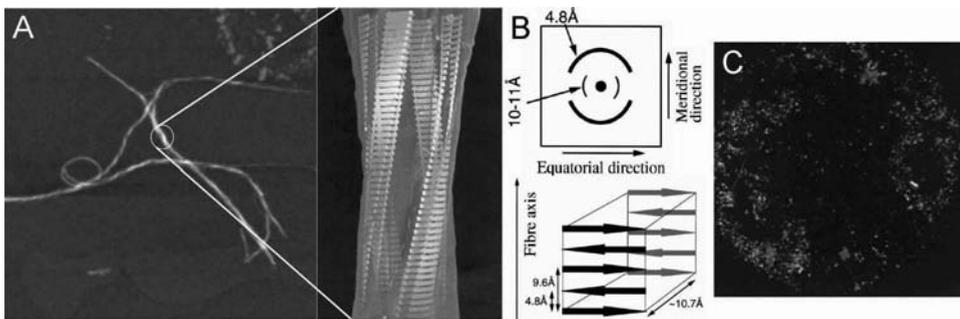


Figure 1.4: Features of amyloid fibrils. (A) Fibrils have a long and unbranched shape. White lines indicate a hypothetic enlargement of a fibril to show protofilaments. (B) Fibrils show a cross- $\beta$  structure. This scheme represents the typical appearance of fibrils when analysed with X-ray fibre diffraction (reprinted from (Serpell 2000)). (C) Amyloid material shows green Congo red birefringence under cross-polarised light. In this picture CR staining of amyloid deposits of  $\beta$ 2-microglobulin is shown (reprinted from (Ivanova et al. 2003)).

In this thesis a characterisation of a partially folded state and transition state populated during folding of the acylphosphatase from *Sulfolobus solfataricus* shall be carried out (see section 2.2.2). The results have implications in the study of biological function in the absence of a structured fold and will be discussed in section 2.2.3. Moreover, the obtained  $\Phi$  values shall be compared in chapter 5 with the results previously obtained on another member of the same superfamily, human muscle acylphosphatase (Chiti et al. 1999b).

### 1.3 Protein misfolding

As mentioned above protein quality control is a crucial point of cell metabolism because incompletely folded proteins must inevitably expose to the solvent at least some regions of structure that are buried in the native state and that are prone to inappropriate interaction with other molecules within the crowded intracellular or extracellular

environments (Dobson 2003). Protein misfolding is the conversion of a protein into a structure that differs from its native state (Chiti and Dobson 2006). This process is related to a set of human diseases (protein misfolding diseases), usually classified in three distinct groups:

1. Diseases in which an impairment in the folding efficiency of a given protein results in a reduced amount of native folded protein. An example is cystic fibrosis, where a mutated variant of a chloride channel (CFTR) populates a misfolded conformation that is degraded instead of being translocated to the cell membrane (Thomas et al. 1995). The absence of the properly functioning protein accounts for the symptoms of the disease (Thomas et al. 1995).
2. Diseases in which misfolding of a given protein results in improper trafficking. In the case of early-onset emphysema, that we cite here as an example, mutations of the gene encoding  $\alpha$ 1-antitrypsin determine the production of a protein that forms polymers. These will be retained in the liver (where  $\alpha$ 1-antitrypsin is produced) and will not reach the lungs, where  $\alpha$ 1-antitrypsin is necessary for a constitutive inhibition of elastase. The absence of such an inhibitory effect gives rise to the disease (Chiti and Dobson 2006).
3. The largest group of protein misfolding diseases is the group of pathological states associated with the conversion of a given peptide or protein from its soluble native state into highly organised fibrillar aggregates (Dobson 2003; Jahn and Radford 2005; Chiti and Dobson 2006). These aggregates are usually referred to as amyloid fibrils when they form *in vivo* outside the cell and intracellular inclusions when they form inside the cell. More than 40 human diseases are associated to this process (for a complete list see (Chiti and Dobson 2006)). Importantly, many of them have high social impact, such as Alzheimer's disease, Parkinson's disease and type II diabetes mellitus. In the following sections structure of fibrils and mechanisms of aggregation are discussed.

### 1.3.1 Definition and structure of amyloid fibrils

Proteins able to form amyloid aggregates do not share any sequence identity and structural homology. Despite that, some important structural features are common in amyloid fibrils formed by different peptides. Amyloid-like fibrils are defined on the basis of peculiar physico-chemical properties when investigated with different techniques (figure 1.4). In particular amyloid fibrils have a long and unbranched shape when observed with atomic force microscopy (AFM) or transmission electron microscopy (TEM) (figure 1.4A). Fibrils usually consist of a number (typically from 2 to 6) of protofilaments, each about 2-5 nm in diameter (Serpell et al. 2000) (see figure 1.5A). The protofilaments twist together and form rope-like fibrils that are typically 7-13 nm wide (Sunde and Blake 1997; Serpell et al. 2000). When analysed by X-ray fiber diffraction (figure 1.4B), the various protein molecules are arranged so that the polypeptide chains form  $\beta$ -strands that run perpendicular to the long axis of the fibril (cross- $\beta$  structure) (Sunde and Blake 1997). Finally, the fibrils have the ability to bind specific dyes such as Thioflavin T (ThT) (Krebs et al. 2005) and Congo red (CR, figure 1.4C) (Nilsson 2004), although the specificity of binding of CR to amyloid fibrils and the resulting green birefringence under cross-polarised light has recently been questioned (Khurana et al. 2001; Bousset et al. 2004b).

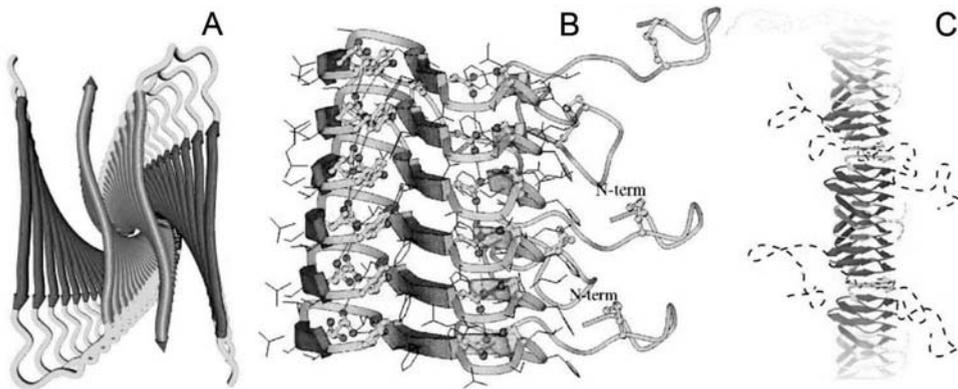


Figure 1.5: Structure of amyloid fibrils and protofilaments. (A) Structure of  $A\beta$  fibrils; reprinted from (Petkova et al. 2006). (B) Structure of human amylin fibrils; reprinted from (Kajava et al. 2005). (C) Structure of Sup35p protofilaments; reprinted from (Krishnan and Lindquist 2005).

High resolution structures of amyloid fibrils have been solved in recent years by use of solid state NMR (SSNMR), site-directed spin labelling coupled to electron paramagnetic resonance (SDSL-EPR) and by X-ray diffraction analysis carried out on nano- or microcrystals of small peptides having characteristics of amyloid fibrils (Tycko 2004; Luhrs et al. 2005; Chiti and Dobson 2006; Petkova et al. 2006). In the case of  $A\beta$ , the peptide whose aggregation is related to Alzheimer's disease, SSNMR studies led some authors to propose that each molecule forms two strands in the core of fibrils, spanning residues 12-24 and 30-40. These strands are not part of the same  $\beta$ -sheet but participate to the formation of two different parallel in register  $\beta$ -sheets that run parallel to the fibril axis (Antzutkin et al. 2000; Petkova et al. 2002). A single protofilament has been proposed to be composed by four  $\beta$ -sheets (i.e. two  $A\beta$  molecules) separated by distances of 10 Å (figure 1.5A). This structure has been confirmed by SDSL-EPR studies that found residues 13-21 and 30-39 highly structured in the fibrils (Torok et al. 2002). A peptide derived from the yeast prion Sup35p (GNNQQNY) and the peptide KFFEAAAKFFE have been converted in three-dimensional crystals with features typical of amyloid fibrils. The structures of these crystals have been solved by X-ray crystallography (Makin et al. 2005; Nelson et al. 2005). In the case of the Sup35p fragment, the crystal consists of pairs of parallel  $\beta$ -sheets in which each individual peptide molecule contributes a single  $\beta$ -strand. The stacked  $\beta$ -strands are parallel and in register in both sheets. The two sheets interact with each other through the side chains of Asn2, Gln4, and Asn6 (Nelson et al. 2005). Following these pioneering studies the X-ray structure of other assembled peptides have been solved (Sawaya et al. 2007).

In some cases different approaches have been applied to solve amyloid fibril structures. A model for fibrils formed by human amylin, whose aggregation is related to type II diabetes mellitus, has been recently proposed on the basis of different experimental evidences, such as protofilament diameter, the cross- $\beta$  structure and the experimental evidence of a parallel and in register arrangement of the  $\beta$ -strands formed by adjacent molecules (Kajava et al. 2005) (figure 1.5B). In the obtained model three different  $\beta$ -strands, formed by residues 12-17, 22-27 and 31-37, participate to the formation of three  $\beta$ -sheets that run parallel to the fibril axis. In the case of the NM region

of the yeast prion Sup35p, 37 single point mutations to cysteine were produced. These variants were labelled with fluorescent probes (Krishnan and Lindquist 2005). The wavelength maximum and the emission intensity of probes bound to different positions were then used to get information about burial of residues and possible interaction in fibrils. On the basis of these experiments a model was drawn in which NM molecules interact via head-to-head (residues 25-38) and tail-to-tail (residues 107-157) interactions (Krishnan and Lindquist 2005). The N-terminal (residues 1-20) and the distal end (residues 158-250) are instead structurally heterogeneous and solvent exposed (Krishnan and Lindquist 2005) (figure 1.5C). This approach has been used in chapter 3 to get insight into the protofibril structure of the acylphosphatase from *Sulfolobus solfataricus*.

### 1.3.2 Mechanisms of amyloid aggregation

If formation of amyloid fibrils is related to the onset of several human diseases, the ability to form amyloid-like aggregates in vitro was shown to be an inherent property of the protein back-bone. Several proteins that are not involved in disease were shown to form amyloid-like aggregates using particular cosolvents, temperature and salts (Guijarro et al. 1998; Chiti et al. 1999c). Nevertheless, the fact remains that different proteins show different tendencies and pathways leading to the formation of amyloid-like fibrils and that these differences can be explained on the basis of their sequence, that is on the basis of the physico-chemical properties of the side chains of their amino acids.

It is now clear that amyloid aggregation is a multi-step process in which different states are transiently populated. The formation of amyloid fibrils has many characteristics of a “nucleation-growth” mechanism (Chiti and Dobson 2006). Conversion of proteins into their fibrillar form follows two phases: (1) a lag phase, in which aggregation nuclei are formed and (2) a growth phase, in which further monomers or oligomers bind to the nuclei (Serio et al. 2000; Pedersen et al. 2004). This mechanism is confirmed by the observation that the addition of aggregation nuclei (seeds) to the sample shortens or abolishes the lag phase (Serio et al. 2000).

In the overall process of amyloid fibril formation various aggregates are thought to form before mature amyloid fibrils accumulate. In the case of A $\beta$ 1-40 and A $\beta$ 1-42 oligomeric species made by 2-4 and 5-6 molecules have been observed, respectively (Bitan et al. 2001; Bitan et al. 2003). In the case of Sup35p oligomers form rapidly and these species only afterwards convert into species with extensive  $\beta$ -sheet structure able to nucleate aggregation (Serio et al. 2000).

Other important species populated prior to the appearance of fibrils are protofibrils. These are isolated or clustered spherical beads 2-5 nm in diameter with  $\beta$ -sheet structure (Chiti and Dobson 2006). These species usually can bind both ThT and CR (Walsh et al. 1999). Species of this type have been observed for  $\alpha$ -synuclein (Conway et al. 2000), amylin (Kayed et al. 2004), transthyretin (Quintas et al. 2001) and the acylphosphatase from *Sulfolobus solfataricus*, one of the proteins object of this thesis (Plakoutsi et al. 2004) (see chapter 3).

The study of oligomers has seen an increasing importance as these species have been shown to be the most toxic for cells (Bucciantini et al. 2002). This is probably due to the fact that these inherently misfolded species expose an array of groups that are

normally buried in globular proteins or dispersed in highly unfolded peptides. This is likely to trigger aberrant events resulting from inappropriate interactions with cellular components, such as membranes, small metabolites, proteins, or other macromolecules. These interactions cause impairment of oxidative stress, ion balance and other factors that lead to cell death (Chiti and Dobson 2006).

It was also shown that, even if the final structure is conserved, different conformational states can be the starting point for amyloid-like aggregation (Bemporad et al. 2006) (see also figure 1.1). (1) A number of systems, including A $\beta$  and  $\alpha$ -synuclein, are largely unfolded prior to aggregation (Bemporad et al. 2006; Chiti and Dobson 2006). Short peptides have been shown to be able to form amyloid-like material (Lopez de la Paz and Serrano 2004). In this case, since the whole protein sequence is fully exposed to the solvent, aggregation is governed by simple physico-chemical factors, such as hydrophobicity, secondary structure propensities and net charge (Chiti et al. 2003) (for a complete description of the determinants of aggregation from unfolded states see section 4.1.1). (2) In most cases globular proteins need to unfold, at least partially, to aggregate into amyloid-like fibrils. It is clear for example that conditions that promote their partial unfolding, such as temperature, low pH, presence of organic cosolvents, increase their propensity to aggregate (Guijarro et al. 1998) (Chiti et al. 2000; Gosal et al. 2005). The aggregation of HypF-N can be initiated by a population of less than 1% of a partially folded conformation in equilibrium with the native one (Marcon et al. 2005). (3) Although the “conformational change hypothesis” can account for the aggregation properties of many proteins, increasing evidence is now accumulating that native proteins retain a significant, albeit small, tendency to aggregate ((Bemporad et al. 2006) and chapter 3). Formation of amyloid-like fibrils of insulin at low pH is preceded by an oligomerization step in which a native-like  $\alpha$ -helical content is retained, while  $\beta$ -sheet rich aggregates form only later on (Bouchard et al. 2000). The acylphosphatase from *Sulfolobus solfataricus* aggregates from an ensemble of native-like conformations into early aggregates that retain enzymatic activity, while unfolding is two orders of magnitude slower than aggregation (Plakoutsi et al. 2004; Plakoutsi et al. 2005).

In conclusion, the initial step of aggregation is the conversion of single molecules populating an aggregation prone state -folded, partially folded or unfolded- into an oligomer in which each monomer resembles the initial state. These oligomers convert afterwards into  $\beta$ -sheet rich species that lead to formation of amyloid fibrils.

## 1.4 Aim of this thesis

### 1.4.1 The acylphosphatase-like family

In this thesis we focus our attention on proteins belonging to the acylphosphatase-like structural family, the acylphosphatase from *Sulfolobus solfataricus* (Sso AcP) and the acylphosphatase from human muscle (mt AcP). Acylphosphatase (AcP) is a small (about 100 residue long)  $\alpha + \beta$  protein belonging to the ferredoxin-like fold (that is a  $\alpha + \beta$  sandwich with antiparallel  $\beta$ -sheet). The structure of the protein is highly conserved throughout the family. All the AcPs so far characterised show the same  $\beta\alpha\beta\beta\alpha\beta$  topology that originates a  $\beta$ -sheet docking against the two helices (figure 1.6A to 1.6H)

(Pastore et al. 1992; Thunnissen et al. 1997; Zuccotti et al. 2004; Miyazono et al. 2005; Pagano et al. 2006). Sso AcP also bears an unstructured, N-terminal 11 residue long segment that plays a major role in the aggregation mechanism of the protein ((Plakoutsi et al. 2006) and chapter 3). AcP is an enzyme (enzyme commission 3.6.1.7) able to hydrolyse acylphosphates, with formation of a phosphate ion and a carboxylate group (Stefani et al. 1997) (figure 1.6I). The mechanism of catalysis has been studied in detail in mt AcP (Taddei et al. 1994; Taddei et al. 1996; Taddei et al. 1997). The catalytic residues consist on an Arg and an Asn residues highly conserved within the family. The catalytic cycle can be summarised as follows. (1) The Arg binds to the phosphate. (2) The Asn residue stabilises the pentacovalent intermediate that forms following nucleophilic attack of a water molecule. (3) The products are released.

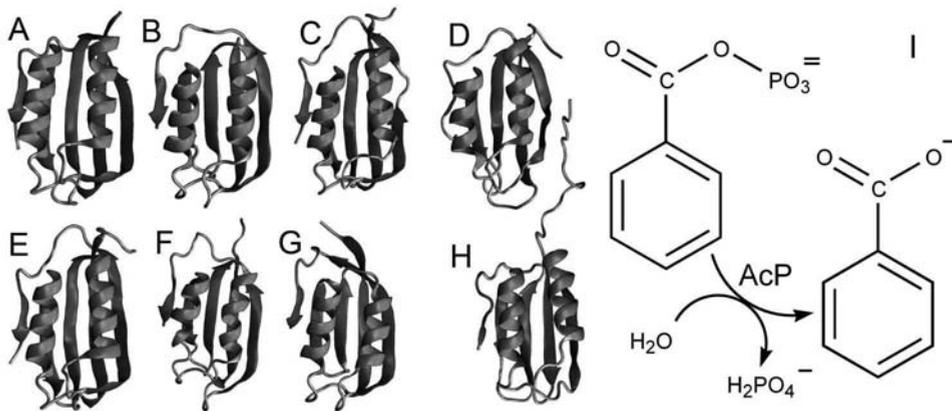


Figure 1.6: The structure of AcPs is highly conserved. AcPs are shown from *Pyrococcus horikoshii* (A, PDB entry 1W2I), *Thermus thermophilus* (B, PDB entry 1ULR), *Drosophila melanogaster* (C, PDB entry 1URR), Homo sapiens (muscular type, D, PDB entry 1APS), Homo sapiens (common type, E, PDB entry 2ACY), *Escherichia coli* (F, PDB entry 2GV1), *Sulfolobus solfataricus* (H, PDB entry 1Y9O). HypF-N from *Escherichia coli* is also shown (G, PDB entry 1GXT), (I) Reaction catalysed by AcP on benzoyl-phosphate, a substrate largely used for AcP enzymatic activity measurements (Stefani et al. 1997).

The function of the protein is yet not well understood. The initial observation that the enzyme is active on 1,3-bisphosphoglycerate suggested for AcP a possible regulative role on glycolysis (Ramponi 1975). Nevertheless, different functions have been proposed for this protein. The evidence that the enzyme is able to hydrolyse the  $\alpha$ -aspartyl-phosphate, which forms during action of membrane pumps, led some authors to propose a role in regulation of membrane transport (Nediani et al. 1996). More recently it was shown the ability of AcP to induce apoptosis in HeLa cells (Giannoni et al. 2000). Since the protein structure is similar to the RNA binding domain of ribonucleoproteins (Swindells et al. 1993), it was proposed that the interaction of this protein with nucleic acids has a physiological role. It was shown also a nuclear migration of mt AcP and its interaction with other DNAses in response to various apoptotic stimuli either in K562 or in Jurkat cells (Chiarugi et al. 1997). Finally, since the AcP levels increase in several cell lines during differentiation and AcP is able to hydrolyse both  $\gamma$  and  $\beta$

phosphate groups of ATP, it was proposed a role for AcP in regulation of differentiation through control of [ATP]/[ADP] levels (Paoli et al. 2000).

#### 1.4.2 Folding and aggregation in the acylphosphatase-like family

In the last decade AcP has been largely used as a model for folding and amyloid-like aggregation studies. The folding process shows high variability in the acylphosphatase-like family. Equilibrium and kinetic fluorescence studies suggest that the folding of mt AcP is a two state process characterised by a rate constant ( $0.23 \text{ s}^{-1}$ ) very low if compared with both other members of the family and with other proteins (van Nuland et al. 1998; Maxwell et al. 2005). The transition state for folding of mt AcP was also characterised (Chiti et al. 1999b; Vendruscolo et al. 2001). The most structured region corresponds to the central part of the  $\beta$ -sheet ( $\beta$ -strand 1 and 3). The second  $\alpha$ -helix as well plays an important role as it appears fully structured in the transition state (Taddei et al. 2000). The other human AcP isoform, common type AcP (ct AcP) folds in a two-state process with a folding rate constant equal to  $2.3 \text{ s}^{-1}$ . Although ct AcP bears a lower conformational stability than mt AcP, this value is ten-fold higher than folding rate for mt AcP, suggesting no correlation between folding rate and conformational stability within the AcP family (Taddei et al. 1999). Interestingly, HypF-N from *E. coli* folds with a rate constant equal to  $70 \text{ s}^{-1}$ , two orders of magnitude higher than the folding rate constant measured for mt AcP (Calloni et al. 2003). Folding proceeds via formation of a partially structured state that forms on the sub-millisecond time-scale (Calloni et al. 2003). The folding of Sso AcP has been characterised in some detail. The protein folds through formation of a partially folded state to reach the native structure with a folding rate constant equal to  $5.4 \text{ s}^{-1}$  (see section 2.1 details). These studies, carried out on structurally related but evolutionary distant proteins, allowed some important parameters for folding to be identified. Indeed, correlations within the acylphosphatase-like family indicate hydrophobicity, relative contact order and  $\alpha$ -helical propensity as important determinants for folding rates (Chiti et al. 1999b; Taddei et al. 2000; Calloni et al. 2003; Bemporad et al. 2004). Moreover, these studies clearly suggested that proline isomerism is not evolutionary conserved as proline residues that slow down folding are not conserved in the family (Bemporad et al. 2004). In chapter 2 we shall present a characterisation of the partially folded state, transiently populated during folding, and of the major transition state for folding of Sso AcP. The obtained results have important consequences in the study of biological function carried out in the absence of folded structures and will be discussed in section 2.3 and in section 5.2.

The amyloid aggregation processes of different AcPs have been studied as well. Interestingly, different conformational states can trigger the amyloid-like aggregation process. In the case of mt AcP and HypF-N, it was shown that a partial unfolding is required to initiate the process (Chiti et al. 1999c; Marcon et al. 2005). By contrast, an acylphosphatase from *Drosophila melanogaster* and Sso AcP have been shown to aggregate from a native-like state (Plakoutsi et al. 2004; Soldi et al. 2006a). The latter mechanism is particularly important as it suggests that native states have significant tendency to aggregate and that evolution has worked to keep under control this process ((Richardson and Richardson 2002) and paragraphs 1.3.2 and 3.1.1). The aggregation mechanism of Sso AcP has been studied (Plakoutsi et al. 2004; Plakoutsi et al. 2005; Plakoutsi et al. 2006). Although it was shown that an edge  $\beta$ -strand and the N-terminal unstructured segment play an important role in the process, the mechanism of aggrega-

tion of this system is still unclear (Plakoutsi et al. 2006). In chapter 3 we get some important clues into this process and we try to propose a possible model that summarises all the experimental evidence so far observed including those described here and in our previous work (see section 3.4). The model will be discussed with respect to the aggregation of other systems that aggregate starting from a native state.

The study of the aggregation of mt AcP has allowed several parameters that favour aggregation to be proposed. An algorithm has been proposed to predict change in aggregation rate upon mutation that takes into consideration hydrophobicity, secondary structure propensities and net charge as parameters (Chiti et al. 2003). Nevertheless, several other parameters have been proposed as determinants for amyloid aggregation (for details see section 4.1 and (DuBay et al. 2004; Pawar et al. 2005)). Among these parameters, the presence of aromatic residues has been proposed as an important determinant for aggregation rate constants (section 4.1). Nevertheless, several authors have shown that aromatic residues play an important role only because of their hydrophobicity and secondary structure propensity (section 4.1). In chapter 4 the possible role for aromaticity in amyloid aggregation is studied using mt AcP as a model. The results are discussed in view of their possible implications in the development of new algorithms.



## Chapter 2

### Enzymatic activity in non-native Sso AcP

#### 2.1 Introduction

Proteins are among the most abundant macromolecules in living systems and carry out a vast number of functions including the catalysis of virtually every chemical transformation occurring in cell biology and the transduction of signals inside and between cells. While it is well known that the flexibility of the native states of folded proteins is crucial in the processes determining their function, increasing evidence is accumulating about the existence of proteins or protein domains that adopt unstructured but functional states under physiological conditions (Dunker et al. 2001; Fink 2005). The mechanisms, however, by which certain proteins are capable of being active and yet natively unfolded is not completely understood (Fink 2005).

In this chapter we shall focus our attention on the acylphosphatase from the archaeon *Sulfolobus solfataricus* (Sso AcP, figure 2.1) and show that this protein retains an ability to function as an enzyme when adopting a non-native state in which the catalytic site is largely unstructured and flexible. Sso AcP is a 101-residue protein belonging to the acylphosphatase-like structural family. The structure of the native state of Sso AcP was recently determined by nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography (Corazza et al. 2006). This protein shares the same  $\beta\alpha\beta\beta\alpha\beta$  topology, typical of the ferredoxin-like fold, with the other acylphosphatases so far characterised (Pastore et al. 1992; Thunnissen et al. 1997; Zuccotti et al. 2004; Miyazono et al. 2005; Pagano et al. 2006). By contrast to related proteins, however, Sso AcP contains an unstructured, 12-residue N-terminal tail (Corazza et al. 2006). Sso AcP is able to hydrolyse benzoyl-phosphate (BP; figure 1.6I), with  $k_{CAT}$  and  $K_M$  values of  $198 \pm 20 \text{ s}^{-1}$  and  $0.36 \pm 0.04 \text{ mM}$ , respectively, at pH 5.3 and 25 °C, and to be competitively inhibited by inorganic phosphate (Corazza et al. 2006). The  $k_{CAT}$  value of the enzyme is low at 25 °C, but increases with temperature and reaches at 81 °C -the living temperature for the Archaeon *Sulfolobus solfataricus*- a value close to those previously reported for the mesophilic enzymes at 25 °C (Corazza et al. 2006). The native state of Sso AcP is thermodynamically very stable with a free energy change of unfolding ( $\Delta G_{U-F}^{H2O}$ ) of  $47 \pm 1 \text{ KJ mol}^{-1}$  at 37 °C (Corazza et al. 2006). The midpoint of thermal unfolding of the protein ( $T_m$ ) is  $100.8 \pm 4.1 \text{ °C}$  and at 81 °C the  $\Delta G_{U-F}^{H2O}$  is as high as  $20.6 \pm 0.3 \text{ KJ mol}^{-1}$ , similar to that of human muscle acylphosphatase (mt AcP) at 28 °C (Corazza et al. 2006).

The folding mechanism of Sso AcP was previously described at pH 5.5 and 37 °C (Bemporad et al. 2004). After removal of the denaturant, the unfolded state of this protein collapses on the microsecond time scale into an ensemble of partially folded con-

formations. This ensemble, which has a free energy of unfolding ( $\Delta G_{U-I}^{H2O}$ ) of  $12.5 \pm 2$  KJ mol<sup>-1</sup>, is capable of binding the fluorescent dye 8-anilino-1-naphthalenesulfonic acid suggesting the presence of hydrophobic clusters exposed to the solvent, and presents a far-UV mean residue ellipticity comparable to that of the fully native state, indicating that a native-like secondary structure is already formed in this state (Bemporad et al. 2004). The partially folded ensemble converts into the fully folded state with a rate constant of  $5.4 \pm 1.0$  s<sup>-1</sup>; a small fraction of molecules (ca. 10%) folds slower with a rate constant of ca. 0.2 s<sup>-1</sup> as their folding process is rate-determined by the *cis* to *trans* conversion of the Leu49-Pro50 peptide bond (Bemporad et al. 2004). The presence of a relatively stable, partially folded state accumulating during folding of Sso AcP offers a very favourable opportunity to study the function of a protein in a conformational state different from the native and folded one.

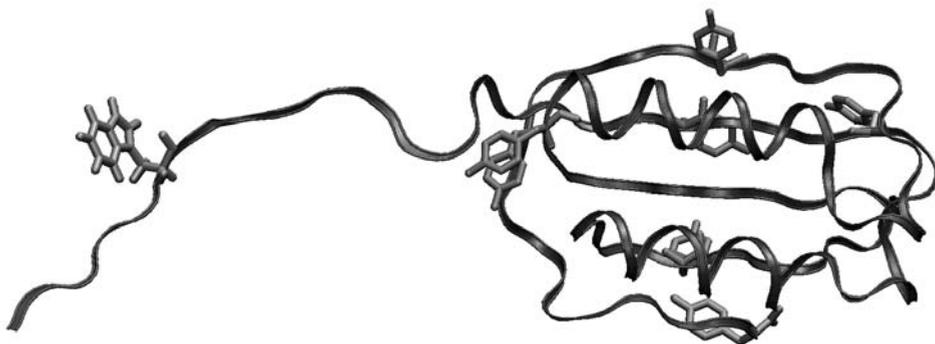


Figure 2.1: Spectroscopic probes of Sso AcP. Trp4, Tyr17, Tyr21, Tyr45, Tyr61, Tyr86, Tyr91 and Tyr101 are depicted in orange. The figure has been drawn with VMD 1.8.3 for win 32 (Humphrey et al. 1996).

In this work the functional properties of the partially folded state of Sso AcP accumulating during folding are investigated using a procedure that allows the recovery of enzymatic activity during folding to be determined in real time. The protein engineering method is then used to obtain information on the degree of structure formation, at the level of the mutated residues, in both the partially folded and transition states of the protein (Matouschek et al. 1989b). We shall show that the partially folded state of Sso AcP accumulating during folding shows enzymatic activity comparable to that of the native state. The experimentally obtained  $\Phi$  values are used as restraints in molecular dynamics simulations to obtain a model of the structures of the partially folded and transition state ensembles (Vendruscolo et al. 2001; Gsponer et al. 2006). These procedures illustrate how this state is made up by an ensemble of conformations displaying a native-like topology, but in which those regions of the sequence forming the active site in the folded protein exhibit high structural heterogeneity. In spite of the high flexibility existing at the level of the active site, which is also indicated by  $\Phi_I^{H2O}$  values close to 0 for mutations of residues in the catalytic loop, the native-like topology and the close proximity between the main substrate binding residue (Arg30) and the main catalytic residue (Asn48) ensures that this conformational state retains enzymatic activity.

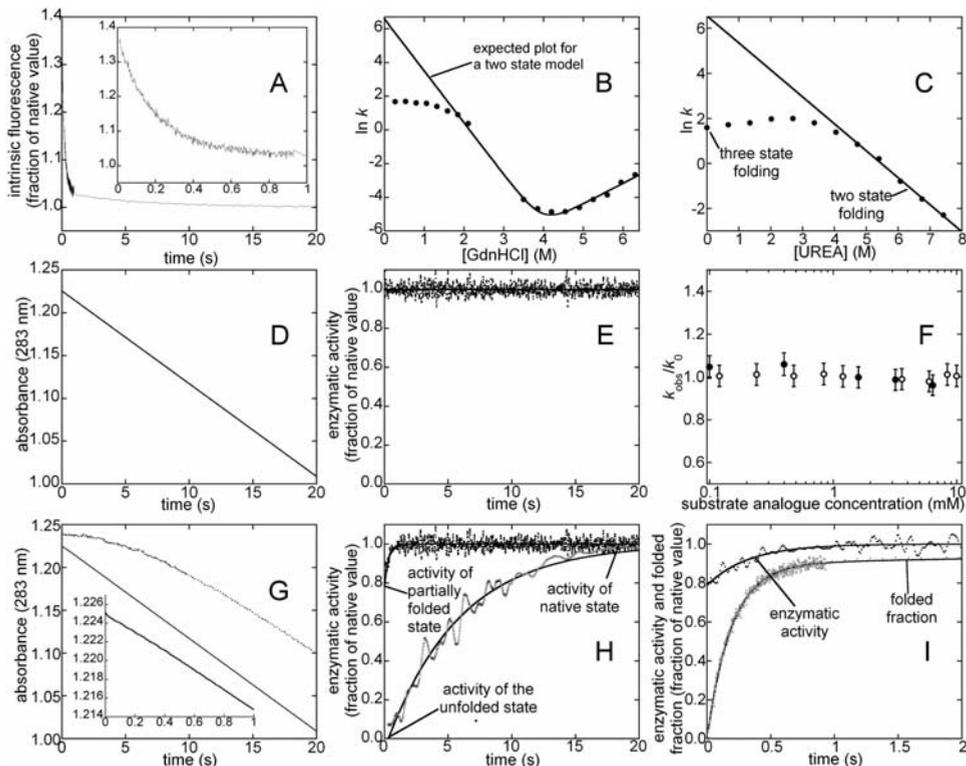


Figure 2.2: (A) Folding trace of Sso AcP recorded using intrinsic fluorescence as a probe in 0.275 M GdnHCl, 50 mM acetate buffer pH 5.5, 37 °C. The inset shows the first second of recording. (B) Observed folding/unfolding rate constant  $\ln k$  versus denaturant concentration; readapted from (Bemporad et al. 2004). The continuous line represents the expected plot for a two-state model. (C) Observed folding rate constant  $\ln k$  versus urea concentration. Above 5 M urea the plot is linear, suggesting two state folding in these conditions. (D) Time course of BP absorbance recorded at 283 nm in the presence of native Sso AcP. (E) Time course of enzymatic activity of native Sso AcP, calculated as the opposite of the first order derivative of the trace reported in panel D (see section 2.4). The continuous line represents the best fit to equation 2.3. (F) Ratio between the main folding rate constant recorded in the presence of phosphate ( $\circ$ ) or phenyl phosphate ( $\bullet$ ) ( $k_{obs}$ ) and that recorded without substrate-analogue in the sample ( $k_0$ ) plotted versus substrate-analogue concentration. (G) BP absorbance recorded at 283 nm after dilution of GdnHCl-unfolded Sso AcP into a refolding buffer; final conditions are 0.275 M GdnHCl (continuous line) and 0.275 M GdnHCl, 7 M urea (dotted line). The inset shows the first second of recording. (H) Development of relative enzymatic activity, calculated as the opposite of the first order derivative of the traces reported in panel G (see section 2.4), during Sso AcP refolding in the absence ( $\bullet$ ) and presence ( $\square$ ) of 7 M urea. The continuous lines represent the best fits to equation 2.3. The activities of fully folded, partially folded and unfolded states are shown. (I) Comparison between the time courses of recovery of native conformation (reported as fraction folded,  $\square$ ) and enzymatic activity (reported as fraction of the native protein activity,  $\blacksquare$ ). After 3.6 milliseconds, when the partially folded ensemble is populated more than 99%, Sso AcP exhibits already 79% of the native enzymatic activity.

## 2.2 Results

### 2.2.1 The partially folded state populated during folding of Sso AcP shows acylphosphatase activity

Sso AcP possesses one tryptophan in the N-terminal tail and seven tyrosines at various positions along the sequence, mainly positioned in the  $\beta$ -sheet (figure 2.1). Figure 2.2A shows the change of intrinsic fluorescence when one volume of Sso AcP unfolded in 5.5 M guanidinium hydrochloride (GdnHCl) is mixed with 19 volumes of refolding buffer. This trace, which is in very good agreement with that previously reported (Bemporad et al. 2004), indicates the presence of three phases in the folding process. The first phase, which occurs within the dead-time (10ms) of the stopped-flow device utilised here, leads to an increase of the intrinsic fluorescence of the protein to a value that is about 40% higher than that of the native state. This initial phase was shown to correspond to the conversion of the fully unfolded state into the partially folded state (Bemporad et al. 2004). The following rapid decrease of fluorescence with a rate constant ( $k_{F \rightarrow N}$ ) of  $5.3 \pm 1.0 \text{ s}^{-1}$ , corresponds to the conversion of this state into the fully native conformation, whereas the second slower decrease, with a rate constant ( $k_{12}$ ) of  $0.18 \pm 0.04 \text{ s}^{-1}$ , corresponds to the *cis-trans* isomerisation of a small fraction of protein molecules with the Leu49-Pro50 peptide bond initially in a non-native *cis* configuration (Bemporad et al. 2004). In addition to the transient hyper fluorescence signal reported in figure 2.2A, the presence of a rapidly formed partially folded state is also indicated by the downward curvature at low GdnHCl concentrations in the Chevron plot showing  $\ln k_{F \rightarrow N}^{H2O}$  versus denaturant concentration (figure 2.2B).

We then studied the time course of recovery of enzymatic activity during Sso AcP folding. Since the substrate BP, unlike its hydrolysis products benzoate and phosphate, has a significant optical absorption at 283 nm, the rate of catalysed BP hydrolysis can be accurately determined by measuring the decay rate of the absorbance at 283 nm ( $-dA_{283}(t)/dt$ ) in the presence of Sso AcP (Chiti et al. 1999a). In a first control experiment, a solution containing native Sso AcP was mixed with the refolding buffer containing a saturating concentration of BP and a non-denaturing concentration of GdnHCl. Final conditions were 0.01 mg ml<sup>-1</sup> protein, 10 mM BP, 0.275 M GdnHCl, 50 mM acetate buffer, pH 5.5, 37 °C. The absorbance arising from BP at 283 nm was monitored in real time and was found to decrease at a constant rate (figure 2.2D). This trace was analysed to yield the time course of enzymatic activity, as described in section 2.4.4. As expected, the enzymatic activity does not change with time (figure 2.2E).

To monitor the time course of enzymatic activity during Sso AcP refolding, a sample of GdnHCl-unfolded Sso AcP was mixed with the refolding buffer containing saturating BP. Final conditions were the same as those described above. The time-dependent changes of absorbance at 283 nm and of the corresponding enzymatic activity were determined (figures 2.2G and 2.2H, respectively). Immediately after mixing, when the partially folded state is maximally populated, the BP absorbance decay is occurring with considerable rate (see also the enlargement in the inset of figure 2.2G). The enzymatic activity is therefore already present, corresponding to  $79.3 \pm 10\%$  of

Table 2.1: Catalytic parameters for a set of Sso AcP variants measured in 50 mM acetate buffer at pH 5.5, 37 °C using BP as a substrate. Values reported in column 3 have been obtained by best fits of data shown in figure 2.3 to equation 2.3. Values reported in column 4 have been obtained combining data of column 1 and 2.

variant	native state $k_{CAT}$ ( $s^{-1}$ )	Partially folded state $k_{CAT}$ ( $s^{-1}$ )	activity in the partially folded state (% of that from native state)
WT	$222 \pm 20$	$175 \pm 30$	$79 \pm 10$
V24A	$124 \pm 10$	$7 \pm 5$	$5.6 \pm 10$
V27A	$13 \pm 1$	$0 \pm 10$	$0.0 \pm 10$
R30A	$7 \pm 1$	n. d.	n. d.
N48A	$2 \pm 2$	n. d.	n. d.
P50A	$137 \pm 14$	$0 \pm 10$	$0.0 \pm 10$
G52A	$178 \pm 18$	$6 \pm 5$	$3.3 \pm 10$
K92A	$169 \pm 20$	$111 \pm 30$	$66 \pm 10$

that of the native protein under the same conditions (Figure 2.2H). The activity then shows a small exponential increase with a rate constant of  $3.53 \pm 1.5 s^{-1}$  (figure 1F). This value is in reasonable agreement with the rate constant of folding determined with intrinsic fluorescence under these conditions ( $5.33 \pm 1.0 s^{-1}$ ) and corresponding to the conversion of the partially folded state into the native state. The kinetic traces of enzymatic activity and intrinsic fluorescence, both normalised to the values of the native protein, are compared in figure 2.2I. After 3.6 milliseconds the activity is equal to 80% of the value of the native protein, but the fraction of native protein is less than 4%.

In another experiment the GdnHCl-unfolded protein was diluted into a refolding solution containing urea. Final conditions were the same described above, except for a final urea concentration of 7 M. Under these conditions the native protein is still thermodynamically more stable than the unfolded state, making it possible to monitor the kinetics of folding. In addition, the plot reporting the folding rate constant versus urea concentration shows a downward curvature in the range of 0-5 M, indicating that in 7 M urea the partially folded state is destabilised and the protein folds according to a two-state model (figure 2.2C). The results show that the activity is absent immediately after mixing, when only the unfolded state is populated (figure 2.2G). Upon refolding, the activity then increases with a rate constant of  $0.13 \pm 0.02 s^{-1}$ , which is in good agreement with the folding rate constant ( $0.16 \pm 0.02 s^{-1}$ ) under these conditions (figure 2.2G). To rule out the possibility that a substantial fraction of the native state forms in the dead time of the stopped flow experiment, a double-jump experiment was carried out in which the GdnHCl-unfolded protein was diluted into the refolding buffer (first jump, refolding) and then, after 10 ms, transferred again to solutions containing GdnHCl at final concentrations ranging from 4.2 to 7 M (second jump, unfolding). Although such final conditions promote the unfolding of native Sso AcP and produce a single exponential change of intrinsic fluorescence (Bemporad et al. 2004), no significant fluorescence changes were observed in any of these experiments, suggesting that the native protein is not present 10 ms after the folding process was initiated when enzymatic activity is present.

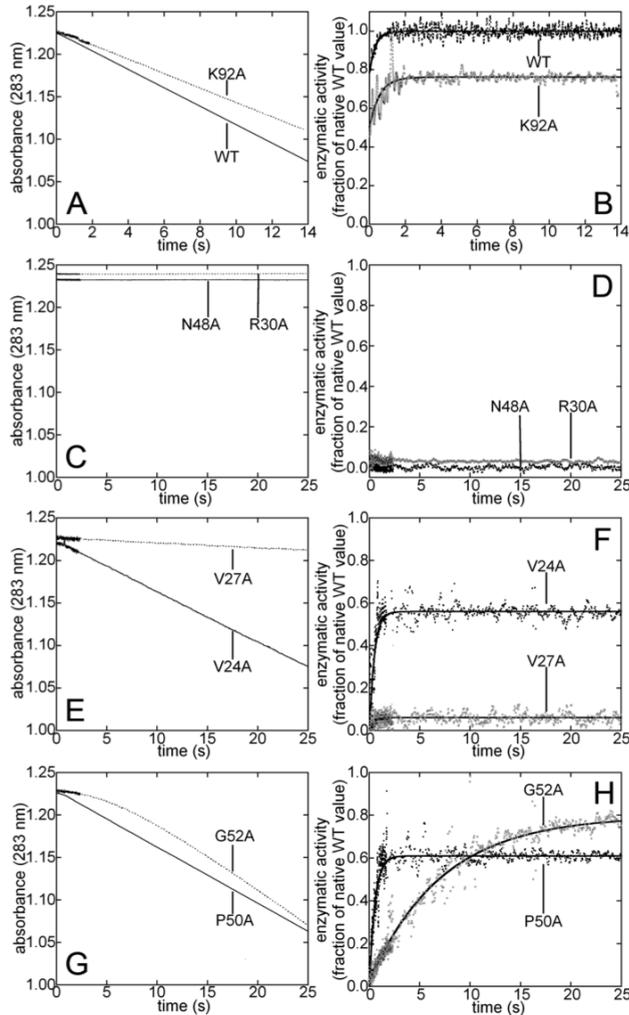


Figure 2.3: The enzymatic activity of the partially folded state of Sso AcP is sensitive to mutations. (A, C, E, G) BP absorbance at 283 nm after dilution of the indicated GdnHCl-unfolded mutants into a refolding buffer: the traces for wild type (A, continuous line), K92A (A, dotted line), N48A (C, continuous line), R30A (C, dotted line), V24A (E, continuous line), V27A (E, dotted line), P50A (G, continuous line) and G52A (G, dotted line) are shown. (B, D, F, H) Development of relative enzymatic activity, calculated as the opposite of the first order derivative of the corresponding traces reported in the corresponding panels on the left (see section 2.4.4), during the refolding of the indicated Sso AcP mutants. The traces for wild type (B, ●), K92A (B, □), N48A (D, ●), R30A (D, □), V24A (F, ●), V27A (F, □), P50A (H, ●) and G52A (H, □) are shown. The continuous lines represent the best fits to equation 2.3. The various mutants are labelled in the panels.

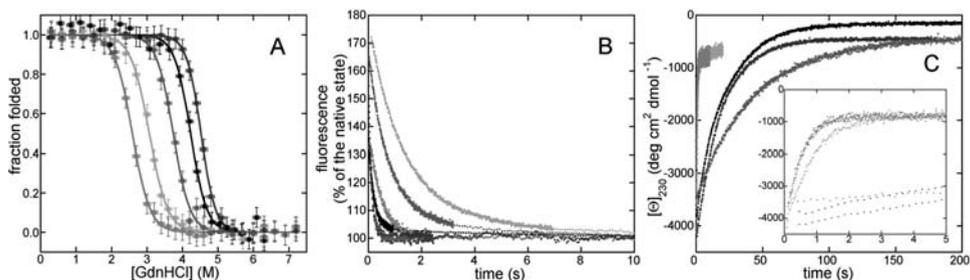


Figure 2.4: Folding thermodynamics and kinetics of Sso AcP variants. (A) Equilibrium unfolding curves for a set of Sso AcP variants in 50 mM acetate buffer, pH 5.5, 37 °C. The continuous lines represent the best fits to the equation reported by Santoro and Bolen (Santoro and Bolen 1988). The obtained parameters of conformational stability are reported in Table 2. (B) Folding traces recorded in 0.275 M GdnHCl, 50 mM acetate buffer at pH 5.5, 37 °C. (C) Unfolding traces recorded in 6 M GdnHCl, 50 mM acetate buffer at pH 5.5, 37 °C. The inset shows the first 5 seconds of recording. In all plots the traces refer to wild type (black), R71A, (blue), A46G (orange), L65A (red) and V20A (green) Sso AcP.

In order to further rule out the possibility that folding of Sso AcP is accelerated by BP and that the enzymatic activity observed at the beginning of the folding process is due to an early substrate-induced folding of the protein, Sso AcP refolding was also followed using intrinsic fluorescence in the presence of phosphate and phenyl phosphate, two competitive inhibitors of Sso AcP that are stable analogues of BP. Experimental conditions were the same as for the enzymatic activity experiments with no added urea. The folding rate constant is not affected by either compound (figure 2.2F). Taken together, these data indicate that the partially folded ensemble accumulating during folding of Sso AcP possesses significant enzymatic activity.

### 2.2.2 The acylphosphatase activity observed in the Sso AcP partially folded state is highly sensitive to mutations

The recovery of enzymatic activity during folding was also recorded for 7 mutants of Sso AcP. The K92A variant showed behaviour similar to that of the wild type protein with enzymatic activity detected for both the partially folded and native states (figures 2.3A and 2.3B; table 2.1). By contrast, the traces recorded for the N48A and R30A variants showed full inactivation of their partially folded and native states (figures 2.3C and 2.3D; table 2.1). These results confirm the key role of these two highly conserved residues in the catalysis of the native state of acylphosphatases (Stefani et al. 1997) and suggest their major role in the catalytic mechanism of the partially folded state as well. The partially structured ensembles of the V24A and V27A mutants do not show significant residual enzymatic activity (figures 2.3E and 2.3F; table 2.1). In the case of V24A, the enzymatic activity increases upon folding with a rate constant of  $3.2 \pm 1.0 \text{ s}^{-1}$ , a value that is, within experimental error, similar to that measured by following folding of the mutant using intrinsic fluorescence (table 2.3). Similarly, two variants with mutations within the 49-52 loop (P50A and G52A) display fully inactive partially folded states, but native structures with significant activity that is recovered with rate

constants highly consistent with those measured for folding using intrinsic fluorescence (figures 2.3G and 2.3H; tables 2.1 and 2.3). Hence, unlike the wild type protein, the recovery of enzymatic activity for the V24A, P50A and G52A mutants occurs concomitantly with the conversion of the partially folded ensemble into the native state. These data show that the ability to hydrolyse the substrate of the partially folded ensemble is more sensitive to mutations than the native state and provide insight into regions of the structure that are more important for enabling enzymatic activity.

Table 2.2: Equilibrium unfolding data for a set of Sso AcP variants.  $\Delta\Delta G_{U-F}^{H2O}$  values have been calculated according to equation 2.4, using the average  $m$  value over the  $m$  values reported in column 3.

variant	position	$C_m$ (M)	$m$ (KJ mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta\Delta G_{U-F}^{H2O}$ (KJ mol <sup>-1</sup> )
WT	-	4.23 ± 0.07	11.3 ± 1.1	-
R15A	β-sheet 1	3.88 ± 0.07	11.9 ± 1.2	3.90 ± 1.10
M16A	β-sheet 1	3.10 ± 0.07	11.0 ± 1.1	12.59 ± 1.13
A18G	β-sheet 1	3.21 ± 0.07	10.3 ± 1.0	11.36 ± 1.12
R19A	β-sheet 1	3.56 ± 0.07	10.4 ± 1.0	7.46 ± 1.11
V20A	β-sheet 1	3.78 ± 0.07	11.7 ± 1.2	5.01 ± 1.11
V24A	loop	3.28 ± 0.07	9.7 ± 1.0	10.58 ± 1.12
V27A	loop	3.78 ± 0.07	11.1 ± 1.1	5.01 ± 1.11
F29L	α-helix 1	3.58 ± 0.07	10.9 ± 1.1	7.24 ± 1.11
R30A	α-helix 1	4.41 ± 0.07	12.2 ± 1.2	-2.00 ± 1.10
A37G	α-helix 1	3.36 ± 0.07	10.1 ± 1.0	9.69 ± 1.12
I42V	loop	4.11 ± 0.07	11.5 ± 1.1	1.34 ± 1.10
A46G	β-sheet 2	3.07 ± 0.07	9.9 ± 1.0	12.92 ± 1.13
N48A	β-sheet 2	4.34 ± 0.07	11.2 ± 1.1	-1.23 ± 1.10
L49A	β-sheet 2	3.73 ± 0.07	11.2 ± 1.1	5.57 ± 1.11
P50A	loop	4.20 ± 0.07	11.8 ± 1.2	0.33 ± 1.10
G52A	loop	3.34 ± 0.07	11.3 ± 1.1	9.91 ± 1.12
V54A	β-sheet 3	3.11 ± 0.07	10.6 ± 1.1	12.47 ± 1.13
A58G	β-sheet 3	3.89 ± 0.07	10.6 ± 1.1	3.79 ± 1.10
E59A	β-sheet 3	3.52 ± 0.07	12.0 ± 1.2	7.91 ± 1.11
Y61A	β-sheet 3	4.18 ± 0.07	14.5 ± 1.4	0.56 ± 1.10
Y61L	β-sheet 3	4.18 ± 0.07	14.0 ± 1.4	0.56 ± 1.10
L65A	α-helix 2	2.59 ± 0.07	10.3 ± 1.0	18.27 ± 1.15
L68A	α-helix 2	2.54 ± 0.07	10.0 ± 1.0	18.82 ± 1.15
R71A	α-helix 2	4.52 ± 0.07	12.6 ± 1.3	-3.23 ± 1.10
I72V	α-helix 2	3.78 ± 0.07	11.8 ± 1.2	5.01 ± 1.11
P76A	loop	3.73 ± 0.07	12.1 ± 1.2	5.57 ± 1.11
P77A	loop	4.19 ± 0.07	10.1 ± 1.0	0.45 ± 1.10
V81A	β-sheet 4	3.13 ± 0.07	10.0 ± 1.0	12.25 ± 1.12
V84A	β-sheet 4	3.26 ± 0.07	10.6 ± 1.1	10.80 ± 1.12
F88A	β-sheet 4	2.92 ± 0.07	12.7 ± 1.3	14.59 ± 1.13
S89A	β-sheet 4	4.17 ± 0.07	11.2 ± 1.1	0.67 ± 1.10
K92A	loop	4.07 ± 0.07	10.1 ± 1.0	1.78 ± 1.10
G93A	loop	4.05 ± 0.07	10.8 ± 1.1	2.00 ± 1.10
F98L	β-sheet 5	3.01 ± 0.07	8.2 ± 0.8	13.59 ± 1.13

### 2.2.3 Investigation of the partially folded and transition states of Sso AcP using $\Phi$ value analysis

To characterise the folding pathway of Sso AcP and obtain structural information on the partially folded and transition states we carried out a  $\Phi$  value analysis using 34 single mutants (see table 2.2 for a complete list). Mutations were chosen to probe (1) the hydrophobic core (18 mutations involve residues contributing to the hydrophobic core of the native protein), (2) the active site (4 of the investigated mutations, namely R30A, N48A, V24A and V27A involve the catalytic site of the acylphosphatases and were found here to abolish or decrease dramatically the enzymatic activity of Sso AcP), (3) the salt bridges that are present on the protein surface and contribute to the enhanced structural stability of the protein (we studied R15A and E59A to probe the cluster formed by Lys14, Arg15, Glu59, Glu62, and Glu94; R19A to probe the cluster formed by Tyr17, Arg19, Tyr45, Lys47, Asp51, and Glu55; R71A to probe the charged interaction with Glu70; R30A to follow the salt bridge with the C-terminal carboxylate of the polypeptide chain (Corazza et al. 2006)).

Table 2.3: Folding and unfolding kinetics data for a set of Sso AcP protein variants.  $\Phi$  values have been calculated according to equation 2.7 and 2.8.

variant	$k_{I \rightarrow F}^{H_2O}$ (s <sup>-1</sup> )	$k_{F \rightarrow U}^{H_2O}$ (s <sup>-1</sup> )	$\Phi_I^{H_2O}$	$\Phi_{\ddagger}^{H_2O}$
WT	5.436 ± 0.272	(6.10 ± 0.30) · 10 <sup>-6</sup>	-	-
R15A	3.012 ± 0.151	(2.07 ± 0.10) · 10 <sup>-5</sup>	-0.20 ± 0.35	0.19 ± 0.23
M16A	7.179 ± 0.359	(2.04 ± 0.10) · 10 <sup>-4</sup>	0.34 ± 0.06	0.28 ± 0.07
A18G	4.249 ± 0.212	(3.21 ± 0.16) · 10 <sup>-5</sup>	0.57 ± 0.05	0.62 ± 0.04
R19A	0.652 ± 0.033	(2.25 ± 0.11) · 10 <sup>-5</sup>	-0.18 ± 0.18	0.55 ± 0.07
V20A	1.477 ± 0.074	(3.30 ± 0.17) · 10 <sup>-6</sup>	0.64 ± 0.09	1.31 ± 0.08
V24A	4.290 ± 0.214	(1.94 ± 0.10) · 10 <sup>-4</sup>	0.10 ± 0.10	0.16 ± 0.09
V27A	6.473 ± 0.324	(4.43 ± 0.22) · 10 <sup>-5</sup>	0.07 ± 0.21	-0.02 ± 0.23
F29L	6.421 ± 0.321	(1.47 ± 0.07) · 10 <sup>-4</sup>	-0.07 ± 0.17	-0.13 ± 0.18
A37G	4.996 ± 0.250	(1.43 ± 0.07) · 10 <sup>-4</sup>	0.14 ± 0.10	0.16 ± 0.10
A46G	0.855 ± 0.043	(1.66 ± 0.08) · 10 <sup>-4</sup>	-0.03 ± 0.09	0.34 ± 0.06
L49A	1.422 ± 0.071	(1.49 ± 0.07) · 10 <sup>-5</sup>	-0.03 ± 0.21	0.59 ± 0.09
G52A	0.181 ± 0.009	(1.21 ± 0.06) · 10 <sup>-5</sup>	-0.06 ± 0.12	0.82 ± 0.03
V54A	0.754 ± 0.038	(1.81 ± 0.09) · 10 <sup>-4</sup>	-0.11 ± 0.10	0.30 ± 0.06
A58G	1.567 ± 0.078	(6.14 ± 0.31) · 10 <sup>-6</sup>	0.15 ± 0.26	1.00 ± 0.05
E59A	2.343 ± 0.117	(5.49 ± 0.27) · 10 <sup>-5</sup>	0.01 ± 0.14	0.28 ± 0.10
L65A	2.354 ± 0.118	(2.78 ± 0.14) · 10 <sup>-4</sup>	0.34 ± 0.04	0.46 ± 0.04
L68A	1.064 ± 0.053	(3.35 ± 0.17) · 10 <sup>-4</sup>	0.23 ± 0.05	0.45 ± 0.04
R71A	10.090 ± 0.504	(7.01 ± 0.35) · 10 <sup>-6</sup>	0.62 ± 0.15	1.11 ± 0.07
I72V	4.364 ± 0.218	(2.22 ± 0.11) · 10 <sup>-5</sup>	0.22 ± 0.18	0.34 ± 0.15
P76A	5.003 ± 0.250	(3.83 ± 0.19) · 10 <sup>-4</sup>	-0.96 ± 0.39	-0.92 ± 0.38
V81A	4.761 ± 0.238	(7.72 ± 0.38) · 10 <sup>-4</sup>	-0.05 ± 0.10	-0.02 ± 0.09
V84A	4.729 ± 0.236	(2.98 ± 1.49) · 10 <sup>-4</sup>	0.04 ± 0.10	0.07 ± 0.10
F88A	5.176 ± 0.259	(2.01 ± 0.10) · 10 <sup>-3</sup>	-0.03 ± 0.08	-0.02 ± 0.08
F98L	4.211 ± 0.211	(7.61 ± 3.80) · 10 <sup>-3</sup>	0.04 ± 0.08	0.08 ± 0.08

An equilibrium GdnHCl-induced denaturation experiment was carried out for each mutant to yield the change in conformational stability upon mutation  $\Delta G_{U-F}^{H_2O}$  (see section 2.4.5). Final conditions were 50 mM acetate buffer, pH 5.5, 37 °C. Figure

2.4A shows representative equilibrium unfolding curves for the wild type protein, the only mutant found to be stabilised relative to the wild type (R71A), and three destabilised mutants, V20A, A46G, L65A. All plots have been analysed with equation 2.4 (see section 2.4.5). The results show that several mutants are destabilised (table 2.2).

The destabilised (or stabilised) mutants with  $\Delta G_{U \rightarrow F}^{H2O}$  values higher than 3.2 KJ mol<sup>-1</sup> or lower than -3.2 KJ mol<sup>-1</sup> were analysed to obtain the  $\Phi$  values of the corresponding mutations for the partially folded ( $\Phi_I^{H2O}$ ) and transition states ( $\Phi_{\ddagger}^{H2O}$ ). For each of these mutants kinetic traces for folding and unfolding were acquired at various denaturant concentrations, using intrinsic fluorescence and far-UV ellipticity as probes for folding and unfolding, respectively. Figures 2.4B and 2.4C show representative traces for folding and unfolding, respectively. All mutants showed, at low denaturant

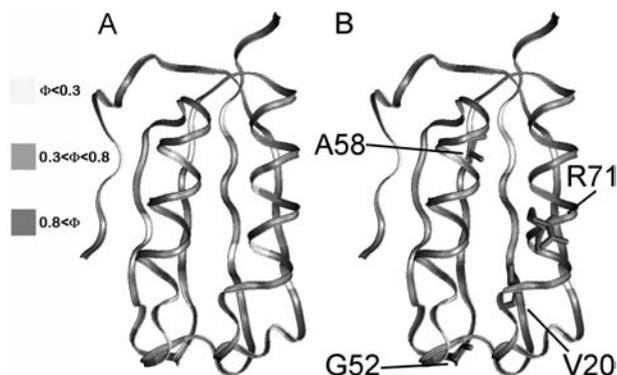


Figure 2.5: Native Sso AcP colour-coded to show the obtained  $\Phi_I^{H2O}$  (left) and  $\Phi_{\ddagger}^{H2O}$  (right) values. Residues are shown in yellow if their  $\Phi$  value is lower than 0.3, orange if the value is between 0.3 and 0.8 and red if the value is greater than 0.8. In the latter case the residue is labelled and shown also in a ribbon representation. The figures have been drawn with VMD 1.8.3 for win32 (Humphrey et al. 1996).

concentrations, a downward curvature in the folding limb of the plot reporting the folding/unfolding rate constant versus denaturant concentration. This deviation from the two-state model is similar to that observed for the wild type protein (figure 2.2B) and suggests that the partially folded ensemble forms in all the mutants that we studied. For each mutant, the various kinetic traces were analysed as described in the (see sections 2.4.6 and 2.4.7) to yield the folding and unfolding rate constants in the absence of denaturant ( $k_{I \rightarrow F}^{H2O}$  refers to the rate of formation of the native state regardless of the on- or off-pathway nature of the partially folded state, while  $k_{F \rightarrow U}^{H2O}$  refers to the unfolding rate constant). A complete list of the  $k_{I \rightarrow F}^{H2O}$  and  $k_{F \rightarrow U}^{H2O}$  values obtained for all the analysed mutants is reported in table 2.3. The thermodynamic and kinetic data were combined to obtain the  $\Phi$  values for the partially folded and transition states. The  $\Phi$  values for the partially folded ensemble ( $\Phi_I^{H2O}$ ) are generally lower than the corresponding ones for the transition state ( $\Phi_{\ddagger}^{H2O}$ ), showing a gain of structure along the folding coordinate (table 2.3). In the partially folded state the catalytic site does not appear to be fully structured. The R30A and N48A variants were not analysed, due to their low  $\Delta G_{U \rightarrow F}^{H2O}$  value. Nevertheless, the  $\Phi_I^{H2O}$  values obtained for the V24A and V27A variants are close to 0, suggesting that the catalytic 22-28 loop does not display a

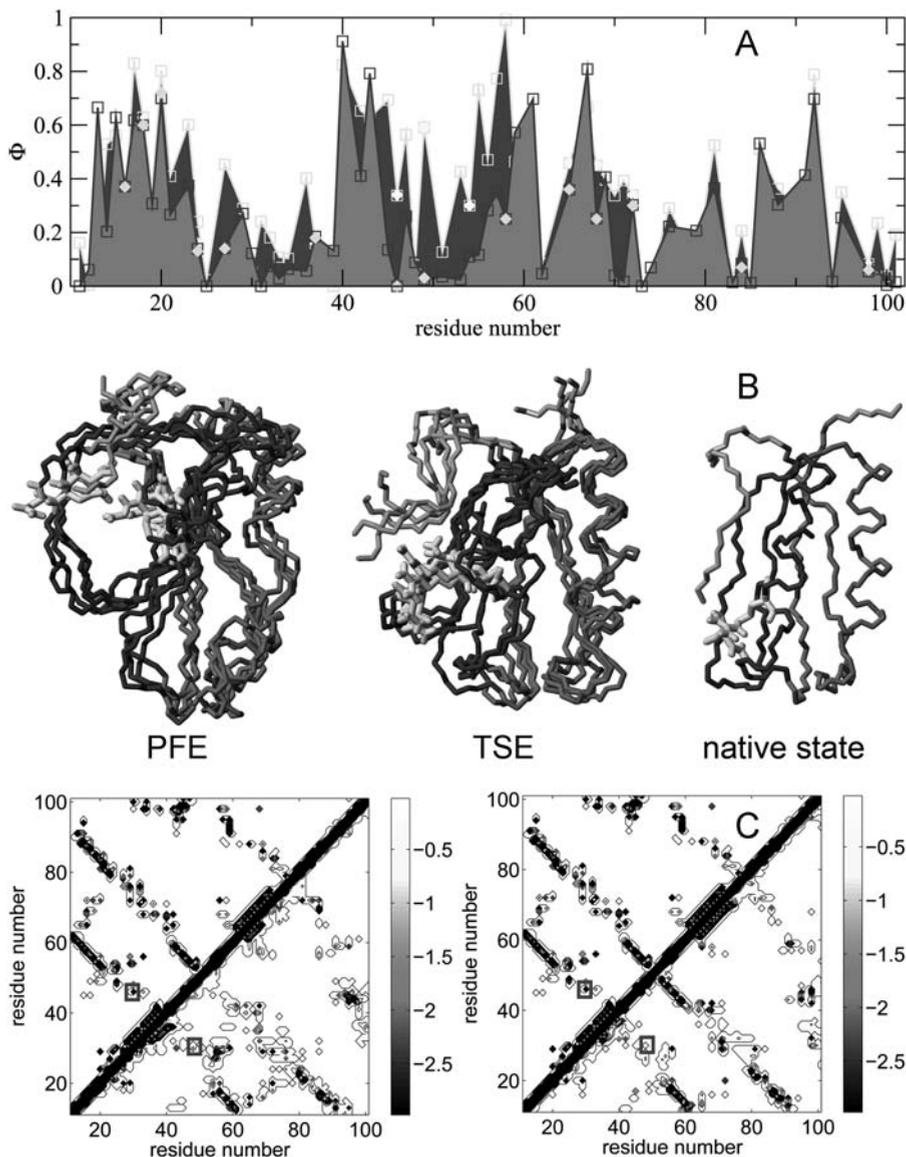


Figure 2.6: Structural properties of the partially folded (PFE) and transition state (TSE) ensembles of Sso AcP. (A) Profiles of the experimentally determined  $\Phi$  values ( $\Phi_{exp}$ ) and those calculated from the simulations ( $\Phi_{calc}$ ) for PFE and TSE of Sso AcP. The ensemble average  $\Phi_{calc}$  values of the PFE and TSE are shown in red and blue, respectively.  $\Phi_{exp}$  values of the PFE and TSE are indicated as green and yellow diamonds, respectively. (B) Comparison of the representative structures of the four biggest clusters in the PFE (left) and TSE (middle) with the X-ray structure of the native state (right); the “scaffold region” (residues 13-23 and 60-90) is shown in red, the region around the catalytic site (residues 24-59) in blue; residues Arg30 and Asn48 are highlighted in yellow. (C) Energy maps of PFE (left) and TSE (right). The pair wise interaction energies of the native state are shown above the diagonal, those of the PFE and TSE below the diagonal. The interaction between Arg30 and Asn48 is highlighted by blue squares.

native like structure in this ensemble. These data suggest that the catalytic properties of the partially folded ensemble do not arise from an overall native-like structural formation of the active site in this state. However, even if the double-jump experiment described above rules out the occurrence of global folding associated with binding, it is possible that the substrate directly participates to this positioning by binding the catalytic residues in the partially folded state and hence determining a local folding of the active site.

Moreover, the analysis of all mutants used in this study enables the folding pathway of Sso AcP to be characterised. The region that appears most structured in the ensemble of partially folded conformations is the interface between the first  $\beta$ -strand and the second  $\alpha$ -helix (figure 2.5). The rest of the molecule shows  $\Phi$  values close to 0. The transition state displays a more compact structure, with generally increased  $\Phi$  values (figure 2.5; table 2.3). In this state, the establishment of native contacts is propagated to the  $\beta$ -hairpin formed by  $\beta$ -strands 2 and 3. Four residues (Val20, Gly52, Ala58 and Arg71) appear to drive structure formation in the transition state ensemble (figure 2.5). Interestingly, Arg71 does not appear important for the overall structure stabilisation, suggesting a specific role for this residue in transition state formation.

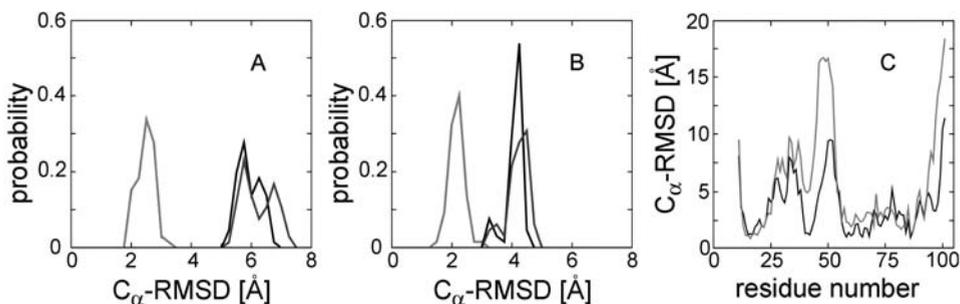


Figure 2.7: The high heterogeneity of the partially folded state is due to the catalytic region. (A, B) Distribution of the  $C_{\alpha}$ -RMSD of the entire sequence (black), the scaffold region (red), and the catalytic region (blue) from the X-ray structure in PFE (A) and TSE (B). (C): The  $C_{\alpha}$ -RMSD per residue of the PFE (red) and TSE (black) from the native state. The region around the catalytic residue Asn48 is structurally very different from the native state in PFE and becomes more native-like in the TSE.

## 2.3 Discussion

### 2.3.1 Structure of the partially folded state and of the transition state

In a collaboration with the University of Cambridge (UK) we have used molecular dynamics simulations with  $\Phi$  value restraints (Vendruscolo et al. 2001; Gsponer et al. 2006) to generate two ensembles of structures representing the partially folded and the transition states, respectively (figure 2.6A and 2.6B and section 2.4.8). The CHARMM22 force-field (MacKerell et al. 1998), an all-atom protein representation, the TIP3P water model and periodic boundary conditions have been used. The partially folded state is characterised by a significant structural heterogeneity and by the presence of several non-native interactions. The native topology is, however, rather well

preserved, as shown by the comparison of the structures of the intermediate and the native state (figure 2.6B) and by the energy maps, which provide an illustration of the most strongly interacting regions (figure 2.6C). The catalytic site (formed by residues in  $\alpha$ -helix 1,  $\beta$ -strand 2 and the loop between  $\beta$ -strand 1 and  $\alpha$ -helix 1) is very flexible, although, importantly, the Arg30 and Asn48 residues, which are the most important for the catalytic activity, remain in more than 20% of the structures closer than 6 Å to each other. Overall, the native architecture is particularly well conserved in the region of  $\alpha$ -helix 2 and  $\beta$ -strand 1 and 4. By contrast, the regions corresponding to  $\alpha$ -helix 1,  $\beta$ -strand 2, 5 and 3 are much less well structured. To quantify these observations, we have considered the protein structure as divided into two parts, the first one corresponding to  $\alpha$ -helix 1 and  $\beta$ -strands 5, 2 and 3 (catalytic region) and the second one corresponding to  $\alpha$ -helix 2 and  $\beta$ -strands 1 and 4 and we calculated the probability distributions of the  $C_\alpha$ -carbon root mean square distance from the native state ( $C_\alpha$ -RMSD) (figure 2.7A and 2.7B). The results show that the high heterogeneity of the partially folded state is due to the catalytic region (blue), which shows distances comparable to that of the entire protein (black), while the remainder of the molecule (red) virtually shows a native-like fold, with an average distance from the native state equal to about 2 Å. We refer to this region as the “scaffold region”. The presence of this scaffold region implies that, although the catalytic residues and the catalytic loop are highly dynamic in the partially folded ensemble, the overall topology of the protein is already formed in this state and this decreases the number of its accessible conformations, that is its entropy. Thus, the structure determination that we present here provides a result that is not apparent from an immediate inspection of the experimental  $\Phi$  values that are all small in this region. This analysis provides a structural basis for rationalising the maintenance of the catalytic activity in the partially folded state. For comparison, the transition state is much more native-like and comprises a particularly well structured region that includes parts of  $\beta$ -strands 1, 2, 3 and 4,  $\alpha$ -helices 1 and 2, bearing a more folded catalytic site (figure 2.6C and 2.7C).

### 2.3.2 Enzymatic activity in the presence of a highly dynamic catalytic site

We have shown that the partially structured ensemble of Sso AcP accumulating during folding prior to the formation of the native state is enzymatically active. This conclusion was reached through the series of experiments that we summarise here. We first used a technique based on the monitoring of the absorbance of the substrate to assess the catalytic activity in real time. Prior to performing experiments on the activity of the intermediate we monitored the time course over 20 seconds of the substrate absorbance under native conditions, which provided the reference rate of disappearance of the substrate (figures 2.2D and 2.2E). We then followed for a similar time the same absorbance signal under refolding conditions, which resulted in a similar trend (figures 2.2G and 2.2H). In particular, after the first four milliseconds of the refolding process less than 1% of the protein molecules are fully folded and about 80% of the native enzymatic activity was recorded (figures 2.2G, 2.2H and 2.2I). As a control experiment we carried out the refolding reaction in 7 M urea. Under these conditions the folding of Sso AcP takes place without intermediates, and, correspondingly, we detected a development of the enzymatic activity with a rate very close to the folding rate (figures 2.2G and 2.2H). As a second control experiment we performed a double-jump experiment that ruled out the presence of a significant fraction of fully folded molecules during the

first 10ms of the refolding process. The possibility of a substrate induced folding was also ruled out through a third control experiment in which two competitive inhibitors of enzymatic activity of Sso AcP were shown to leave the folding rate unaffected (figure 2.2F).

A few systems have been shown to bear enzymatic activity in the absence of folded structure. An intermediate in the folding of ribonucleaseT1 is characterised by extensive secondary and tertiary structure, a hydrophobic core with low solvent accessibility and partial enzymatic activity (Kiefhaber et al. 1992). A monomeric chorismate mutase obtained by topological redesign of a dimeric helical bundle enzyme from *Methanococcus jannaschii* shows properties typical of a molten globule state. This protein is enzymatically active and its activity is coupled with a substrate-induced folding (Vamvaca et al. 2004). Moreover, the complex formed upon binding of substrate retains high flexibility (Pervushin et al. 2007). Two variants of dihydrofolate reductase are enzymatically active and show molten globule features. They gain native-like structure in the presence of methotrexate and NADPH (Uversky et al. 1996). The catalytic site of Sso AcP that we characterised here is highly heterogeneous in the partially folded ensemble. This conclusion follows from the observation that  $\Phi$  values of residues in the catalytic 22-28 loop are close to 0. Moreover, the overall structural analysis carried out with all of the experimentally determined  $\Phi$  values confirms that the catalytic site of the protein is not yet fully structured, although it may become so upon binding of the ligand. Importantly, the folding is not accelerated in the presence of substrate analogues, suggesting that no global substrate-induced folding occurs, even though a local reorganisation of the catalytic region cannot be ruled out.

These findings indicate that an enzyme can be an efficient catalyst even in the absence of its stable native conformation and provides clues to the characterisation of the structural and dynamic features of a protein that allow the catalysis to take place in the absence of a fully structured catalytic site. Indeed, we have shown that the presence of a scaffold region, whose structure is already formed in the partially folded state, determines the topology of the entire molecule; thus, the catalytic residues, albeit highly dynamic, remain in close proximity and can hydrolyse the substrate. We have found that the activity of the partially folded state is highly sensitive to mutations, suggesting that in the partially structured ensemble a small number of native contacts stabilises the overall structure and enables the protein to hydrolyse the substrate; substitutions may therefore cause a higher degree of flexibility and allow a full inactivation of the enzyme due to an increase in entropy.

The importance of conformational changes and flexibility in enzyme catalysis has been widely recognised (Osborne et al. 2001; Eisenmesser et al. 2002; Benkovic and Hammes-Schiffer 2003; Poulsen et al. 2003; Garcia-Viloca et al. 2004). The detection of enzymatic activity in the absence of a structured catalytic site and the ability to determine the distribution of structures in these non-native states represent a key step forward in the elucidation of the protein dynamics that are required for enzyme catalysis. In addition, the identification of a highly dynamic functional site in the presence of a scaffold region that restricts the conformational space of the flexible region suggests a potentially important concept in molecular biology. It has been shown, for example, that it is possible to change the catalytic activity of an existing protein scaffold by substituting several loops and then introducing point mutations to tune the enzyme activity (Park et al. 2006). Moreover, it has been proposed that the order of formation of na-

tive contacts during folding recapitulates the emergence of topology in molecular evolution (Nagao et al. 2005). The fact that the catalytic site of Sso AcP is highly heterogeneous in the partially folded ensemble and folds afterwards -i.e. concomitantly to the formation of the native state- indicates the presence of an efficient evolutionary mechanism in which the scaffold of a protein is maintained and the regions or residues directly involved in function (for example in catalysis or binding) are allowed to mutate without compromising the overall stability of the structure. This mechanism may help to increase the rate of development of new activities and could explain how proteins with the same topology can possess very different functions.

### 2.3.3 Biological function in the absence of a three-dimensional fold

It has been proposed that a considerable fraction of eukaryotic proteins are either fully unstructured or contain significant portions of their sequence -i.e. regions longer than 50 residues- in an unstructured state (Dunker et al. 2001; Fink 2005). Intriguingly, these natively unfolded polypeptide chains are mostly involved in fundamental biological processes, such as transcription, translation and regulation of the cell cycle (Nakayama et al. 2001). The existence of such natively unfolded regions, however, is at first sight surprising as proteins have a high propensity to aggregate into deleterious misfolded structures under these conformational states (Dobson 2003). It has been suggested that, because of their high flexibility, these polypeptide chains are able to bind many substrates and interact with many targets (Uversky 2002); this is consistent with the observation that many “hubs” (i.e. proteins with a large number of interaction partners) in protein interaction networks are constituted by proteins either completely or partially disordered in solution (Dunker et al. 2005). It has also been proposed that, since natively unfolded proteins need to fold before binding, they can couple high specificity with a low affinity (Dunker et al. 2001). These proteins overcome steric restrictions, giving rise to interaction surfaces larger than those obtained from a native state (Dunker et al. 2005) and their flexibility increases the rate of specific macromolecular association (Uversky 2002). The disorder of these peptides speeds up their turnover and this favours rapid response to cell signalling in fundamental points of the protein network (Fink 2005). The widespread presence of natively unfolded proteins in living organisms suggests that proteins do not have to adopt necessarily compact globular structures to be functional, at least for the molecular recognition of their targets.

In this work we have reported evidence that a conformational state structurally distant from the native structure -particularly in those loops or residues that form the substrate binding and catalytic site- is able to bind substrates, carry out catalysis and release products. These findings extend the spectrum of possible biological functions carried out in the absence of a folded state to include enzyme catalysis. If confirmed, the results that we have presented will suggest an extension of the paradigm that specific biological functions can only be associated with unique three-dimensional folds, and thus provide an important conceptual tool for a better understanding of the complex network of biological functions, protein-protein interactions and regulatory mechanisms that are at the basis of the function of the cell.

## 2.4 Materials and methods

### 2.4.1 Mutagenesis

The gene encoding wild type Sso AcP was initially inserted in a pGEX-2T plasmid. Mutants of Sso AcP were produced by using the Quick Change site-directed mutagenesis kit® from Stratagene. In particular, the following protocol was applied to achieve desired mutation (figure 2.8):

1. Wild type plasmid was extracted from DH5- $\alpha$  *E. coli* cells using the QIA quick extraction kit (see figure 2.8B).

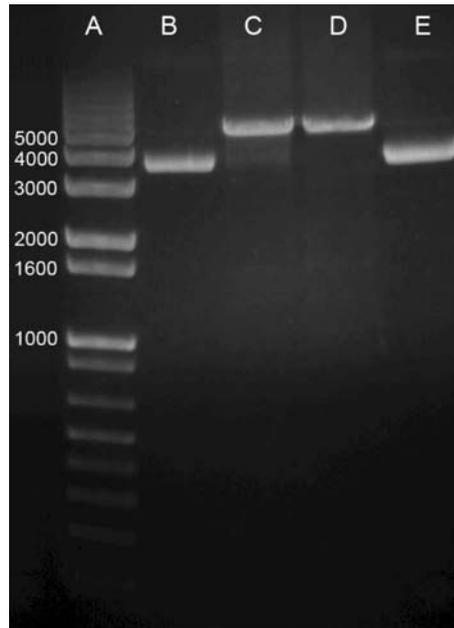


Figure 2.8: Agarose gel at 1% (w/v) showing the steps of mutagenesis. (A) 1Kb standard; length values corresponding to bands are shown on the left. (B) The extracted, supercoiled, template plasmid. (C) Result of the amplification reaction: the product is not supercoiled; a band is visible corresponding to supercoiled plasmid due to the presence of the template. (D) Result of treatment with *Dpn* I; the band of the template is now not visible. (E) Plasmid extracted from transformed cells. The plasmid is supercoiled.

2. A mutagenic PCR was carried out with 10 ng of pGEX-2T plasmid carrying the gene encoding wild type Sso AcP as template. Primers were designed to contain the desired mutation in the middle with about 15 bases of correct sequence on both sides. The melting temperature  $T_m$  was calculated to be  $\geq 78$  °C according to the following formula:

$$T_m = 81.5 + 0.41 \cdot (0.41 \cdot \%GC) - \frac{675}{N} \quad (2.1)$$

where %GC is the relative content of G and C bases and  $N$  is the primer length in bases. 125 ng of both primers and 2.5 UI of Pfu Turbo DNA polymerase were used. 19 polymerisation cycles were carried out with the following pro-

tolcol: (1) 30 seconds at 95 °C; (2) one minute at 55 °C; (3) 10 minutes at 68 °C (figure 2.8C).

3. After amplification reaction, the sample was treated with *Dpn* I (10 UI/ $\mu$ l) at 37 °C for one hour to digest the parental supercoiled double strand DNA (figure 2.8D).
4. XL1-Blue supercompetent cells were transformed with 1 ml of the sample. Cells were grown on LB plates containing 0.1  $\mu$ g ml<sup>-1</sup> ampicillin.
5. The plasmid carrying the desired mutation was extracted from colonies (figure 2.8E). Presence of the desired mutation was assessed by DNA sequencing. Cells were stored in glycerol at -80 °C.



Figure 2.9: SDS-polyacrilamide gel electrophoresis showing the different steps of purification. (A) Sigma wide range standard; mass values in kDa corresponding to bands are shown on the left. (B and C) Pellet (B) and supernatant (C) separated by centrifugation after cell lysis; the band at 37 kDa corresponds to the GST-Sso AcP fusion protein (see text). (D) Flow-through harvested at the bottom of the column after application of the supernatant containing GST-Sso AcP fusion protein. The 37 kDa band has disappeared as fusion protein is bound to the resin. (E and F) Two washing steps with PBS (E) and TRIS (F) buffer. After the second step non-specific proteins are not present in the column. (G) Sso AcP eluted after thrombin cleavage. (H) GST eluted after washing the column with glutathione.

#### 2.4.2 Protein expression and purification

Expression and purification of wild type protein and mutants were carried out by affinity chromatography as previously described (Modesti et al. 1995). In particular the following protocol was applied (figure 2.9):

1. Plasmid carrying the gene encoding the Sso AcP protein variant was extracted and transformed into BL21 *E. coli* competent cells.
2. Cells were grown over night in a LB medium with 0.1  $\mu$ g ml<sup>-1</sup> ampicillin.
3. Expression of Glutathione-S-transferase/Sso AcP fusion protein (GST-Sso AcP) was induced with isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG, Inalco,

Milan, Italy) at a final concentration equal to 50 mg ml<sup>-1</sup>. After three hours of expression cells were separated from medium by centrifugation (15 minutes at 7,000 × g) and resuspended with a 20 mM sodium phosphate and 250 mM NaCl buffer at pH 7.3 (PBS) with ethylenediaminetetraacetic acid (EDTA) 1mM, β-mercaptoethanol 1 mM and phenylmethylsulfonyl fluoride (PMSF) 0.1 mM. Cells were stored at -20 °C.

4. Cell lysis was carried in three steps: (1) thawing cells; (2) adding lysozyme to a final concentration of 1 mg ml<sup>-1</sup>; (3) 6 sonication cycles of 30 seconds. The obtained sample was centrifuged for 40minutes at 39,000 × g (figure 2.9B and 2.9C).
5. The supernatant was applied to a column containing glutathione-agarose resin (Sigma) (figure 2.9D).
6. After two washing steps with buffers (figure 2.9E and 2.9F), 10 ml of a 50 mM 2-amino-2-hydroxymethyl-1,3-propanediol (TRIS) and 150 mM NaCl buffer at pH 8.0 containing 50 UI of human thrombin (Sigma) were added to the column. Thrombin cleavage was carried out over night.
7. The eluted protein was concentrated with centriplus (Millipore) and checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and electrospray mass spectrometry (figure 2.9G and 2.9H). Protein concentration *c* was measured according to the Lambert-Beer law:

$$c = \frac{A_{280}}{\epsilon_{280} \cdot l} \quad (2.2)$$

Where  $A_{280}$  is the absorbance at 280 nm measured in a JascoV-630 spectrophotometer (Tokio, Japan) spectrophotometer,  $\epsilon_{280}$  is the extinction coefficient calculated as reported (Gill and von Hippel 1989) and  $l$  is the cell length. Protein was stored at -20 °C.

#### 2.4.3 Enzymatic activity essay

Enzymatic activity of native Sso AcP was measured in a continuous optical test at 283 nm using benzoyl-phosphate (BP) as a substrate ((Ramponi et al. 1966) and figure 1.6I) with a Lambda 4V Perkin Elmer spectrophotometer (Wellesley, Massachusetts). Experimental conditions were 2.0 µg ml<sup>-1</sup> Sso AcP, 5.0 mM BP, 50 mM acetate buffer at pH 5.5, 37 °C. BP was synthesised as previously described (Camici et al. 1976) and freshly dissolved before enzymatic activity measurements.

#### 2.4.4 Development of enzymatic activity during folding

A Bio-logic SFM-3 stopped-flow device (Claix, France) coupled with an absorbance detection system and thermostated with a RTE-200 water circulating bath from Neslab (Newington, New Hampshire) was used to measure the recovery of enzymatic activity during folding. Sso AcP was initially unfolded at a concentration equal to 0.4 mg ml<sup>-1</sup> in 5.5 M GdnHCl (Sigma-Aldrich), 50 mM acetate buffer, pH 5.5, at 37 °C. 20 µl aliquots of this sample were mixed with 380 µl of a solution containing 5.27 mM BP, 50 mM acetate buffer, pH 5.5. Final conditions in the assay test were 0.02 mg ml<sup>-1</sup> Sso AcP, 0.275 M GdnHCl, 10 mM BP, 50mM acetate buffer, pH5.5, 37 °C. The experimental dead time ranged from 4 to 20 ms. The cuvette length was 1 cm. The signal at 283 nm ( $A_{283}$ ) was acquired during protein refolding. Since the decrease of this signal is

proportional to the enzyme activity, for each obtained trace the decay rate of absorbance at 283 nm was calculated, after a smoothing on a 0.02 s sliding window, as the opposite of the first order derivative ( $-dA_{283(t)}/dt$ ). The obtained values were normalised to the activity of the native wild type protein, plotted versus time and fitted to the following equation:

$$-\frac{d}{dt}A_{283(t)} = C_{tot} \left[ k_{CAT}^F - (k_{CAT}^F - k_{CAT}^I) \cdot e^{-k_{I \rightarrow F} \cdot t} \right] \quad (2.3)$$

where  $t$  is the time,  $-dA_{283(t)}/dt$  is the enzymatic activity measured during folding,  $C_{tot}$  is the total protein concentration,  $k_{CAT}^F$  and  $k_{CAT}^I$  are  $k_{CAT}$  of native and partially folded states and  $k_{I \rightarrow F}$  is the main folding rate. Derivation of equation 2.3 is reported in appendix A (see section A.4).

#### 2.4.5 Equilibrium GdnHCl-induced unfolding curves

For each mutational variant of Sso AcP, 28 samples containing 0.2 mg ml<sup>-1</sup> of the tested protein, 50mMacetate buffer, pH 5.5 and a GdnHCl concentration ranging from 0 to 7.2 M, were incubated for 2 h at 37 °C to reach equilibrium. After this time, the circular dichroism (CD) signal at 222 nm was acquired for all samples using a 0.1 cm path-length cuvette in a Jasco J-810 CD spectropolarimeter (Great Dunmow, Essex, United Kingdom) thermostated with a C25P Thermo Haake water circulating bath (Karlsruhe, Germany). The mean residue ellipticity was plotted versus GdnHCl concentration. The obtained plots were fitted to a two state transition according to the equation described by Santoro and Bolen (Santoro and Bolen 1988) to obtain the free energy difference between the unfolded and the native states in the absence of denaturant ( $\Delta G_{U-F}^{H_2O}$ ), the dependence of the  $\Delta G_{U-F}^{H_2O}$  on GdnHCl concentration ( $m$  value) and the midpoint of denaturation ( $C_m$ ).

The fraction folded at each GdnHCl concentration was calculated as described (Chiti et al. 1998). For each mutant, the change in conformational stability upon mutation,  $\Delta \Delta G_{U-F}^{H_2O}$  was calculated according to

$$\Delta \Delta G_{U-F}^{H_2O} = (C_m - C'_m) \cdot \bar{m} \quad (2.4)$$

where  $C'_m$  and  $C_m$  are the mid-denaturation concentrations for the considered mutant and the wild type, respectively, and  $\bar{m}$  is the average  $m$  value over the wild type and 34 mutants that we studied here. The average  $m$  value was used following (Matouschek and Fersht 1991), since the  $m$  value obtained by the best fit of a single mutant to the Santoro & Bolen model arises from the few points in the transition zone of the plot (see figure 2.4A) and is therefore highly sensitive to the experimental error.  $C_m$ ,  $\Delta G_{U-F}^{H_2O}$  and individually calculated  $m$  values are reported in table 2.2.

#### 2.4.6 Folding kinetics

Folding experiments were carried out using the Bio-logic SFM-3 stopped flow device (Claix, France) equipped with a fluorescence detection system and thermostated with a RTE-200 water circulating bath from Neslab (Newington, New Hampshire). An excitation wavelength of 280 nm and a band-pass filter to monitor emitted fluorescence above 320 nm were used. The cuvette path-length was 0.15 cm. 20  $\mu$ l aliquots of 0.4 mg ml<sup>-1</sup> protein unfolded in 5.5 M GdnHCl were mixed with 380  $\mu$ l of refolding buffer. Final condi-

tions were 0.02 mg ml<sup>-1</sup> protein, 50 mM acetate buffer, pH 5.5, a GdnHCl concentration ranging from 0.2 to 2.5 M, 37 °C. The experimental dead time was 10.4 ms. The obtained traces were fitted to a double exponential equation of the following form:

$$f(t) = A_1 \cdot e^{(-k_{12}t)} + A_2 \cdot e^{(-k_{I \rightarrow F}t)} + q \quad (2.5)$$

where  $f(t)$  is the fluorescence as a function of time  $t$ ,  $A_1$  and  $A_2$  are amplitudes,  $q$  is the equilibrium signal and  $k_{12}$  and  $k_{I \rightarrow F}$  are folding constants. Derivation of equation 2.5 is reported in appendix A (section A.2.1). The main folding rate constant was then plotted versus denaturant concentration to extrapolate the folding rate constant in the absence of denaturant ( $k_{I \rightarrow F}^{H_2O}$ ) as reported in appendix A (section A.2.3). However,  $k_{I \rightarrow F}^{H_2O}$  refers to the rate of formation of the native state regardless of the on- or off-pathway nature of the partially folded state. In another set of experiments 20  $\mu$ l aliquots of 0.4 mg ml<sup>-1</sup> wild type Sso AcP unfolded in 5.5 M GdnHCl were mixed with 380  $\mu$ l of refolding buffer containing different concentrations of inorganic phosphate and phenyl phosphate. Final conditions were 0.02 mg ml<sup>-1</sup> Sso AcP, 0.275 M GdnHCl, 50 mM acetate buffer, pH 5.5, a phosphate and phenyl phosphate concentration ranging from 0.1 to 10 mM, 37 °C. The resulting traces were fitted to equation 2.5.

#### 2.4.7 Unfolding kinetics

Unfolding experiments were carried out using a Bio-logic SFM-20 stopped flow device (Claix, France) equipped with a Jasco J-810 circular dichroism detection system (Great Dunmow, Essex, United Kingdom) and thermostated with a C25P Thermo Haake water circulating bath (Karlsruhe, Germany). The signal was recorded at 230 nm with a slit window of 4.0 nm. The cuvette path-length was 0.2 cm. 85  $\mu$ l aliquots of 1.4 mg ml<sup>-1</sup> protein in 1.0 M GdnHCl were mixed with 215  $\mu$ l of a solution containing 8.0 M GdnHCl, 50 mM acetate, pH 5.5. Final conditions were 0.4 mg ml<sup>-1</sup> protein, 50 mM acetate buffer, pH 5.5, 6.0 M GdnHCl, 37 °C. The experimental dead time was 74 ms. To obtain the unfolding rate constant ( $k_{F \rightarrow U}$ ), 6 to 10 traces were averaged and fitted to a single exponential equation of the following form:

$$[\Theta]_{(t)} = A \cdot e^{(-k_{F \rightarrow U}t)} + q. \quad (2.6)$$

where  $[\Theta]_{(t)}$  is the CD signal as a function of time  $t$ ,  $A$  is the amplitude and  $q$  is the equilibrium signal. Derivation of equation 2.6 is reported in appendix A (section A.2.2). The unfolding rate constant in the absence of denaturant ( $k_{F \rightarrow U}^{H_2O}$ ) was obtained with a linear extrapolation method using a previously obtained slope (Bemporad et al. 2004).

#### 2.4.8 $\Phi$ value analysis

A  $\Phi$  value analysis was carried out on the partially folded state and the transition state ensemble populated during the Sso AcP folding process. The  $\Phi$  values for the transition state ensemble (Matouschek and Fersht 1991),  $\Phi_{\ddagger}^{H_2O}$ , were calculated according to

$$\Phi_{\ddagger}^{H_2O} = 1 - \left[ \frac{-RT \ln \left( \frac{k_{F \rightarrow U}^{H_2O}}{k_{F \rightarrow U}} \right)}{(C'_m - C_m) \cdot m} \right] \quad (2.7)$$

where  $C'_m$  and  $C_m$  are the mid-denaturation concentrations for the considered mutant

and the wild type, respectively,  $R$  is the ideal gas constant ( $8.314 \text{ J mol}^{-1} \text{ K}^{-1}$ ),  $T$  is the temperature in K (310.15 K),  $k'_{F \rightarrow U}{}^{H2O}$  and  $k_{F \rightarrow U}{}^{H2O}$  are the unfolding rate constants in the absence of denaturant for the mutant and the wild type protein, respectively. Derivation of equation 2.7 is shown in appendix A (see section A.3).  $\Phi$  values for the partly folded state,  $\Phi_I{}^{H2O}$  (Matouschek et al. 1992) were calculated according to

$$\Phi_I{}^{H2O} = 1 - \left[ \frac{-RT \ln \left( \frac{k'_{I \rightarrow F}{}^{H2O} \cdot k_{F \rightarrow U}{}^{H2O}}{k_{I \rightarrow F}{}^{H2O} \cdot k'_{F \rightarrow U}{}^{H2O}} \right)}{(C'_m - C_m) \cdot \bar{m}} \right] \quad (2.8)$$

where  $k'_{I \rightarrow F}{}^{H2O}$  and  $k_{I \rightarrow F}{}^{H2O}$  are the folding rate constants in the absence of denaturant for the mutant and the wild type protein, respectively. Derivation of equation 2.8 is shown in appendix A (see section A.3).



## Chapter 3

### Aggregation studies on Sso AcP

#### 3.1 Introduction

##### 3.1.1 Aggregation from native states

As introduced in section 1.3.2 the “conformational change hypothesis” can account for the aggregation of most peptides. Nevertheless, evidence is now emerging that native folded states retain a significant, albeit small, propensity to aggregate (section 1.3.2 and (Bemporad et al. 2006; Chiti and Dobson 2006)). Indeed, edge  $\beta$ -strands are potentially dangerous as they are already in the right conformation to interact with any other  $\beta$ -strand they encounter. This can be the initial step that triggers the formation of amyloid-like aggregates. This natural tendency to aggregate of proteins and edge  $\beta$ -strands is kept under control by several evolutionary strategies (Richardson and Richardson 2002; Monsellier and Chiti 2007; Monsellier et al. 2007). For example,  $\alpha$ -helix proteins cover their  $\beta$ -sheet ends with loops of different length and structure. In many cases, edge  $\beta$ -strands are protected by presence of particular structure, such as  $\beta$ -bulges or particular residues, such as proline, that force the strand in a conformation with low propensity to aggregate (Richardson and Richardson 2002). Finally, in some cases edge  $\beta$ -strands are very short and this prevents their aggregation (Richardson and Richardson 2002).

Despite the existence of evolutionary strategies to keep under control this process, some systems are able to aggregate starting from an ensemble of native-like conformations (Bemporad et al. 2006; Chiti and Dobson 2006). In the case of insulin, for example, aggregation at low pH is preceded by an oligomerization step in which a native-like content of  $\alpha$ -helical structure is almost completely retained, and aggregates with a morphology reminiscent of amyloid protofibrils and with a high content of  $\beta$ -structure appear only later in the process (Bouchard et al. 2000). In the case of ataxin-3, the protein associated with spinocerebellar ataxia type-3, a polyglutamine insertion strongly enhances aggregation propensity but does not affect native state stability. This led the authors to propose a model for amyloid aggregation in which the pathways of unfolding and misfolding are distinct and separate (Chow et al. 2004). In addition, the S6 protein from *Thermus thermophilus* adopts a quasi-native state at pH 2.0, 0.4 M NaCl, and 42 °C. Under stirring, the protein grows into fibrils after several days. Interestingly, kinetic analysis revealed that longer lag phases in aggregation correlate with faster unfolding rates in a number of variants, suggesting that the native-like state rather than an ensemble of highly fluctuating conformations, participates in the nucleation of the fibrillation process (Pedersen et al. 2004). Finally, in the case of the yeast

prionUre2p, it was shown that a native-like conformation is retained also in the fibrils, suggesting a native-like state as the initial conformer that initiates amyloid-like aggregation (Bousset et al. 2002; Bousset et al. 2004a; Bousset et al. 2004b).

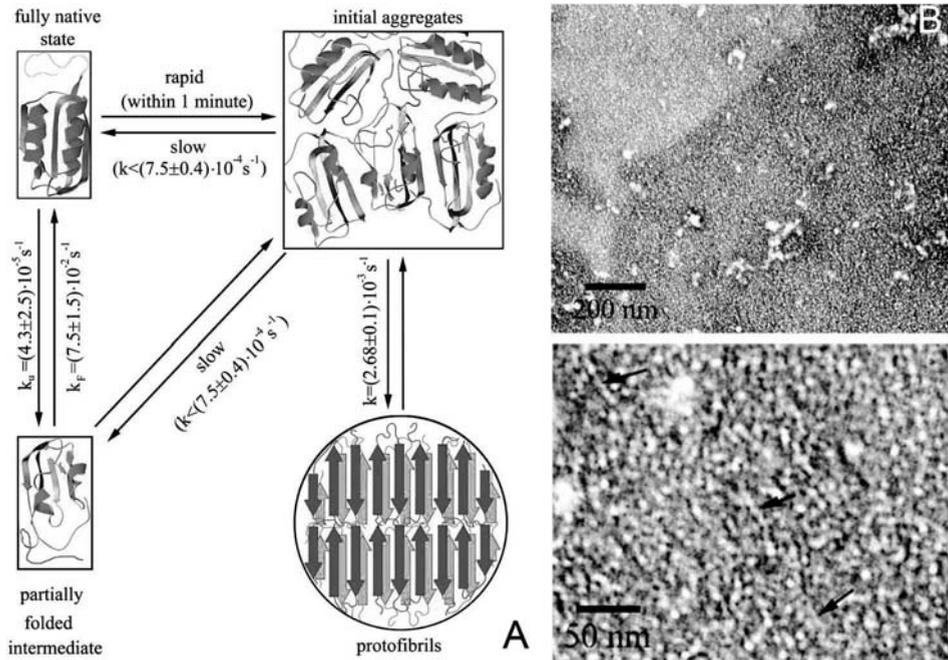


Figure 3.1: Aggregation of Sso AcP; reprinted from (Plakoutsi et al. 2005). (A) The aggregation process of Sso AcP. An ensemble of native-like conformations gives rise to the formation of early aggregates in which the native structure is retained. These convert afterwards in amyloid-like protofibrils. (B) Electron micrograph at two different magnifications of negatively stained Sso AcP aggregates formed after one hour in 20% (v/v) TFE, 50 mM sodium acetate at pH 5.5 and 25 °C. Small aggregates consisting of globules and short thin fibrils are visible throughout the samples and form the background of the grid. These species have diameters of 3-5 nm. Elongated protofibrils are indicated by arrows.

The ability to aggregate of native conformations is not surprising if we consider that native states are actually ensembles of a multitude of conformers (Lindorff-Larsen et al. 2005a). Some of these conformers will be only transiently populated but could be significant for aggregation just as they are for the hydrogen exchange of their main-chain amide groups (Chiti and Dobson 2006).

### 3.1.2 Aggregation of Sso AcP

Two members of the acylphosphatase family (section 1.4.1) aggregate starting from an ensemble of native-like conformations (Bemporad et al. 2006). The acylphosphatase from *Drosophila melanogaster* (AcP Dro2) forms amyloid-like fibrils under conditions in which the protein has initially a secondary structure, hydrodynamic diameter, catalytic activity, and packing around hydrophobic residues indistinguishable from

those of the native state (Soldi et al. 2006a). Importantly, conformational stabilities measured in native and aggregation promoting conditions do not significantly differ and the protein does not need to unfold to initiate aggregation (Soldi et al. 2006a).

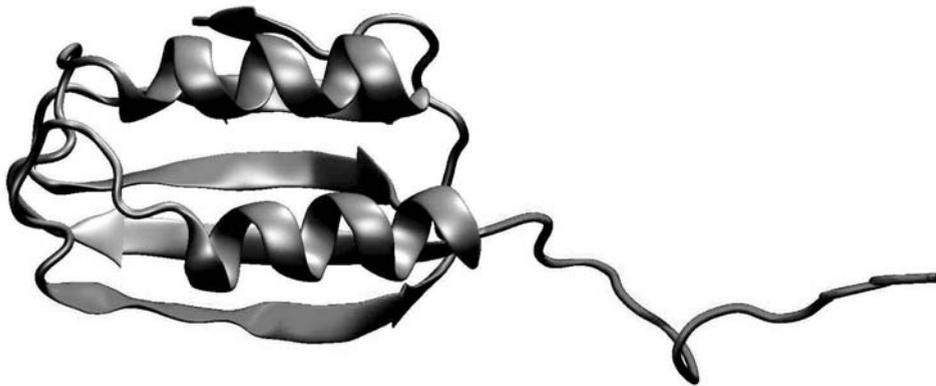


Figure 3.2: Structure of Sso AcP. Ribbon representation of the Sso AcP structure (Corazza et al. 2006). The regions that were shown to be important in amyloid-like aggregation are depicted in red. These correspond to N-terminal segment and fourth  $\beta$ -strand (Plakoutsi et al. 2006). The figure has been drawn with VMD 1.8.3 for win32 (Humphrey et al. 1996).

The acylphosphatase from *Sulfolobus solfataricus* (Sso AcP) aggregates starting from an ensemble of native-like conformations as well. The protein is able to give rise to amyloid-like protofibrils in about one hour, in 15-25 % 2,2,2-Trifluoroethanol (TFE) at pH 5.5 and 25 °C (Plakoutsi et al. 2004; Plakoutsi et al. 2005). Importantly, in these conditions, before aggregation occurs, Sso AcP has a considerable enzymatic activity (Plakoutsi et al. 2004) and native-like far- and near-UV circular dichroism (CD) spectra (Plakoutsi et al. 2006). Moreover, folding is faster than unfolding (Plakoutsi et al. 2004). Aggregation occurs from this native-like state in two phases (see figure 3.1).

1. In a first phase the Sso AcP molecules interact giving rise to an aggregated species, referred to as early aggregates, that retain native-like CD spectra and enzymatic activity ((Plakoutsi et al. 2005) and figure 3.1A). Importantly, this species does not bind to Congo red (CR) and Thioflavin T (ThT) dyes, suggesting no amyloid-like conformation (Plakoutsi et al. 2005). Finally, early aggregates do not show any increase in  $\beta$ -structure, as observed by means of Fourier transform infrared (FTIR) spectroscopy (Plakoutsi et al. 2005).
2. Only afterwards, in a second phase, the early aggregates convert into amyloid-like protofibrils (figure 3.1B). These species do not show enzymatic activity (Plakoutsi et al. 2005). Importantly, they bind both ThT and CR dyes and possess extensive  $\beta$ -structure (Plakoutsi et al. 2005). The rate of this phase does not seem to depend on protein concentration, suggesting that this phase is an intra-molecular reorganisation rather than an elongation phase (Plakoutsi et al. 2005). Finally, protofibrils appear as thin filaments with a diameter of 3-5 nm (figure 3.1B), when analysed with transmission electron microscopy (TEM) (Plakoutsi et al. 2005).

In a recent study, we tried to investigate in detail the mechanism of aggregation of Sso AcP and the regions that promote aggregation of the protein. As introduced in section 1.3.2 Sso AcP possesses an 11 residue, unstructured N-terminal segment ((Corazza et al. 2006) and figure 3.2). Importantly, neither the unstructured segment nor the globular part of Sso AcP without segment ( $\Delta N11$  Sso AcP) are able to aggregate in conditions that promote aggregation of the wild type protein (Plakoutsi et al. 2006). This clearly suggests that the N-terminal segment plays a major role in promoting aggregation of Sso AcP. Moreover, limited proteolysis data, coupled to hydrogen/deuterium exchange experiments and equilibrium unfolding experiments carried out on a set of Sso AcP protein variants, allowed the fourth  $\beta$ -strand of the molecule to be identified as an important region in promoting the aggregation of the molecule (Plakoutsi et al. 2006). Importantly, this strand is an edge  $\beta$ -strand of Sso AcP, confirming the importance of protecting these regions in prevention of native state aggregation.

Despite the identification of these two regions as major determinants of the amyloid-like aggregation of Sso AcP (see figure 3.2), the aggregation mechanism of the protein is still unclear. Moreover, it was shown that  $\Delta N11$  Sso AcP has a conformational stability higher than wild type protein (Plakoutsi et al. 2006). This could suggest that this segment induces aggregation of the molecule through a destabilisation effect. Finally, no information on the structure of early aggregates and protofibrils has been so far collected. In the following sections we shall get further insight into the role of the N-terminal segment in the process studying the effect of changing its position in the sequence on the aggregation rate and mechanism. Moreover, we shall investigate the ability of this segment to affect wild type and  $\Delta N11$  Sso AcP behaviour in aggregation promoting conditions. The role of the destabilisation induced by the N-terminal segment and the dependence of the rates of the two aggregation phases on Sso AcP concentration shall be studied. Finally, we shall use a fluorescent probe, acrylodan, to get further insight into the regions buried in early aggregates and protofibrils. The obtained results allow to rule out several models reported in appendix B and shall be discussed on section 3.3 to obtain a model for the aggregation of Sso AcP that recapitulates all the experimental evidences collected so far on this system.

## 3.2 Results

### 3.2.1 Sso AcP aggregates regardless of the position of the N-terminal segment

As mentioned above, the aggregation properties of Sso AcP suggest a major role in the process for the 11 residue unstructured N-terminal segment and for the fourth  $\beta$ -strand, positioned at the edge of the protein (figure 3.2 and (Plakoutsi et al. 2006)). However, the role played by these two regions in the aggregation process is still unclear (Plakoutsi et al. 2006). To get insights into the mechanism of amyloid-like aggregation of Sso AcP, we have produced a mutant in which the unstructured segment is moved from N-terminus to C-terminus. Importantly, in this mutant the primary sequence of both segment and the globular part of Sso AcP do not change. However, N-terminus and C-terminus are far from each other (figure 3.2) and this different positioning offers an unique opportunity to check possible intra-molecular interactions be-

tween unstructured segment and globular part of Sso AcP. We will refer to this mutant as C-tail Sso AcP.

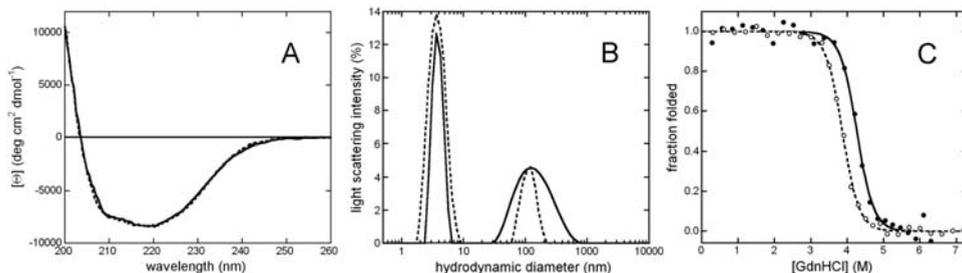


Figure 3.3: Properties of native C-tail Sso AcP. (A) Far UV circular dichroism spectra of wild type (continuous line) and C-tail (dashed line) Sso AcP in 10 mM TRIS buffer at pH 8.0, 25 °C. Mean residue ellipticity is shown versus wavelength. The overlapping of spectra suggests similar secondary structure contents in native states of these protein variants. (B) Dynamic light scattering spectra of wild type (continuous line) and C-tail (dashed line) Sso AcP in 10 mM TRIS buffer at pH 8.0, 25 °C. An apparent diameter equal to  $3.7 \pm 0.1$  nm has been measured for both monomers. (C) Equilibrium unfolding curves of wild type (filled circles) and C-tail (empty circles) Sso AcP carried out in 50 mM acetate buffer at pH 5.5, 37 °C. The relative amount of native protein (folded fraction) is reported versus denaturant concentration. Continuous and dashed lines represent best fits of experimental data to the Santoro & Bolen model for wild type and C-tail Sso AcP, respectively ((Santoro and Bolen 1988) and section A.1.1).

Since the aggregation of Sso AcP starts from an ensemble of native-like conformations we have checked the effect of the mutation on the native state of the protein. Figure 3.3A shows a comparison between far-UV CD spectra of wild type Sso AcP and C-tail Sso AcP recorded in 10 mM TRIS buffer at pH 8.0 and 25 °C. Both spectra are typical CD spectra of  $\alpha + \beta$  globular proteins. Moreover, the overlapping of these spectra suggests that moving the 11 residue segment from N-terminus to C-terminus does not affect the secondary structure content of the globular part of the molecule. In a second experiment we have measured the hydrodynamic diameter of wild type Sso AcP and C-tail Sso AcP in 10 mM TRIS buffer at pH 8.0 and 25 °C (figure 3.3B). Both proteins show a peak at  $3.7 \pm 0.1$  nm. This value is consistent, within the experimental error, with the average diameter determined by 1H-NMR (Corazza et al. 2006). This result shows that the mutation inserted in C-tail Sso AcP does not affect the compactness of the native state. Then, we have measured the conformational stability of C-tail Sso AcP in an equilibrium unfolding experiment carried out in 50 mM acetate buffer at pH 5.5 and 37 °C. The obtained plot is reminiscent of a two-state cooperative transition. Thus, it has been analysed with the method provided by Santoro & Bolen ((Santoro and Bolen 1988) and section A.1.1; figure 3.3C). Results of this analysis show an  $m$  value equal to  $11.4 \pm 0.5$  KJ mol<sup>-1</sup> M<sup>-1</sup> for both wild type and C-tail Sso AcP. The concentration of middle denaturation ( $C_m$ ) is  $4.2 \pm 0.1$  M and  $3.8 \pm 0.1$  M for wild type and C-tail Sso AcP, respectively. The free energy change upon denaturation ( $\Delta G_{U-F}^{H2O}$ )  $43.0 \pm 2$  KJ mol<sup>-1</sup> for C-tail Sso AcP. The  $\Delta G_{U-F}^{H2O}$  value for wild type Sso AcP determined in the same conditions is equal to  $48.0 \pm 2.0$  KJ mol<sup>-1</sup> (Bemporad et al. 2004; Corazza et al. 2006). This shows that moving the unstructured segment from N-terminus to C-terminus induces

a slight destabilisation of the protein. Finally, we have measured the enzymatic activity of C-tail Sso AcP in 50 mM acetate buffer at pH 5.5 and 25 °C using benzoyl-phosphate (BP) as a substrate. The obtained values are  $190 \pm 20 \text{ s}^{-1}$  for wild type Sso AcP (Corazza et al. 2006) and  $152 \pm 20 \text{ s}^{-1}$  for C-tail Sso AcP. This decrease in enzymatic activity is probably due to a steric effect induced by moving the unstructured segment to the C-terminus, which is much closer to the catalytic site of Sso AcP than N-terminus (Corazza et al. 2006). Taken together, these results show that moving the unstructured segment from the N-terminus to the C-terminus does not affect the structural parameters of the globular part of Sso AcP.

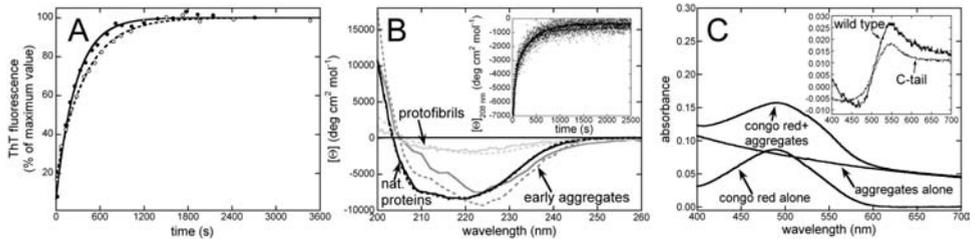


Figure 3.4: Aggregation of C-tail Sso AcP. (A) ThT fluorescence during aggregation of wild type (●) and C-tail (○) Sso AcP 34  $\mu\text{M}$  in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE, 25 °C. The ThT signal has been normalised to the plateau value. Lines represent best fits of experimental data to equation 3.3. (B) Circular dichroism spectra, reported as mean residue ellipticity versus wavelength, of wild type (continuous lines) and C-tail (dashed lines) Sso AcP. Spectra are shown for native proteins (black lines) in 10mMTRIS buffer at pH 8.0 and 25 °C, early aggregates (dark grey lines) and protofibrils (light grey lines) in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE 25 °C. Early aggregates spectra have been extrapolated using equation 3.2 as reported in section 3.4. The inset shows the change in mean residue ellipticity at 208 nm over aggregation time for wild type (●) and C-tail (○) Sso AcP 34  $\mu\text{M}$  in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE, 25 °C. Continuous lines represent best fits of experimental data to equation 3.2. (C) CR staining of C-tail Sso AcP. Spectra for CR alone, aggregates alone and CR red in the presence of protofibrils are labelled. The inset shows the spectra obtained subtracting the contributions of CR alone and aggregates alone from the spectrum of CR in the presence of aggregates, for wild type (continuous line) and C-tail (dashed line) Sso AcP. The presence of the peak at 540 nm suggests ordered aggregates.

We have studied the behaviour of C-tail Sso AcP in conditions that induce amyloid-like aggregation of wild type protein, that is 34  $\mu\text{M}$  protein in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE, 25 °C (figure 3.4). Both first and second phase have been investigated. The results show that, in aggregation conditions, C-tail Sso AcP induces the same increase in ThT fluorescence as the wild type protein ((Plakoutsi et al. 2004; Plakoutsi et al. 2005) and figure 3.4A). Aggregation rate constants, determined by best fits of experimental data to equation 3.3 are  $(3.7 \pm 0.4) \cdot 10^{-3} \text{ s}^{-1}$  for wild type Sso AcP and  $(2.5 \pm 0.3) \cdot 10^{-3} \text{ s}^{-1}$  for C-tail Sso AcP (figure 3.4A). Moreover, the species populated by C-tail Sso AcP at the plateau of the ThT kinetic experiment binds to CR dye inducing the same shift in peak wavelength as the wild type protein (Plakoutsi et al. 2004) (figure 3.4C). To monitor formation of early aggregates, aggregation of C-tail Sso AcP has been followed also with circular dichroism. Similarly to wild type protein, recorded traces show two distinct phases (inset in figure 3.4B). The first phase

corresponds to the formation of early aggregates. Aggregation rate constants are  $(2.5 \pm 0.3) \cdot 10^{-2} \text{ s}^{-1}$  for wild type Sso AcP and  $(2.6 \pm 0.3) \cdot 10^{-2} \text{ s}^{-1}$  for C-tail Sso AcP. The second phase corresponds to the process monitored by ThT kinetics. Aggregation rate constants determined by best fits of experimental data to equation 3.2 are  $(3.0 \pm 0.3) \cdot 10^{-3} \text{ s}^{-1}$  for wild type Sso AcP and  $(2.9 \pm 0.3) \cdot 10^{-3} \text{ s}^{-1}$  for C-tail Sso AcP. These values are consistent, within the experimental error, with values determined by ThT kinetics. CD spectrum of early aggregates formed by C-tail Sso AcP shows properties similar to the spectrum recorded for early aggregates formed by wild type Sso AcP, with a single negative peak at about 224 nm (figure 3.4B). Taken together, these data show that the positioning of the unstructured segment of Sso AcP does not affect the aggregation process. These observations rule out possible models for the aggregation mechanism of the protein in which a specific intra-molecular interaction is supposed to be the fundamental step that leads to the formation of an aggregation prone monomer (see models B.2.5 and B.2.11 in appendix B). In fact, N-terminus and C-terminus are far from each other and moving the segment to a different position should affect any specific interaction.

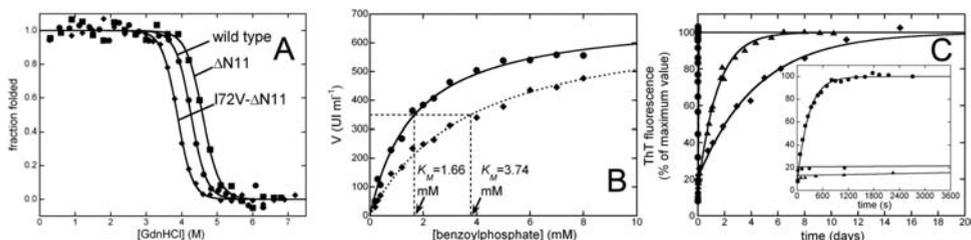


Figure 3.5: Investigation on the effect on Sso AcP aggregation of destabilisation induced by N-terminal segment. (A) Equilibrium unfolding curves of wild type (●)  $\Delta$ N11 (■) and I72V- $\Delta$ N11 (◆) Sso AcP carried out in 50 mM acetate buffer at pH 5.5, 37 °C. The relative amount of native protein (folded fraction) is reported versus denaturant concentration. Continuous lines represent best fits of experimental data to the Santoro & Bolen model (Santoro and Bolen 1988). Elimination of a methyl group from the hydrophobic core of the Sso AcP globular part results in a protein variant more destabilised than the protein lacking the unstructured N-terminus. (B) Determination of  $K_i$  of Sso AcP and phosphate. Michaelis-Menten plot of 5  $\mu$ M  $\Delta$ N11 Sso AcP in 50 mM acetate buffer pH 5.5 with 20% (v/v) TFE in the absence (●) and in the presence (◆) of 1.5 mM phosphate. Apparent  $K_M$  values are shown. (C) ThT fluorescence during aggregation of Sso AcP in different conditions. Traces are shown for wild type 34  $\mu$ M in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25 °C (●), wild type 34  $\mu$ M in 44.1 mM acetate and 4.8 mM phosphate buffer at pH 5.5, 20% (v/v) TFE and 25 °C (▲), I72V- $\Delta$ N11 34  $\mu$ M in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25 °C (◆). The inset shows the first hour of recording. Continuous lines represent best fits of experimental data to equation 3.3.

### 3.2.2 N-terminal segment does not induce Sso AcP aggregation via a destabilising effect

As mentioned above, we previously showed that the Sso AcP protein variant lacking the 11 residue N-terminal tail ( $\Delta$ N11 Sso AcP) is characterised by a conformational stability higher than wild type Sso AcP (figure 3.5A and (Plakoutsi et al. 2006)).  $\Delta G_{U-F}^{H2O}$  values are equal to  $48.0 \pm 2.0 \text{ KJ mol}^{-1}$  and  $52.0 \pm 1.7 \text{ KJ mol}^{-1}$  for wild type (Bemporad et al. 2004) and  $\Delta$ N11 (Plakoutsi et al. 2006) Sso AcP, respectively. Since the latter protein variant is not able to aggregate in conditions that promote aggrega-

tion of wild type Sso AcP, it is possible that the role played by the unstructured segment in the aggregation process is related to its destabilising effect on the protein. To verify this hypothesis, we have studied Sso AcP aggregation both in conditions that decrease the conformational stability of  $\Delta$ N11 protein variant and in conditions that increase the conformational stability of wild type protein.

In order to destabilise  $\Delta$ N11 Sso AcP we have applied a protein engineering strategy (figure 3.5). In particular, we have introduced an isoleucine to valine single point mutation in the hydrophobic core of the molecule to eliminate a methyl group from Ile72. We will refer to the obtained mutant as I72V- $\Delta$ N11 Sso AcP. The equilibrium unfolding curve carried out in 50 mM acetate buffer at pH 5.5 and 37 °C on this mutant shows that the elimination of the methyl group results in a significant destabilisation of the protein (figure 3.5A). The  $\Delta G_{U-F}^{H_2O}$  value is equal to  $43.1 \pm 1.6$  KJ mol<sup>-1</sup>. Thus, I72V- $\Delta$ N11 Sso AcP is a protein variant without unstructured segment with a conformational stability lower than wild type protein. The change in ThT fluorescence over time after dilution of this mutant in the buffer that induces aggregation of wild type protein (see above) is shown in figure 3.5B. This mutant induces the same change in fluorescence as wild type protein does. However, it is important to observe that the process is three orders of magnitude slower than the one observed for wild type protein. The aggregation rate constant, determined by best fit of experimental data to equation 3.3, is equal to  $(8.1 \pm 0.8) \cdot 10^{-6}$  s<sup>-1</sup>. The species populated at the plateau of the kinetic experiment binds to CR dye (data not shown).

Stabilisation of wild type protein has been achieved taking advantage of the catalytic properties of Sso AcP. This protein is an enzyme able to hydrolyse phosphoanhydridic bonds of acylphosphates (Corazza et al. 2006). All proteins belonging to acylphosphatase superfamily follow standard Michaelis-Menten kinetic theory and phosphate ion is a well known competitive inhibitor of their activity (Stefani et al. 1997). In the presence of phosphate the amount of native protein will increase and the resulting stabilisation  $\Delta\Delta G_{U-F}^{P_i}$  can be calculated as follows:

$$\Delta\Delta G_{U-F}^{P_i} = RT \ln \left( 1 + \frac{C_{P_i}}{K_i} \right) \quad (3.1)$$

where  $C_{P_i}$  is the phosphate concentration,  $K_i$  is the affinity constant of phosphate,  $R$  is the ideal gas constant and  $T$  is the temperature. Derivation of this equation is shown in appendix A (see section A.1.2). Equation 3.1 allows the phosphate concentration to be calculated that induces the desired stabilisation on Sso AcP. We have determined the affinity constant of Sso AcP for phosphate ion in the aggregation promoting conditions. In these conditions  $K_i$  is equal to  $1.12 \pm 0.1$  mM (data not shown). Then we have followed ThT fluorescence in the presence of 34  $\mu$ M wild type protein in 44.1 mM acetate and 4.8 mM phosphate buffer at pH 5.5, 20% (v/v) TFE and 25 °C (figure 3.5B). In these conditions wild type protein has the same conformational stability as  $\Delta$ N11 Sso AcP, as determined by equation 3.1, while the overall ionic strength does not vary. The fluorescence of the dye increases to reach a plateau in a single exponential phase. Aggregation rate constant, determined by best fit of experimental data to equation 3.3, is equal to  $(2.6 \pm 0.3) \cdot 10^{-6}$  s<sup>-1</sup>. The species populated at the plateau of the kinetic experiment binds to CR dye (data not shown).

These experiments show that phosphate slows down amyloid-like aggregation of wild type Sso AcP. This is probably due to the ability of this ion to bind to the catalytic site and to decrease the conformational fluctuations (Soldi et al. 2006b). Moreover, de-

stabilising  $\Delta$ N11 Sso AcP induces its aggregation. Nevertheless, aggregation rate constant determined for the stabilized wild type protein is significantly higher than the one obtained for destabilised  $\Delta$ N11 Sso AcP. These observations rule out models for the aggregation mechanism of the protein in which the fundamental force that leads to formation of early aggregates is the destabilisation induced by the N-terminal unstructured segment on the globular part of Sso AcP (see models B.2.10 and B.2.13 in appendix B).

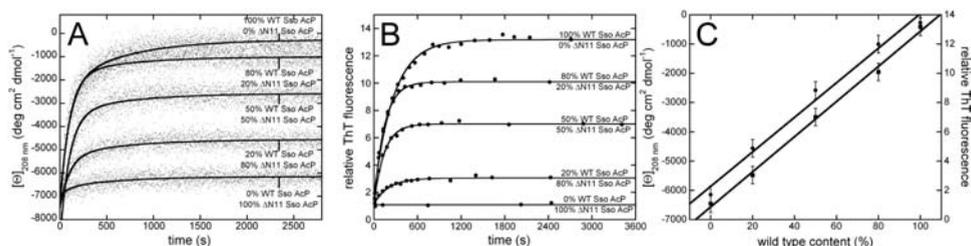


Figure 3.6: Aggregation of wild type Sso AcP in the presence of  $\Delta$ N11 Sso AcP. (A) Aggregation of 34  $\mu$ M Sso AcP in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25 °C monitored by means of mean residue ellipticity at 208 nm. Although the total amount of protein is constant, the traces show different relative amounts of protein with N-terminal segment (wild type Sso AcP) and without N-terminal segment ( $\Delta$ N11 Sso AcP). Continuous lines represent best fits of experimental data to equation 3.2. (B) Aggregation of 34  $\mu$ M Sso AcP in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25 °C monitored by means of ThT fluorescence. The signal is shown relative to ThT fluorescence in the presence of the blank solution, i. e. 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25 °C. Although the total amount of protein is constant, the traces show different relative amounts of protein with N-terminal segment (wild type Sso AcP) and without N-terminal segment ( $\Delta$ N11 Sso AcP). Continuous lines represent best fits of experimental data to equation 3.3. (C) Plateau circular dichroism (●) and ThT fluorescence (■) versus wild type relative content in the experiments reported in panels (A) and (B). Continuous lines represent best linear fits of experimental data.

### 3.2.3 A specific inter-molecular interaction between N-terminal segment and globular Sso AcP leads to the formation of early aggregates

We showed in our previous experiments that neither the globular part nor the N-terminal segment of Sso AcP are able to aggregate when separated from the remaining of the molecule (Plakoutsi et al. 2006). However, whether or not the unstructured segment gives rise to specific interactions in the early aggregates and the intermolecular or intra-molecular nature of these interactions is still unclear. To get insight into the aggregation mechanism of Sso AcP we have studied the behaviour, in conditions that promote aggregation, of solutions containing different relative amounts of wild type Sso AcP,  $\Delta$ N11 Sso AcP and four short peptides.

In a first set of experiments we have checked if the aggregation nucleus whose formation is led by the N-terminal segment is able to sequester the protein variant lacking this portion (figure 3.6). That is, we have investigated the effect of mixing wild type Sso AcP with  $\Delta$ N11 Sso AcP on the two phases of the process. In this experiment the Sso AcP concentration remains equal to 34  $\mu$ M in all tested samples. However, the relative amount of the two protein variants varies ranging from 0% to 100% of wild

type. These two latter conditions represent control experiments. Figure 3.6A shows aggregation kinetics followed by mean residue ellipticity at 208 nm. Mixing different relative amounts of the protein variant carrying the N-terminal segment with the one lacking this segment does not affect the rate of formation of protofibrils, as determined by best fits of experimental data to equation 3.2 (data not shown). There is instead a significant correlation between aggregation rate constant  $k_1$  and wild type content. Most importantly, the secondary structure content present at the end of the kinetic experiment, as inferred from the value reached at the plateau of the trace, increases as the relative amount of  $\Delta N11$  Sso AcP in the sample increases. Similar results have been obtained following the process by ThT fluorescence (figure 3.6B). The aggregation rate constant  $k_2$ , determined by best fits of experimental data to equation 3.3, does not show significant changes as the relative amounts of the protein variants change. However, an increase of the wild type content in the sample results in a higher fluorescence at the plateau of the experiment (figure 3.6B). Plateau values measured in ThT and mean residue ellipticity kinetics have been plotted versus wild type relative content in figure 3.6C. The figure shows a significant linear correlation between these values and the content of protein carrying the N-terminal segment. This confirms that only wild type Sso AcP is able to aggregate and that this variant is not able to hijack the protein lacking the N-terminal segment into initial aggregates. Thus, presence of this segment is required not only in nucleation of the process, but also in the elongation of early aggregates. Moreover, these observations rule out models based on the idea that the N-terminal segment leads to the formation of large early aggregates in which globular portion of the molecules interact as well (see model B.2.3 and B.2.4 in appendix B).

To get further information into the role of the N-terminal segment we have studied the effect on Sso AcP aggregation of the four following molecules:

1. An 11 residue peptide bearing the sequence of the Sso AcP N-terminal segment, MKKWS DTEVFE. We will refer to this peptide as tail-11.
2. A 14 residue peptide bearing the sequence of the Sso AcP N-terminal segment and the initial three residues of the first  $\beta$ -strand, MKKWS DTEVFEMLK. We will refer to this molecule as tail-14.
3. An 11 residue peptide bearing the same amino-acid content as tail-11 but a different sequence, TMFKDWESEKV. We will refer to this molecule as scrambled peptide.
4. An 11 residue peptide designed to be soluble and charged at pH 5.5. The sequence of the peptide is KSRAHNGKSAQ. We will refer to this molecule as control peptide.

The peptide tail-14 has been used to check the possibility that in the aggregation promoting conditions a partial unfolding occurs that exposes to the solvent a portion of the molecule longer than the segment. The scrambled peptide has been used to check if the possible effects induced by tail-11 are due to its primary sequence or to its overall physico-chemical properties. Finally, the control peptide has been used to control possible effects on the aggregation due to molecular crowding.

We have investigated in detail possible effects induced by tail-11 and tail-14 peptides on the behaviour shown by  $\Delta N11$  Sso AcP. In fact, this protein variant does not undergo any conformational modification in conditions that promote aggregation of wild type protein. The hydrodynamic diameter of the molecule shows a small decrease

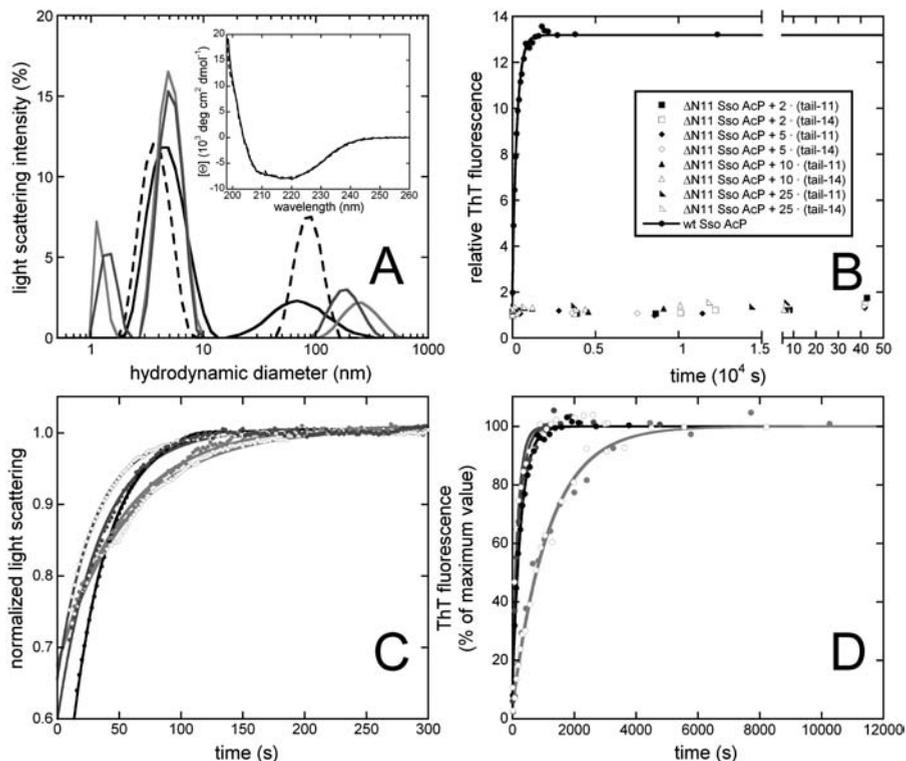


Figure 3.7: Effect of peptides on the aggregation of Sso AcP. (A) Dynamic light scattering spectra of 34  $\mu\text{M}$   $\Delta$ N11 Sso AcP in 10mM TRIS buffer at pH 8 and 25  $^\circ\text{C}$  (black continuous line), in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25  $^\circ\text{C}$  (black dashed line), in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25  $^\circ\text{C}$  with a 4 fold molar excess of tail-11 (red line) and scrambled (blue line) peptides. The inset shows circular dichroism spectra of 34  $\mu\text{M}$   $\Delta$ N11 Sso AcP in 10 mM TRIS buffer at pH 8 and 25  $^\circ\text{C}$  (continuous line) and in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25  $^\circ\text{C}$  (dashed line). (B) ThT fluorescence, reported as fold-increase relative to the blank (50 mM acetate buffer at pH 5.5, 20% (v/v) TFE) during incubation of  $\Delta$ N11 Sso AcP in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25  $^\circ\text{C}$  in the presence of different molar excesses of tail-11 (filled symbols) and tail-14 (empty symbols) peptides. The trace for wild type Sso AcP in the same conditions ( $\bullet$ ) is shown for comparison. Continuous line represents best fit of experimental data to equation 3.3. (C) First phase of aggregation of 34  $\mu\text{M}$  wild type Sso AcP in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25  $^\circ\text{C}$  in the presence of 4 fold molar excesses of peptides. Static light scattering signal recorded at 208 nm normalised to the maximum value is shown versus time in this panel. Traces are shown for Sso AcP alone ( $\bullet$ ), Sso AcP and tail-11 peptide ( $\circ$ ), Sso AcP and tail-14 peptide ( $\bullet$ ), Sso AcP and scrambled peptide ( $\bullet$ ), Sso AcP and soluble peptide ( $\circ$ ). Continuous lines represent best fits of experimental data to equation 3.4. (D) Second phase of aggregation monitored by ThT fluorescence normalised to the plateau value, of 34  $\mu\text{M}$  wild type Sso AcP in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE and 25  $^\circ\text{C}$  in the presence of 10 fold molar excesses of peptides. Traces are shown for Sso AcP alone ( $\bullet$ ), Sso AcP and tail-11 peptide ( $\circ$ ), Sso AcP and tail-14 peptide ( $\bullet$ ), Sso AcP and scrambled peptide ( $\bullet$ ), Sso AcP and soluble peptide ( $\circ$ ). Continuous lines represent best fits to equation 3.3 of experimental data.

when a sample containing native protein is diluted into a 50 mM acetate buffer at pH 5.5, TFE 20% (v/v) and 25 °C. This is probably due to a change in solvent conditions and suggests that no partial unfolding or decrease in compactness (figure 3.7A) occurs. The secondary structure content of the protein, as assessed by far-UV CD, does not significantly change (inset to figure 3.7A). Thus, we have checked the effect of tail-11 and scrambled peptide on the hydrodynamic diameter of the protein.

The results show that incubating  $\Delta$ N11 Sso AcP in the presence of an equimolar amount of both peptides does not induce increase in protein dimensions compatible with dimeric or oligomeric species (figure 3.7A). The observed slight increase relative to the control can be due to imprecision in data fitting induced by the presence of the peptides or to the interaction between peptide and  $\Delta$ N11 molecule. Since  $\Delta$ N11 Sso AcP is stable in the aggregation conditions, we have checked whether increasing the relative amount of peptides is able to induce aggregation of the protein. In particular, we have followed the ThT fluorescence induced by samples containing 34  $\mu$ M  $\Delta$ N11 Sso AcP in 50 mM acetate buffer at pH 5.5, 20% (v/v) TFE, 25 °C and in the presence of tail-11 and tail-14 molar excesses ranging from 1 fold to 25 fold. The results show that, while the aggregation of wild type protein reaches a plateau in about one hour, there are no significant changes in ThT fluorescence when incubating  $\Delta$ N11 Sso AcP in the presence of these peptide excesses for a week (figure 3.7B). These observations allow to rule out aggregation mechanisms based on a bridging effect of the N-terminal segment between two Sso AcP molecules (models B.2.2, B.2.3, and B.2.9 in appendix B) and models based on intra-molecular interactions between unstructured segment and globular part of Sso AcP that give rise to an aggregation prone state (models B.2.5, B.2.6, B.2.11 and B.2.12 in appendix B). Finally, the absence of oligomers when  $\Delta$ N11 Sso AcP is incubated in aggregation conditions rules out models in which two molecules interact through their globular parts (models B.2.7 and B.2.8 in appendix B).

The effect of peptides has been also studied on the aggregation of wild type Sso AcP. We have followed both first phase of aggregation via static light scattering signal at 208 nm and second phase via ThT fluorescence (see section 3.4). In the presence of a fourfold molar excess of tail-11 and tail-14 formation of early aggregates is significantly slower (figure 3.7C). Aggregation rate constants  $k_1$ , determined by best fits of experimental data to equation 3.4, are equal to  $(3.4 \pm 0.3) \cdot 10^{-2} \text{ s}^{-1}$ ,  $(1.8 \pm 0.2) \cdot 10^{-2} \text{ s}^{-1}$  and  $(1.7 \pm 0.2) \cdot 10^{-2} \text{ s}^{-1}$  for wild type alone, wild type in the presence of a fourfold molar excess of tail-11 and wild type in the presence of a fourfold molar excess of tail-14, respectively. Most importantly, the scrambled peptide is not able to induce the same effect on the process (figure 3.7C). Aggregation rate constant  $k_1$  for wild type protein in the presence of a fourfold molar excess of scrambled peptide is  $(3.0 \pm 0.3) \cdot 10^{-2} \text{ s}^{-1}$ . This value is, within the experimental error, compatible with aggregation rate constant of wild type protein alone. Finally, we have also checked the effect of molecular crowding on formation of early aggregates. Aggregation rate constant  $k_1$  for wild type protein in the presence of a fourfold molar excess of control peptide is  $(3.4 \pm 0.3) \cdot 10^{-2} \text{ s}^{-1}$ , showing no effects on the process. Similar conclusions can be made for the second phase of aggregation, which has been studied by means of ThT kinetics (figure 3.7D). Adding a 10 molar excess of tail-11 or tail-14 peptides significantly slows down formation of protofibrils. Aggregation rate constants  $k_2$ , determined by best fits of experimental to data equation 3.3, are equal to  $(3.7 \pm 0.3) \cdot 10^{-3} \text{ s}^{-1}$  and  $(8.5 \pm 0.8) \cdot 10^{-4} \text{ s}^{-1}$  for wild type alone and wild type in the presence of a 10 fold molar excess of either peptide, respec-

tively. In fact, no significant differences have been found between the effects of tail-11 and tail-14. By contrast, the scrambled peptide does not show a similar effect on the aggregation of the protein. Aggregation rate constant  $k_2$  for wild type Sso AcP in the presence of a 10 fold molar excess of scrambled peptide is  $(4.7 \pm 0.4) \cdot 10^{-3} \text{ s}^{-1}$ , showing a light, albeit significant, acceleration of the process.

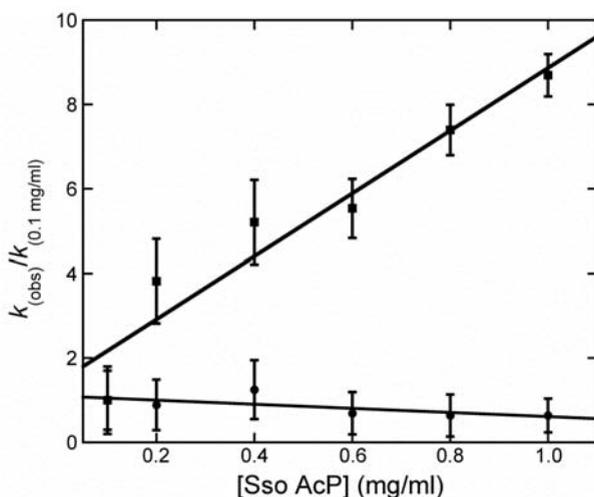


Figure 3.8: Dependence of  $k_1$  and  $k_2$  on Sso AcP concentration. In this figure the ratio is shown between the observed first phase ( $\blacksquare$ ) and second phase ( $\bullet$ ) aggregation rate constants at a given protein concentration and the value measured at  $0.1 \text{ mg ml}^{-1}$ . Continuous lines represent best linear fits of experimental data.

However, this effect is probably due to the crowding in the solution as the control peptide shows a similar effect. Aggregation rate constant  $k_2$  for wild type Sso AcP in the presence of a 10 fold molar excess of control peptide is  $(5.8 \pm 0.5) \cdot 10^{-3} \text{ s}^{-1}$ .

Taken together, these results show that tail-11 and tail-14 peptides compete with the N-terminal Sso AcP segment for binding to a specific sequence as aggregation is affected by the presence of either molecule and that this interaction leads to amyloid aggregation. The interaction seems to be an inter-molecular one as peptides induce a deceleration instead of an acceleration of the process.

### 3.2.4 Formation of initial aggregates depends on protein concentration

The analysis of the experiments reported in figure 3.6A suggests that formation of early aggregates depends on protein concentration while formation of protofibrils does not. To get further insight into the aggregation process, we have studied the dependence of the two aggregation rate constants  $k_1$  and  $k_2$  on protein concentration. In this experiment only wild type protein is present in the samples, albeit in different concentrations. The process has been followed by means of mean residue ellipticity at 208 nm (data not shown) and the recorded traces have been analysed with equation 3.2. Figure 3.8 shows the ratio between measured aggregation rate constants at a given protein concentration,  $k_1$  (squares) and  $k_2$  (circles), and the corresponding value measured at the lowest investigated concentration,  $0.1 \text{ mg ml}^{-1}$ . The results confirm that, as we previously showed

(Plakoutsi et al. 2005), for protein concentrations ranging from 0.1 to 1 mg ml<sup>-1</sup>  $k_2$  values do not significantly change. However, a different trend has been observed for formation of early aggregates. Aggregation rate constant  $k_1$  linearly increases as Sso AcP concentration increases. This experiment suggests that formation of early aggregates is an inter-molecular process whose rate depends on the concentration of free molecules in the solvent. By contrast, the second phase of the process, the formation of protofibrils, is likely to be an intra-molecular process in which the native-like fold of the protein converts into a ThT binding state characterised by increase in  $\beta$ -structure.

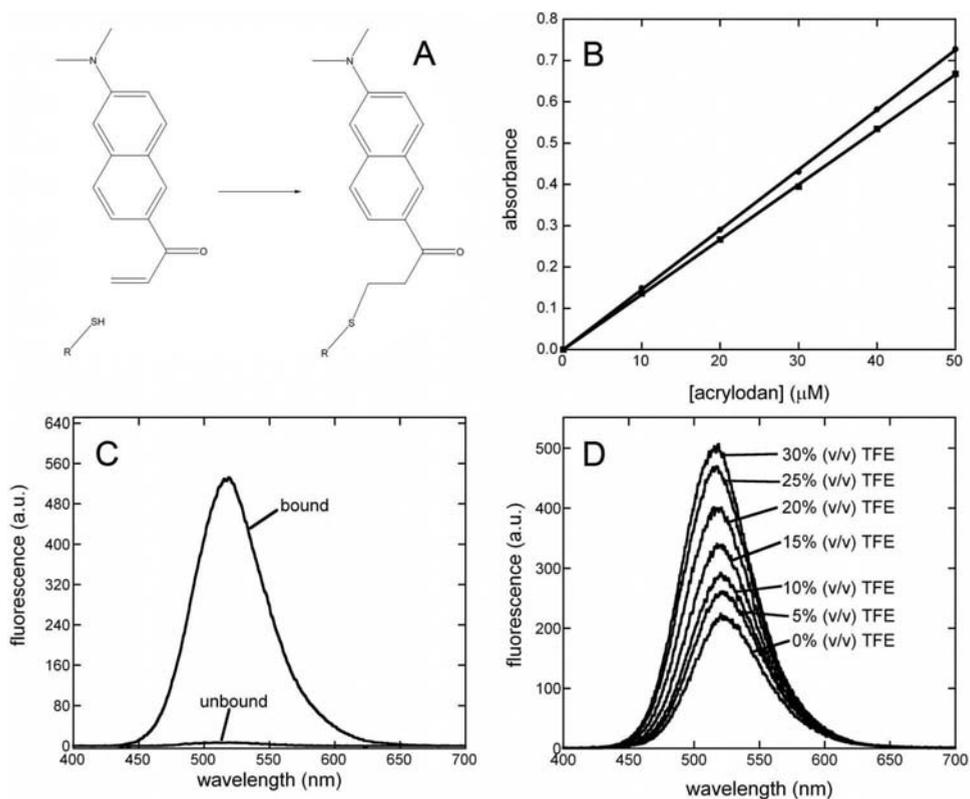


Figure 3.9: Properties of 6-acryloyl-2-dimethylaminonaphthalene (acrylodan). (A) Structure of acrylodan and its reaction with sulphhydryl group of cysteine. (B) Determination of the molar extinction coefficient of acrylodan. The figure shows dependence of absorbance at 360 nm (●) and 390 nm (■) on acrylodan concentration. Measurements have been performed in 50 mM phosphate buffer at pH 7, 25 °C. (C) Fluorescence spectra of 34  $\mu$ M acrylodan bound and unbound to glutathione in 50 mM acetate buffer at pH 5.5 and 25 °C. (D) Fluorescence spectra of 34  $\mu$ M acrylodan in 50 mM acetate buffer at pH 5.5 and different TFE concentrations, ranging from 0% (v/v) to 30% (v/v).

## 3.2.5 Experiments with acrylodan

The experiments presented in the previous sections give kinetic results that allow the role of the tail in the aggregation of Sso AcP to be clarified. Moreover, they can be used to propose a possible model for the aggregation mechanism of this protein (see section 3.3). Nevertheless, these experiments do not give any information about the most structured regions in the early aggregates and amyloid-like protofibrils.

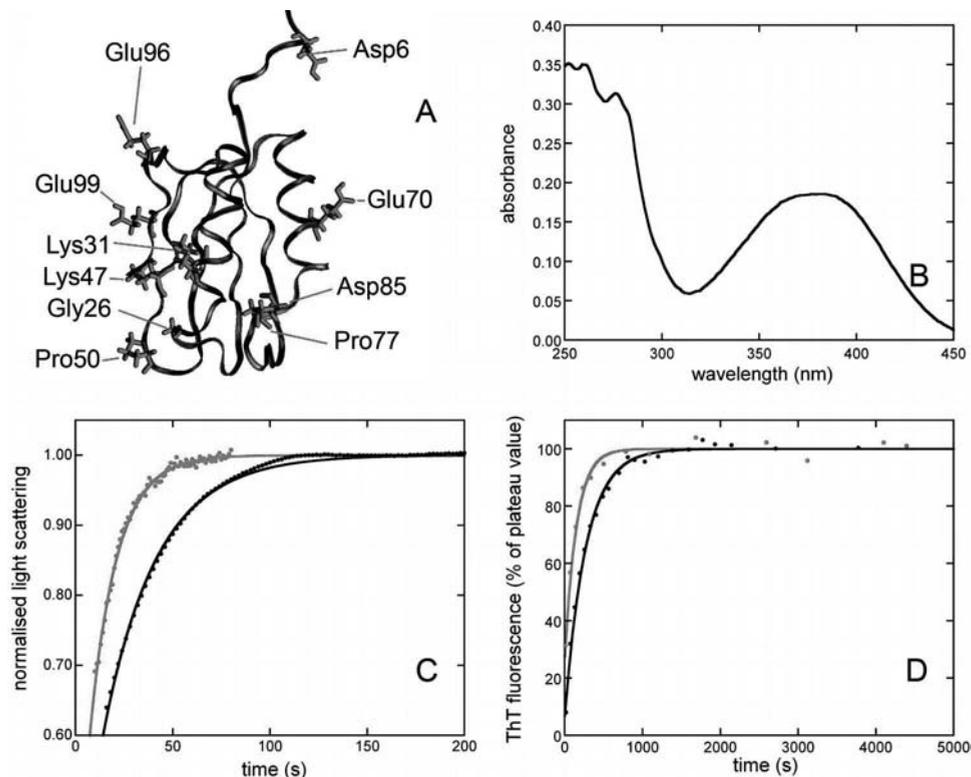


Figure 3.10: Preliminary aggregation studies on acrylodan labelled Sso AcP protein variants. (A) ribbon representation of Sso AcP to show the labelled residues. (B) Absorbance spectrum of a cysteine variant of Sso AcP labelled with acrylodan. E70C variant is shown. The degree of labelling, calculated as reported in section 3.4, is equal to  $1.10 \pm 0.05$ . (C) and (D) Preliminary aggregation studies on labelled proteins. Data for E70C (●) and wild type (●) are shown. First phase (C) has been monitored with normalised light scattering; second phase (D) has been monitored with ThT fluorescence.

To perform a structural analysis on the species transiently populated along the aggregation pathway of Sso AcP, we performed fluorescence studies using 6-acryloyl-2-dimethylaminonaphthalene (acrylodan; figure 3.9A). This molecule emits fluorescence between 400 and 600 nm when excited at 390 nm and has peculiar properties. First, it is able to react with sulphhydryl groups of cysteine residues while it does not react with side chains of other amino acids. This is particularly important in the view that wild type Sso AcP does not possess any cysteine and therefore that different amino acids can

be substituted with a cysteine through a single point mutation. This feature allows labelling with acrylodan in specific positions of the sequence to be performed without risk of non-specific binding. Thus, the fluorescence of the bound probe can be studied during aggregation of the protein. The presence of the probe can be verified by absorbance measurements as the probe has an extinction coefficient at 360 nm equal to  $14500 \pm 200 \text{ M}^{-1} \text{ cm}^{-1}$  (see figure 3.9B). A second important property of this molecule is that fluorescence of acrylodan bound to cysteine residues is more than 100-fold higher than fluorescence of unbound acrylodan (figure 3.9C). This is fundamental in order to avoid artefacts in the measurements due to the presence of unbound probe that has not been removed from the sample.

Table 3.1: Aggregation kinetic data for a set of Sso AcP cysteine variants labelled with acrylodan. Aggregation rate constants  $k_1$  have been determined by best fits of experimental data to equation 3.4. Aggregation rate constants  $k_2$  have been determined by best fits of experimental data to equation 3.3.

protein variant	position	$k_1 \text{ (s}^{-1}\text{)}$	$k_2 \text{ (s}^{-1}\text{)}$
WT	-	$(3.4 \pm 0.4) \cdot 10^{-2}$	$(3.7 \pm 0.4) \cdot 10^{-3}$
D6C	tail	$(3.6 \pm 0.4) \cdot 10^{-2}$	$(2.2 \pm 0.2) \cdot 10^{-3}$
G26C	loop $\alpha 1$ - $\beta 1$	$(2.9 \pm 0.6) \cdot 10^{-1}$	$(3.2 \pm 0.4) \cdot 10^{-3}$
K31C	$\alpha 1$	$(7.4 \pm 0.7) \cdot 10^{-2}$	$(2.4 \pm 0.2) \cdot 10^{-2}$
K47C	$\beta 2$	$(1.1 \pm 0.2) \cdot 10^{-1}$	$(2.0 \pm 0.3) \cdot 10^{-2}$
P50C	loop $\beta 2$ - $\beta 3$	$(5.6 \pm 0.6) \cdot 10^{-2}$	$(1.1 \pm 0.2) \cdot 10^{-3}$
E70C	$\alpha 2$	$(2.2 \pm 0.3) \cdot 10^{-2}$	$(6.7 \pm 0.7) \cdot 10^{-3}$
P77C	loop $\alpha 2$ - $\beta 4$	$(2.2 \pm 0.3) \cdot 10^{-1}$	$(2.0 \pm 0.3) \cdot 10^{-2}$
D85C	$\beta 4$	$(4.0 \pm 1.0) \cdot 10^{-1}$	$(1.2 \pm 0.2) \cdot 10^{-2}$
E96C	loop $\beta 4$ - $\beta 5$	$(8.4 \pm 0.8) \cdot 10^{-2}$	$(1.4 \pm 0.2) \cdot 10^{-3}$
E99C	$\beta 5$	$(1.1 \pm 0.2) \cdot 10^{-1}$	$(2.2 \pm 0.2) \cdot 10^{-2}$

The fluorescence of acrylodan is sensitive to the local environment. In particular, when the molecule is in a polar environment it emits fluorescence with a peak of about 520 nm. When the probe is instead in an apolar environment, its fluorescence shows significant blue shift (with a peak that reaches 470 nm) and increase in intensity (Krishnan and Lindquist 2005; Sun et al. 2007). This allows the local environment of the regions labelled with the probe to be investigated. In particular, in the case of amyloid aggregation studies, this feature allows the region buried in the different phases of aggregation to be found. However, it is important to observe that the fluorescence of acrylodan is not only sensitive on the solvent exposure of the labelled moiety. The signal depends also on the physico-chemical properties of the solvent. Figure 3.9D shows the dependence of acrylodan fluorescence spectra on TFE concentration. The graph shows that increasing the concentration of this cosolvent, which possesses two aliphatic carbons, induces an increase and a blue shift of the probe fluorescence. This behaviour must be taken into account during analysis of the data as TFE is the solvent used to induce aggregation of Sso AcP.

We have produced 10 variants of Sso AcP carrying a cysteine in different positions (figure 3.10A, table 3.1). Residues to be mutated have been chosen to have the following properties: (1) Mutated residues are exposed to the solvent. Solvent accessibility is important to reach the equilibrium of the labelling reaction in which the entire protein population is labelled. Moreover, the aggregation of Sso AcP starts from an ensemble

of native-like conformations and the regions of the sequence that initiate the process are likely to be solvent exposed in the native state. (2) Mutated residues span different secondary structure elements of Sso AcP. Mutations have been chosen in order to investigate the two edge  $\beta$ -strands, the two  $\alpha$ -helices and the N-terminal unstructured segment. (3) Mutated residues are positioned in loops. Four of the 10 investigated positions are loop residues. Labelling reaction has been carried out as reported in section 3.4.10. Presence and purity of the labelled protein has been verified by MALDI-TOF mass spectrometry; labelling degree has been calculated with equation 3.5. In all cases the degree of labelling has been higher than 98% (figure 3.10B). Purity of samples is important to avoid the presence of two subpopulations of the protein in the sample that can show two different aggregation mechanisms.

Table 3.2: Peaks of acrylodan fluorescence spectra of cysteine mutants of Sso AcP. Data for native proteins have been acquired in 10mMTRIS buffer at pH 8.0 and 25 °C. Data for early aggregates and protofibrils have been acquired at the beginning and at the plateau of the ThT kinetics, respectively. Conditions have been 20% (v/v) TFE and 50 mM acetate buffer at pH 5.5, 25 °C.

protein variant	Peak in native protein (nm)	Peak in early aggregates (nm)	Peak in protofibrils (nm)
D6C	504.2 ± 0.5	503.5 ± 0.5	496.5 ± 0.5
G26C	508.4 ± 0.5	495.6 ± 0.5	495.3 ± 0.5
K31C	513.9 ± 0.5	496.5 ± 0.5	494.0 ± 0.5
K47C	514.8 ± 0.5	498.1 ± 0.5	492.2 ± 0.5
P50C	520.3 ± 0.5	505.0 ± 0.5	496.2 ± 0.5
E70C	517.1 ± 0.5	503.9 ± 0.5	501.2 ± 0.5
P77C	503.7 ± 0.5	499.1 ± 0.5	497.2 ± 0.5
D85C	519.5 ± 0.5	501.3 ± 0.5	500.3 ± 0.5
E96C	519.3 ± 0.5	503.1 ± 0.5	496.9 ± 0.5
E99C	518.2 ± 0.5	499.9 ± 0.5	497.7 ± 0.5

In order to verify that the presence of the probe on the protein surface does not affect the amyloid-like aggregation process, a set of preliminary experiments has been carried out on the labelled protein variants (table 3.1). In particular:

- **Static light scattering kinetics:** To monitor the first phase of the process, aggregation of 34  $\mu$ M labelled protein variants in 20% (v/v) and 50 mM acetate buffer at pH 5.5 and 25 °C has been followed by static light scattering signal at 208 nm (figure 3.10C). The results (table 3.1) show that all but one variants show a significant, in some cases remarkable, increase of the aggregation rate constant  $k_1$  determined with equation 3.4. This effect is not surprising as the presence of a large, hydrophobic moiety on the surface of the globular part of Sso AcP increases the ability to interact of two Sso AcP molecules. Moreover, it is possible that the mutation introduced has a destabilising effect on native Sso AcP and it has been observed that the extent of destabilisation is correlated with the aggregation rate of the protein (Plakoutsi et al. 2006).
- **ThT fluorescence kinetics:** To monitor the second phase of the process, aggregation of 34  $\mu$ M labelled protein variants in 20% (v/v) and 50 mM acetate buffer at pH 5.5 and 25 °C has been followed by ThT fluorescence at 485 nm (figure 3.10D). As verified for the first phase, in most cases kinetic rate constants  $k_2$  determined by

best fits to equation 3.3 show that the presence of the probe on the protein surface speeds up the process (table 3.1). However, this increase in aggregation rate  $k_2$  is likely to be an effect of the increase of  $k_1$  constant, as we have shown that the rate of the second phase does not depend on peptide concentration (see section 3.2.4). Importantly, all protein variants show the same increase in ThT fluorescence that has been observed for wild type protein, suggesting that the species populated at the plateau of the ThT kinetic experiment is not affected by the presence of acrylodan.

- **CR staining:** The species populated at the plateau of the ThT kinetics have been stained with CR. In all cases presence of a peak at 550 nm can be observed (data not shown). Similar behaviour was shown for wild type protein (Plakoutsi et al. 2004), suggesting that protein variants labelled with acrylodan give rise to protofibrils with regular structure.

Taken together, these results show that the presence of the probe on the protein surface speeds up the aggregation process but does not affect the aggregation mechanism that leads to the formation of protofibrils.

Once investigated the aggregation properties of the cysteine variants produced and labelled, we have studied fluorescence of the probe during aggregation (table 3.2). In particular, the following spectra have been acquired for each mutant:

- **native protein:** A spectrum between 400 nm and 700 nm with an excitation wavelength of 390 nm has been acquired for each mutant. Conditions are 34  $\mu$ M protein in 10 mM TRIS at pH 8.0 and 25 °C (table 3.2). In these conditions the protein variants are enzymatically active (data not shown). Thus, this spectrum is important to verify the fluorescence of the probe in conditions in which the protein is native and monomeric.
- **early aggregates:** For each mutant a set of spectra has been acquired over time in conditions that induce aggregation of Sso AcP, that is 34  $\mu$ M protein in 20% (v/v) TFE and 50 mM acetate buffer at pH 5.5, 25 °C. Fluorescence has been recorded between 450 nm and 550 nm, with an excitation wavelength of 390 nm (table 3.2). Since the fluorescence at every wavelength has shown exponential behaviour, the value at the beginning of the experiment (corresponding to the early aggregates) has been extrapolated using equation 3.3.
- **protofibrils:** Finally, the spectrum of protofibrils has been obtained using the same set of spectra used for early aggregates. For each wavelength, the plateau value of the probe fluorescence has been measured in order to obtain the plateau spectrum (table 3.2).

Native proteins show a fluorescence peak similar to the peak of acrylodan bound to glutathione in aggregation promoting conditions ( $518.0 \pm 0.5$  nm). In these conditions acrylodan is completely exposed to the solvent. This confirms that in the native states, after labelling, the probe is still on the protein surface. In the early aggregates we observe, as expected, a blue shift in the fluorescence peak. This shift ranges from 13 nm for G26C to 20 nm for P50C (table 3.2). In protofibrils, further blue shift of the fluorescence peak can be observed (table 3.2). Nevertheless, it must be noticed that the observed changes are not position dependent. All investigated regions show similar behaviours. This clearly suggests that the structural reorganisation that the globular part of Sso AcP undergoes during aggregation is a general one. The first phase of aggregation consists of a general collapse of the structure in which the entire globular Sso AcP is sequestered from the solvent. Moreover, the observed values are not remarkable, albeit significant. In the case of Sup35p, residues labelled with acrylodan and buried in

protofilaments show fluorescence peak equal to about 470 nm (Krishnan and Lindquist 2005). In the case of prion PrP all labelled residues show fluorescence peaks in fibrils that range from 460 nm to 470 nm (Sun et al. 2007). This is compatible with the fact that in the two mentioned cases experiments were performed on mature fibrils, while in the case of Sso AcP we have worked on prefibrillar aggregates. Finally, the increase in fluorescence during aggregation is not remarkable. This is probably due to the presence of water molecules in the aggregates, which act as quencher of fluorescence and thus decrease its quantum yield. During the second phase, part of these molecules is expelled from aggregates and the fluorescence increases. However, it is also possible that the light is scattered by the aggregates that form clusters and that the probe in the aggregates is excited to a lower extent than in the native monomeric state.

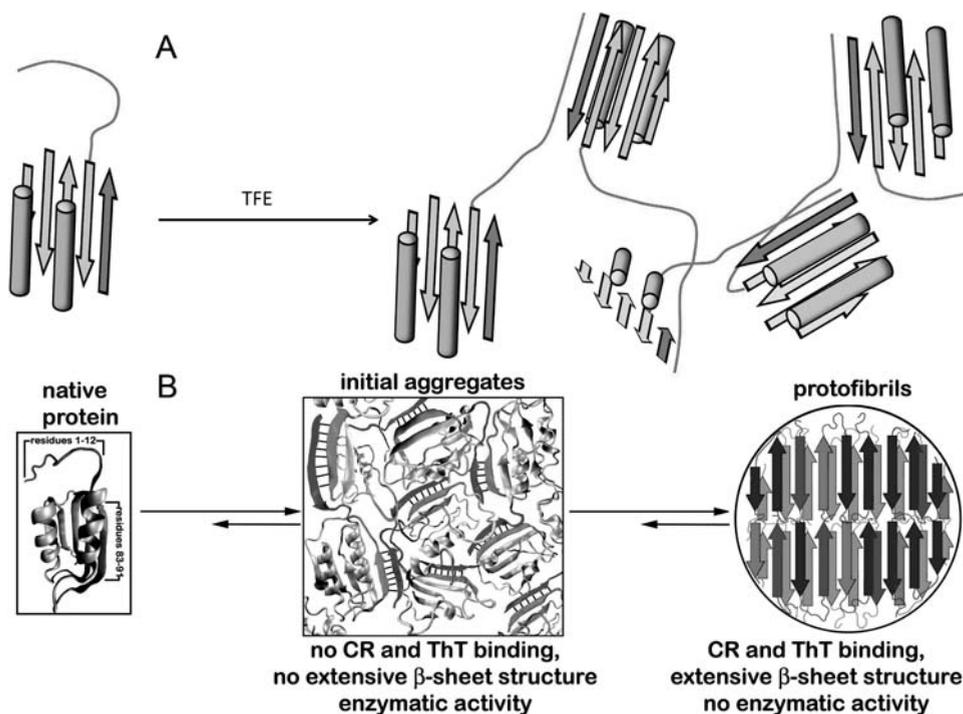


Figure 3.11: A possible model for aggregation of Sso AcP. (A) in the first phase the unstructured N-terminal segment interacts with a region of the globular part of Sso AcP (for instance the fourth  $\beta$ -strand is shown) and this interaction leads to the formation of the early aggregates. Fourth  $\beta$ -strand and unstructured segment are depicted in red. (B) The early aggregates are characterised by a global collapse of the structure and all the regions of the sequence retain a native-like conformation. In the second phase an intra-molecular reorganisation leads to the formation of protofibrillar aggregates that bind ThT, CR and possess extensive  $\beta$ -sheet structure. The unstructured N-terminal segment and the fourth  $\beta$ -strand are depicted in blue and red, respectively.

### 3.3 Conclusions: a possible aggregation mechanism for Sso AcP

In this chapter we have investigated the aggregation mechanism of Sso AcP. The results can be summarised as follows. (1) The N-terminal unstructured segment of Sso AcP is able to induce aggregation of the molecule regardless of its position (section 3.2.1). (2) The N-terminal unstructured segment of the protein is able to induce aggregation only when it is bound to the globular part of the molecule. In fact, this segment is not able to restore aggregation of  $\Delta N11$  Sso AcP, even if it is added in large excess (section 3.2.3). (3) The aggregation properties of the N-terminal segment of Sso AcP are not due to the destabilising effect that this peptide has on the molecule as a stabilised variant of the protein with N-terminal segment aggregates faster than a destabilised protein variant lacking the segment (section 3.2.2). (4) A peptide carrying the sequence of N-terminal segment slows down the aggregation of wild type Sso AcP and this feature is due to its primary sequence, as a scrambled peptide does not show the same properties (section 3.2.3). (5) The first phase of the process depends on protein concentration while the second one does not (section 3.2.4). This suggests inter-molecular recognition as a major event in the first phase and intra-molecular reorganisation as the mechanism that leads to the formation of protofibrils in the second phase. (6) During aggregation of Sso AcP the whole sequence collapses and participates to the formation of the early aggregates. Water molecules are present in the aggregates and are expelled only later on (section 3.2.5).

These results allow a possible model for the aggregation of Sso AcP to be proposed. The model (this model is not discussed in appendix B) can be summarized in three points:

1. Once diluted in aggregation promoting conditions, Sso AcP undergoes conformational modifications that lead to the formation of a native-like ensemble. It is possible that in this ensemble some regions buried in the native state are exposed to the solvent. Otherwise, it is possible that some regions undergo conformational modifications and become more flexible and prone to aggregate. According to previous experiments carried out in our lab the fourth  $\beta$ -strand plays a major role in the aggregation (Plakoutsi et al. 2006). It is possible that this region is important in the initial conformational modification event.
2. The native-like ensemble gives rise to the early aggregates. Our experiments suggest that the major event in this phase is a specific inter-molecular interaction between the unstructured N-terminal segment of one molecule of Sso AcP and the globular part of another molecule (figure 3.11A). The interaction is an inter-molecular one because a peptide corresponding to the N-terminal segment slows down the process. Moreover, if the interaction were intra-molecular, the fact that shifting the N-terminal segment does not affect the process would be in contrast with the evidence that the interaction is specifically due to the sequence of the segment. The interaction cannot be between two unstructured segments as we previously showed that the segment alone is not able to aggregate (Plakoutsi et al. 2006). Finally, the interaction is specifically due to the particular sequence of the segment. This is clearly shown by the fact that a scrambled peptide is not able to slow down the process as the peptide corresponding to the N-terminal segment does. Figure 3.11A shows an interaction between the unstructured segment and the fourth  $\beta$ -strand. Al-

though this interaction cannot be ruled out on the basis of our experiments, it must be noticed that it is possible that residues on this segment are fundamental only for the initial conformational modification that leads to native-like ensemble, and that in this state another region gets exposed to the solvent and flexible. Thus, figure 3.11A proposes only one of the possible interactions between unstructured segment and globular part. The early aggregates that form do not possess extensive  $\beta$ -structure and do not bind CR and ThT (Plakoutsi et al. 2004; Plakoutsi et al. 2005). Sso AcP molecules are instead still native-like and enzymatically active (Plakoutsi et al. 2004; Plakoutsi et al. 2005). Our experiments with acrylodan suggest that the entire globular part of Sso AcP is collapsed in their structure (figure 3.11B) and that water molecules are still present between Sso AcP molecules.

3. In the second phase of aggregation early aggregates undergo intra-molecular modifications (figure 3.11B). This is shown by the fact the kinetic rate constant of this phase does not depend on protein concentration. During this phase Sso ACP molecules loose their native-like topology (Plakoutsi et al. 2005). The aggregates loose enzymatic activity and acquire extensive  $\beta$ -sheet structure and ability to bind CR and ThT dies (Plakoutsi et al. 2004; Plakoutsi et al. 2006). Finally, our data show that part of the water molecules are expelled during this phase as acrylodan fluorescence increases during the process.

The model presented in this section for the aggregation mechanism of Sso AcP is consistent with all the experimental evidences presented in this chapter and in our previous work (Plakoutsi et al. 2004; Plakoutsi et al. 2005; Plakoutsi et al. 2006). Although the model is in contrast with the evidence that in most fibrils parallel  $\beta$ -sheets form between the same segments of different molecules (Tycko 2004; Kajava et al. 2005; Krishnan and Lindquist 2005; Chiti and Dobson 2006), an interaction between two unstructured segments is impossible in our case on the basis of our experiments and on the basis of the experiments that we previously carried out (Plakoutsi et al. 2006). The target region of the segment remains still unclear. Nevertheless, only further experiments will allow to find out, at atomic level, the regions that interact in the early aggregates and in the protofibrils formed by Sso AcP. However, these experiments are particularly important as they give detailed information for studies of aggregation starting from native states. Moreover, the data presented here are relevant for the study of protein dynamics in solution.

### 3.4 Materials and methods

#### 3.4.1 Materials

Guanidine hydrochloride, 2,2,2-trifluoroethanol, ThT and CR were purchased from Sigma. The synthetic peptides were purchased from Genscript Corporation (Piscataway, New Jersey, USA) and had amidated C-termini. The purity of samples was in all cases  $\geq 95\%$ . Benzoyl-phosphate was synthesised as previously described (Camici et al. 1976). Oligonucleotides for site directed mutagenesis were purchased from MWG (Ebersberg, Germany). 6-acryloyl-2-dimethylaminonaphthalene (acrylodan) was purchased from Molecular Probes.

### 3.4.2 Mutagenesis, protein expression and purification

Wild type Sso AcP and  $\Delta$ N11 Sso AcP were expressed and purified as previously described ((Bemporad et al. 2004; Plakoutsi et al. 2006) and section 2.4.2). The genes encoding for C-tail Sso AcP and I72V- $\Delta$ N11 Sso AcP were obtained starting from the pGEX-2T plasmid carrying the gene of  $\Delta$ N11 Sso AcP and using Quick Change Site-Directed Mutagenesis Kit from Stratagene (see section 2.4.1). In the case of C-tail Sso AcP three consecutive insertions were made to reach the final insertion of 11 residues. The desired mutations were verified by DNA sequencing. The encoded proteins were expressed and purified as wild type Sso AcP (Bemporad et al. 2004). Purity of samples was checked by SDS-PAGE and mass spectrometry. Protein concentration was calculated using extinction coefficients at 280 nm (Gill and von Hippel 1989) and Lambert-Beer law.

### 3.4.3 Far-UV Circular dichroism

Far-UV CD spectra were acquired at 25 °C with a Jasco J-810 spectropolarimeter (Tokyo, Japan) equipped with a thermostated cell holder and a quartz cell of 1 mm path length. Spectra of wild type Sso AcP,  $\Delta$ N11 Sso AcP and Sso AcP were acquired at a protein concentration equal to 34  $\mu$ M in a 10 mM TRIS buffer at pH 8.0 and 25 °C. Spectra of  $\Delta$ N11 Sso AcP were also acquired at a protein concentration equal to 34  $\mu$ M in 50 mM acetate buffer at pH 5.5 with 20% (v/v) TFE and 25 °C. In another set of experiments the change of signal at 208 nm over time was recorded in the 1mmcell using the J-810 instrument during aggregation of three types samples. (1) Samples containing different relative amounts of wild type and  $\Delta$ N11 Sso AcP. In this case the total protein concentration was 34  $\mu$ M in all cases. (2) Samples containing wild type Sso AcP in the presence of four fold molar excesses of the four peptides. Sso AcP concentration was 34  $\mu$ M. (3) Samples containing different amounts of wild type Sso AcP ranging from 0.1 to 1.0 mg ml<sup>-1</sup>. The recorded traces were analysed using the following equation:

$$[\Theta]_{208(t)} = A_1 \cdot \exp(-k_1 \cdot t) + A_2 \cdot \exp(-k_2 \cdot t) + q \quad (3.2)$$

where  $[\Theta]_{208(t)}$  represents the CD signal at 208 nm as a function of time,  $q$  represents the plateau signal,  $A_1$ ,  $A_2$ ,  $k_1$  and  $k_2$  are amplitudes and rate constants of the first and second aggregation phase, respectively. Best fits of experimental data to equation 3.2 allowed quantitative estimations of aggregation rate constants to be obtained.

### 3.4.4 Thioflavin T fluorescence

crAggregation kinetics followed by ThT fluorescence were performed as previously reported (Plakoutsi et al. 2004; Plakoutsi et al. 2005; Plakoutsi et al. 2006). The experiments were carried out in 50 mM acetate buffer at pH 5.5 with 20% (v/v) TFE and 25 °C on the following samples: (1) 34  $\mu$ M wild type Sso AcP alone; (2) 34  $\mu$ M wild type Sso AcP in the presence of 10 fold molar excesses of the four peptides; (3) 34  $\mu$ M  $\Delta$ N11 Sso AcP in the presence of the four peptides in molar excesses ranging from 1 fold to 25 fold. In another experiment ThT fluorescence was measured in the presence of wild 34  $\mu$ M type Sso AcP incubated in a 44.1 mM acetate and 4.8 mM phosphate buffer at pH 5.5, 20% (v/v) TFE and 25 °C. Two different types of plots were produced. (1) In figures 3.4 and 3.5 ThT fluorescence intensity in the absence of protein was subtracted from all the fluorescence measurements in the presence of the tested sample and

the resulting values were normalised so that the final fluorescence intensity at the endpoint of the kinetic trace was 100%. (2) In figures 3.6 and 3.7 ThT fluorescence intensity in the presence of the tested sample was directly reported as fold increase relative to signal in the presence of blank solution. The obtained plots were analysed using the following equation:

$$f(t) = A \cdot \exp(-k_2 \cdot t) + q \quad (3.3)$$

where  $f(t)$  represents the fluorescence as a function of time,  $A$  is the amplitude of the observed phase and  $k_2$  is the kinetic rate constant for the conversion of early aggregates into amyloid-like protofibrils. Best fits of experimental data to equation 3.3 gave quantitative estimations of aggregation rate constant  $k_2$ .

#### 3.4.5 Aggregation kinetics followed by static light scattering

The presence of peptides resulted in a significant increase of the absorbance of samples and thus in a reduction of the signal to noise ratio of the traces recorded with the Jasco spectropolarimeter. As a consequence, formation of early aggregates in the presence of these molecules could not be followed by means of mean residue ellipticity. Formation of early aggregates in the presence of peptides was followed with static light scattering at 208 nm in a Jasco V-630 spectrophotometer (Tokio, Japan) with a 0.1 cm cuvette. The signal was followed after dilution of wild type Sso AcP into an aggregation buffer. Final conditions were 34  $\mu\text{M}$  Sso AcP in 50 mM acetate buffer at pH 5.5 with 20% (v/v) TFE at 25 °C. In another set of experiments the signal was followed for wild type protein in the presence of a fourfold molar excess of the tested peptides. The obtained trace was normalised to the observed linear decrease in signal due to precipitation of sample. The plot was analysed with the following equation:

$$A_{208(t)} = A \cdot \exp(-k_1 \cdot t) + q \quad (3.4)$$

where  $A_{208(t)}$  is the absorbance as a function of time,  $A$  is the amplitude of the observed phase,  $k_1$  is the rate of formation of the early aggregates and  $q$  is the equilibrium value. Best fits of experimental data to equation 3.4 allowed the aggregation rate constant  $k_1$  to be measured.

#### 3.4.6 Congo red staining

Species populated at the plateau of the ThT kinetics were tested for CR binding. Aliquots of 60  $\mu\text{l}$  of each protein solution were mixed with 440  $\mu\text{l}$  of solutions containing 20  $\mu\text{M}$  CR, 5 mM phosphate buffer, 150 mM NaCl, pH 7.4. After sample equilibration, optical absorption spectra were acquired from 400 to 700 nm with the Jasco spectrophotometer. A 5 mm path length cuvette was used. Solutions without tested sample and solutions without CR were used as controls.

#### 3.4.7 Enzymatic activity measurements

Enzymatic activity of wild type and C-tail Sso AcP protein variants was measured in a continuous optical test at 283 nm using BP as a substrate (Ramponi et al. 1966) with a

Lambda 4V Perkin Elmer spectrophotometer (Wellesley, Massachusetts). Experimental conditions were 2.0  $\mu\text{g ml}^{-1}$  Sso AcP, 5.0 mM BP, 50 mM acetate buffer at pH 5.5, 37 °C. BP was freshly dissolved before enzymatic activity measurements. In another set of experiments enzymatic activity of 2.0  $\mu\text{g ml}^{-1}$   $\Delta\text{N11}$  Sso AcP was measured at 25 °C in 50 mM acetate buffer at pH 5.5, TFE with 20% (v/v) and a BP concentration ranging from 0.2 to 8 mM. The experiment was carried out both in the absence and in the presence of 1.5 mM phosphate ion. Affinity constant for phosphate was calculated using standard Michaelis-Menten theory.

#### 3.4.8 Equilibrium unfolding

28 samples of the tested protein variant (8.5  $\mu\text{M}$ ) were incubated for 1 hr in 50mM acetate at pH 5.5, 37 °C, and different concentrations of GdnHCl ranging from 0 to 7.2 M. The values of mean residue ellipticity at 222 nm ( $[\Theta]_{222}$ ) of the samples were measured with the Jasco J-810 CD. The cell length was 1 mm. The plot of  $[\Theta]_{222}$  versus GdnHCl concentration was fitted to a two-state transition equation, as described ((Santoro and Bolen 1988) and section A.1.1) to determine the free energy change upon unfolding in the absence of denaturant ( $\Delta G_{U-F}^{H_2O}$ ), the dependence of  $\Delta G_{U-F}^{H_2O}$  on GdnHCl concentration ( $m$  value), and the midpoint of unfolding ( $C_m$ ).

#### 3.4.9 Dynamic light scattering

Hydrodynamic diameter was measured for wild type Sso AcP, C-tail Sso AcP and  $\Delta\text{N11}$  Sso AcP in 10 mM TRIS buffer at pH 8.0 and 25 °C. In another set of experiments DLS measurements were performed on 34  $\mu\text{M}$   $\Delta\text{N11}$  Sso AcP in the absence and in the presence of equimolar amounts of tail-11 and tail-14 peptides. Experimental conditions were 50mM acetate buffer at pH5.5, TFE with 20% (v/v). Size distributions by light scattering intensity were acquired with a Zetasizer Nano S DLS device from Malvern Instruments (Malvern, Worcestershire, UK). Low volume 12.5  $\times$  45 mm disposable cuvettes were used. A Peltier thermostating system maintained the temperature at 25 °C. The viscosity and refractive index parameters were set for each solution. The buffer and stock protein solutions were centrifuged (16,000 $\times$ g for 5 minutes) and filtered with 0.02 mm Anotop 10 filters (Whatman, Maidstone, UK) before the measurements.

#### 3.4.10 Cysteine labelling

Ten Sso AcP protein variants were produced as reported above (see section 3.4.2). These were labelled with 6-acryloyl-2-dimethylaminonaphthalene (acrylodan). The following protocol was applied for each mutant:

1. Reaction between acrylodan and free cysteine was carried out with 300  $\mu\text{M}$  protein in a solution containing 3mM acrylodan, 600  $\mu\text{M}$  tricarboxyethylphosphine (TCEP), 10% (v/v) dimethylformamide (DMF) 50 mM phosphate at pH 7. Samples were incubated for three hours at 25 °C. The use of TCEP is recommended to avoid contemporaneously dimerisation of the protein, which carries a free cysteine on its surface, and reaction between the reducing agent and cysteine. Unlike  $\beta$ -mercaptoethanol and DTT, TCEP is able

to reduce dimerised cysteine without binding the protein. Thus, TCEP does not compete with acrylodan for labelling reaction.

2. Samples were centrifuged for 8 minutes at 16,000×g in order to separate supernatant from the pellet, that contained both precipitated probe and Sso AcP variant.
3. Unlabelled probe was removed from samples through size exclusion chromatography. Sephadex™ G-25 medium in a PD-10 desalting column (GEHealthcare) was used. Labelled protein was diluted into 10mMTRIS buffer at pH 8.0, in which Sso AcP is stable.
4. Labelling of the protein was assessed in two ways. (1) AMALDI-TOF mass spectrum was acquired for each mutant to show the presence of the labelled protein. In all cases no unlabelled protein was detected, suggesting that at the equilibrium only labelled protein was present. (2) An absorbance spectrum was acquired for each labelled protein between 250 nm and 500 nm using the Jasco V-630 spectrophotometer (see figure 3.10B). The concentration of bound acrylodan was calculated from the signal at 360 nm using Lambert-Beer law and a molar extinction coefficient  $\epsilon_{360}$  equal to 14500 M<sup>-1</sup> cm<sup>-1</sup>. The contribution of the probe at 280 nm was then subtracted by comparing the acquired spectrum with a reference spectrum of acrylodan alone. Finally, the protein concentration was calculated as reported above (section 3.4.2). The degree of labelling  $r$  was calculated according to the following equation:

$$r = \frac{[acrylodan]}{[SsoAcP]} \quad (3.5)$$

where  $[acrylodan]$  and  $[SsoAcP]$  are molar concentrations of probe (supposing that all the acrylodan molecules are bound to the protein) and Sso AcP. In all cases  $r$  resulted  $\geq 98\%$ .

5. Labelled samples were stored at -80 °C.



## Chapter 4

### Role of $\pi$ -stacking in amyloidoses

#### 4.1 Aromatic residues and amyloidoses

##### 4.1.1 Introduction

As introduced in section 1.3, a wide range of human diseases is associated with the conversion of specific peptides or proteins from their soluble state into highly organised aggregates known as amyloid fibrils (Stefani and Dobson 2003; Dobson 2004). These include Alzheimer's disease, type 2 diabetes mellitus, and several systemic amyloidoses. The fibrillar aggregates in these diseases show some typical features, such as a long and unbranched morphology, a cross- $\beta$  X-ray pattern diffraction pattern (Sunde and Blake 1997; Jimenez et al. 1999), and peculiar tinctorial properties upon binding with Congo Red (CR) and Thioflavin T (ThT) ((Klunk et al. 1989; LeVine III 1995) and section 1.3.1). While it has been widely demonstrated that under appropriate conditions many, if not all, polypeptide chains can convert into amyloid-like fibrils (Guijarro et al. 1998; Chiti et al. 1999c), it is also clear that they do so with very different propensities. Therefore, an understanding of the parameters that modulate the aggregation propensity of a polypeptide chain and of the mechanism by which it forms fibrils is fundamental to gain insight into the pathogenesis of protein deposition diseases and to better understand the process of amyloid formation of polypeptide chains more generally. A great effort has been expended in the past few years to predicting the key determinants of the aggregation propensity and the aggregation-prone regions of a given sequence. The hydrophobic content of a sequence has been suggested as a determinant of the aggregation rate of an unstructured polypeptide chain (Calamai et al. 2003; Chiti et al. 2003; DuBay et al. 2004). Other authors have shown that sequences designed to share an identical pattern of alternating polar and non-polar residues are able to promote aggregation into amyloid-like fibrils (West et al. 1999). Indeed, this particular pattern is highly prone to form  $\beta$ -sheets, the underlying type of secondary structure observed in amyloid fibrils. The role of the propensity to form secondary structure has been extensively investigated by many other investigators. It has been shown that several proteins forming amyloid structures under physiological conditions present a  $\alpha$ -helix in a segment that has rather a high propensity to form a  $\beta$ -strand according to secondary structure predictions (Kallberg et al. 2001). Various mutants of the activation domain of human procarboxypeptidase A2 (ADA2h) designed to increase the local stability of the two helical regions have been found to be less prone to form fibrils and this is due to a decreased aggregation propensity of the unfolded state (Villegas et al. 2000). It has also been shown that destabilising the  $\alpha$ -

conformer of a given sequence is not enough to start the aggregation and that only an increase of the  $\beta$ -sheet propensity can favour aggregation (Ciani et al. 2002).

Moreover, it has been demonstrated that a highly significant inverse correlation exists between the rates of aggregation in a set of protein mutants under denaturing conditions and their overall net charge, clearly suggesting the protein charge as a major determinant for amyloid formation (Chiti et al. 2002a). Accordingly, it has been demonstrated that the tetrapeptide KFFE is able to aggregate whereas KFFK and EFFE are not (Tjernberg et al. 2002). Taking into account the parameters mentioned above, some algorithms have been proposed to predict the effects of mutations on the aggregation rate of an unstructured polypeptide chain (provided that the mutation falls within the aggregation prone regions) (Chiti et al. 2003; Tartaglia et al. 2004), the absolute aggregation rate (DuBay et al. 2004), and the aggregation-prone regions of an unfolded protein (Fernandez-Escamilla et al. 2004; Pawar et al. 2005) solely on the basis of its primary structure.

#### 4.1.2 A possible role for $\pi$ -stacking in amyloid-like aggregation

It was also suggested that the presence of aromatic residues, particularly phenylalanine and tyrosine, promotes amyloid formation and stabilises the resulting amyloid fibrils (Azriel and Gazit 2001; Chelli et al. 2002; Gazit 2002; Porat et al. 2003; Porat et al. 2004). Usually, the interactions formed by several aromatic ring planes that are parallel to each other are referred to as  $\pi$ -stacking (Gazit 2002). A possible role of  $\pi$ -stacking in protein aggregation has been initially prompted by the observation that substitution of Phe23 with alanine in the most amyloidogenic segment of amylin (the segment NFGAILSS, corresponding to residues 22-29) results in a dramatically decreased aggregation propensity (Azriel and Gazit 2001). A molecular dynamics simulation has suggested that Phe23 may potentially allow a coherent association between sheets by cementing the macromolecular assemblies due to its low conformational flexibility when interacting with other aliphatic residues (Zanuy et al. 2004). Given the high frequency of aromatic residues in aggregation-prone fragments deriving from disease-related proteins, it was concluded that aromatic residues play a major role in the molecular recognition, which is likely to be a fundamental step in amyloid formation (Gazit 2002). Based on these data, an algorithm which predicts the change of aggregation rate upon mutation has been developed by taking into account aromaticity as an additional parameter (Tartaglia et al. 2004).

As evidence was accumulating on the possible role of  $\pi$ -stacking in amyloid formation, other authors have reported results that induce the importance of aromatic-aromatic interactions in amyloid assembly to be reconsidered (Tracz et al. 2004). A Phe to Leu substitution in the NNFGAILSS amylin segment (residues 21-29) does not prevent aggregation and formation of fibrils as observed with Fourier transform infrared spectroscopy (FTIR), transmission electron microscopy (TEM), and CR staining (Tracz et al. 2004). The F23L variant of the NFGAIL segment also results in amyloid formation at low and high pH values (Tracz et al. 2004). Substitution of the phenylalanine with an alanine results in a segment that is less prone to form fibrils than the corresponding wild type and F23L variants (Tracz et al. 2004). Aromatic residues are characterised by a high hydrophobicity and propensity to form  $\beta$ -sheet structure. Both

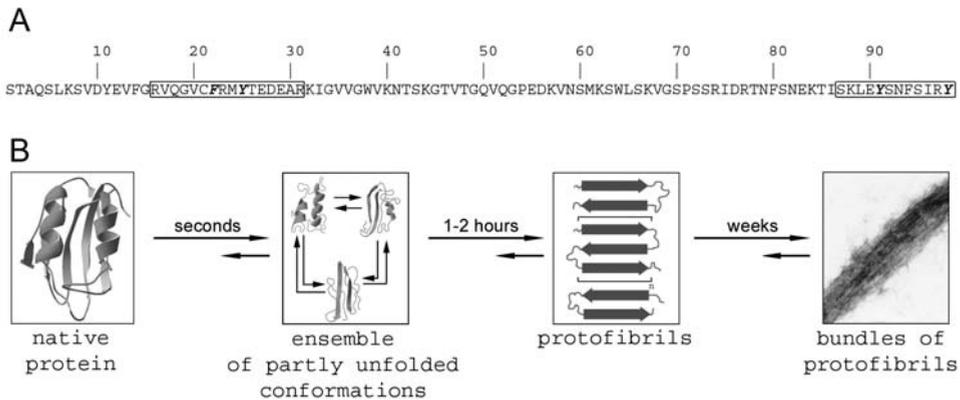


Figure 4.1: (A) The sequence of mt AcP. Regions in rectangles have been demonstrated to be important for the aggregation of the protein from a partially unfolded ensemble (Chiti et al. 2002b). Apart from Phe94, which does not seem to play a role in aggregation, the aromatic residues belonging to these two segments and mutated in the present analysis are shown in bold and italic. (B) The proposed aggregation pathway of mt AcP. Aggregation starts from an ensemble of partially unfolded conformations and involves formation of  $\beta$ -structured and ThT-binding protofibrils that convert later into long protofilaments and fibrils; reprinted, with permission, from (Chiti et al. 1999c), copyright 1993-2005 by the National Academy of Sciences of the USA. No detailed structural model has been proposed for the protofibrils; for descriptive purposes they are depicted in this figure as parallel  $\beta$ -sheets rich oligomers.

leucine and alanine residues are non-aromatic; the only difference is that leucine possesses a hydrophobicity and a  $\beta$ -sheet propensity greater than those of alanine; based on these results it was proposed that aromatic residues only affect aggregation because of these features rather than for their ability to form  $\pi$ -stacking. These conflicting reports raise the question as to whether aromaticity performs any role in the mechanism of amyloid formation and on the nature of the stabilising interactions in the resulting amyloid fibrils.

#### 4.1.3 mt AcP as a model for investigating $\pi$ -stacking

To answer to these questions we have focused our attention on human muscle acylphosphatase (mt AcP), a 98 residue enzyme introduced above (section 1.4.1 and (Pastore et al. 1992; Stefani et al. 1997)) whose sequence is shown in figure 4.1A. As introduced in section 1.4.2, although mt AcP is not associated with any known human disease, it can form aggregates which show an extensive  $\beta$ -sheet structure, as revealed by circular dichroism (CD) and FTIR spectroscopy, tinctorial properties typical of amyloid, such as a yellow-green birefringence under cross-polarised light in the presence of CR and a high fluorescence in the presence of ThT and a fibrillar appearance as detected with TEM (Chiti et al. 1999c). Initially, when incubated in a solution containing moderate concentrations of 2,2,2-Trifluoroethanol (TFE), mt AcP converts, on a time scale of a few seconds, into an ensemble of partially unfolded conformations with a far-UV CD spectrum indicative of extensive  $\alpha$ -helical content. Within 1-2 h the protein shows a transition from this state to an aggregated state that appears to be rich in

$\beta$ -sheet structure, with increased fluorescence with ThT, but lacking evidence for extended fibrils. No structural models have been proposed for these aggregates and it is not still clear whether their strands are parallel or anti-parallel. After 1-2 months electron micrographs show isolated as well as bundles of 3-5 nm wide protofilaments, which display CR birefringence (Chiti et al. 1999c). This sequential appearance of species during mt AcP aggregation is shown in figure 4.1B. A protein engineering approach has allowed the regions of the sequence that promote the conversion of the partially unfolded ensemble into  $\beta$ -structured aggregates to be determined (Chiti et al. 2002b). All the mutations that significantly alter the aggregation rate have been found in two regions of the primary structure corresponding to residues 16-31 and 87-98 (figure 4.1A). These two segments correspond to two insoluble peptides when dissected from the remainder of the sequence (Chiti et al. 2002b). Moreover, they have values of  $\beta$ -sheet propensity and hydrophobicity above the average values calculated from the entire mt AcP sequence (Chiti et al. 2002b) and appear to be solvent-exposed and/or flexible in the initial partially unfolded state (Monti et al. 2004).

The ability of mt AcP to convert into  $\beta$ -structured oligomers and fibrils, under conditions in which the protein is initially in a partially unfolded state, and the presence of aromatic residues within the two segments that promote such conversion make mt AcP a good model-system to study the importance of aromaticity in amyloid aggregation. Here we have determined the aggregation rate for a series of single point mutants of mt AcP in which aromatic residues have been substituted with other residues. The data have been analysed to assess the importance of aromatic residues in aggregation and to distinguish between aromaticity and other more generic effects in the possible aggregation-promoting action of these residues.

## 4.2 Results

### 4.2.1 Strategy employed

Two regions of the sequence of mt AcP have previously been found to promote the conversion of the partially unfolded ensemble populated in the presence of moderate concentrations of TFE into structured amyloid aggregates (Chiti et al. 2002b). These encompass residues 16-31 and 87-98 (figure 4.1A). These two segments contain five aromatic residues: Phe22, Tyr25, Tyr91, Phe94, and Tyr98. Phe94 was suggested to play a critical role in the folding mechanism of the molecule and in the stabilisation of the folding transition state and native structures (Chiti et al. 1999b; Vendruscolo et al. 2001). However, it does not seem to play a significant role in promoting aggregation directly, probably because the side chain of this residue is buried in the partially unfolded ensemble (Chiti et al. 2002b). In contrast, the four remaining aromatic residues appear to significantly influence aggregation, with the rate of assembly changing when they are mutated to other residues (Chiti et al. 2002b). For each of these four residues a set of single mutants has been produced with the wild type residue substituted with a large hydrophobic (leucine), a small hydrophobic (alanine), a hydrophilic (serine or glutamine), and a charged (arginine) residue. Table 1 reports a list of the produced variants.

The rate of conversion of the TFE-denatured state into aggregates rich in  $\beta$ -sheet and able to bind ThT was measured for each variant. This was achieved by incubating

all variants separately in 50 mM acetate buffer, 25% TFE, pH 5.5, 25 °C and monitoring the formation of aggregates using the ThT assay. Under this condition wild type and destabilised variants of mt AcP are known to denature rapidly on the time scale of a few seconds (Chiti et al. 1999c). The process of aggregation that is therefore monitored in the following minutes and hours consists in the conversion of the denatured ensemble into aggregates. The apparent change of aggregation rate following mutation is hence fully attributable to the effect of the amino acid substitution on this process rather than on the conformational destabilisation of the native state.

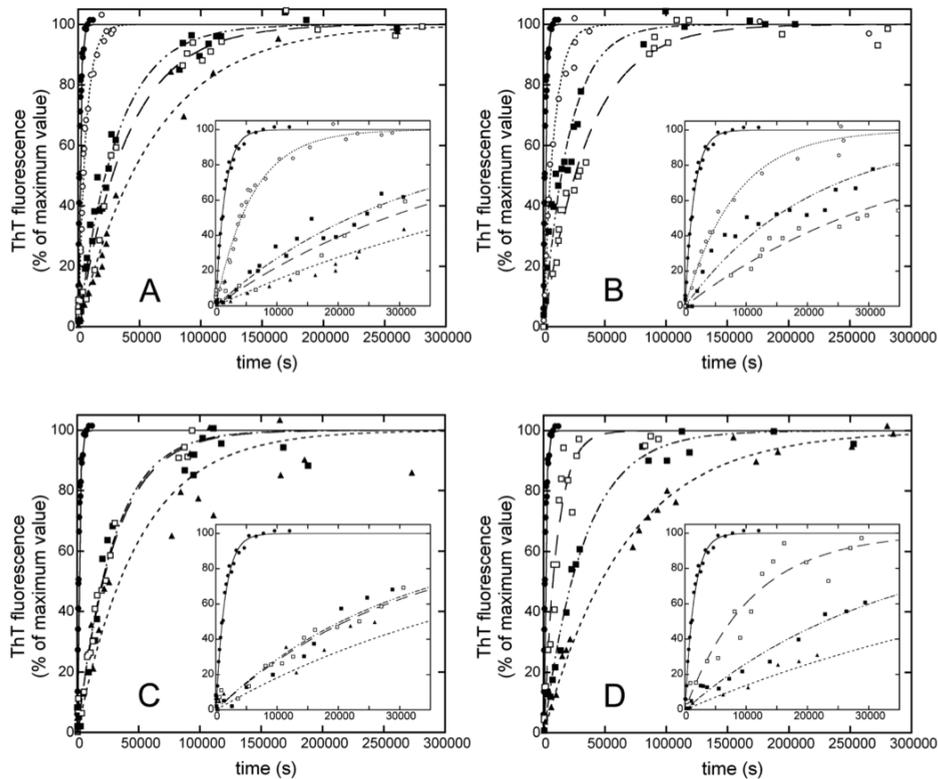


Figure 4.2: Aggregation kinetics for mutants of Phe22 (A), Tyr25 (B), Tyr91 (C), and Tyr98 (D). All panels show data for the wild type protein (●) and for mutations to leucine (○), alanine (■), a hydrophilic residue (□), and arginine (▲). The insets show the first day of recording. The lines represent best fits of collected data to single exponential equations (see section 4.5.2 and equation 4.1). The obtained rate constants  $\nu$  ( $s^{-1}$ ) are shown in table 4.1.

#### 4.2.2 Phenylalanine 22

Four mutants have been purified for this residue, with substitutions to leucine, alanine, serine, and arginine (table 4.1). The results show that the substitution of the phenyla-

lanine at this position results in a dramatically decreased aggregation rate (figure 4.2A).

The kinetic constants measured for all of the four mutants are significantly lower than that of the wild type (table 4.1). The most intense effect is observed for the F22R mutant, while the least effective deceleration is obtained when substituting phenylalanine with leucine. Mutations to alanine and serine result in similar effects, although substitution of the wild type residue with a hydrophilic one displays a slightly greater change in the aggregation rate.

Table 4.1: Kinetic data for a set of aromatic mutants of mt AcP. Observed  $\ln(v_{mut}/v_{wt})$  values have been obtained from best fits of experimental data to equation 4.1. Theoretical  $\ln(v_{mut}/v_{wt})$  values have been obtained from equation 4.2.

protein variant	experimental $\nu$ ( $s^{-1}$ )	observed $\ln(v_{mut}/v_{wt})$	theoretical $\ln(v_{mut}/v_{wt})$
WT	$(8.9 \pm 1.2) \cdot 10^{-4}$	-	-
F22L	$(1.7 \pm 0.2) \cdot 10^{-4}$	$-1.7 \pm 0.3$	-0.87
F22A	$(3.2 \pm 0.4) \cdot 10^{-5}$	$-3.3 \pm 0.3$	-2.31
F22S	$(2.5 \pm 0.3) \cdot 10^{-5}$	$-3.6 \pm 0.3$	-2.14
F22R	$(1.5 \pm 0.2) \cdot 10^{-5}$	$-4.1 \pm 0.3$	-5.02
Y25L	$(1.2 \pm 0.2) \cdot 10^{-4}$	$-2.0 \pm 0.3$	0.00
Y25A	$(4.8 \pm 0.6) \cdot 10^{-5}$	$-2.9 \pm 0.3$	-1.80
Y25S	$(2.7 \pm 0.3) \cdot 10^{-5}$	$-3.5 \pm 0.3$	-1.69
Y91A	$(3.4 \pm 0.4) \cdot 10^{-5}$	$-3.3 \pm 0.3$	-2.31
Y91Q	$(3.2 \pm 0.4) \cdot 10^{-5}$	$-3.3 \pm 0.3$	-2.16
Y91R	$(2.0 \pm 0.3) \cdot 10^{-5}$	$-3.8 \pm 0.3$	-5.00
Y98A	$(3.0 \pm 0.4) \cdot 10^{-5}$	$-3.4 \pm 0.3$	-1.66
Y98Q	$(8.5 \pm 1.1) \cdot 10^{-5}$	$-2.3 \pm 0.3$	-2.41
Y98R	$(1.5 \pm 0.2) \cdot 10^{-5}$	$-4.1 \pm 0.3$	-4.57

#### 4.2.3 Tyrosine 25

This residue has been substituted with leucine, alanine, and serine. Results are similar to those obtained for phenylalanine 22, with all substitutions resulting in slower aggregation (figure 4.2B; table 4.1). By comparing the Y25L and Y25S mutants we can see that the latter is much less prone to aggregate. Indeed, the kinetic constants  $\nu_{mut}$  are reduced by factors equal to 7.4 and 33, respectively. Similar decelerations were obtained for the F22L and F22S variants.

#### 4.2.4 Tyrosine 91

Tyr91 is located in the second aggregation-promoting segment (figure 4.1A). For this residue we have collected the kinetic profiles for the Y91A, Y91Q, and Y91R mutants (figure 4.2C). We have also designed an Y91L mutant, but it was not analysed due to purification problems. The results of the analysis of these mutants are shown in Figure 2C. The replacement of Tyr91 with a small hydrophobic residue or with a hydrophilic one results, within experimental error, in an identical effect (table 4.1). Interestingly, the order of aggregation speed is similar for the variants involving Phe22 and Tyr91: substitution of either residue to alanine or a hydrophilic group leads to decelerations

that are fairly similar to each other and comparable in the two sets of variants. In addition, in both cases, replacement to arginine leads to the most remarkable deceleration effect, with the aggregation reaction reaching equilibrium after many days.

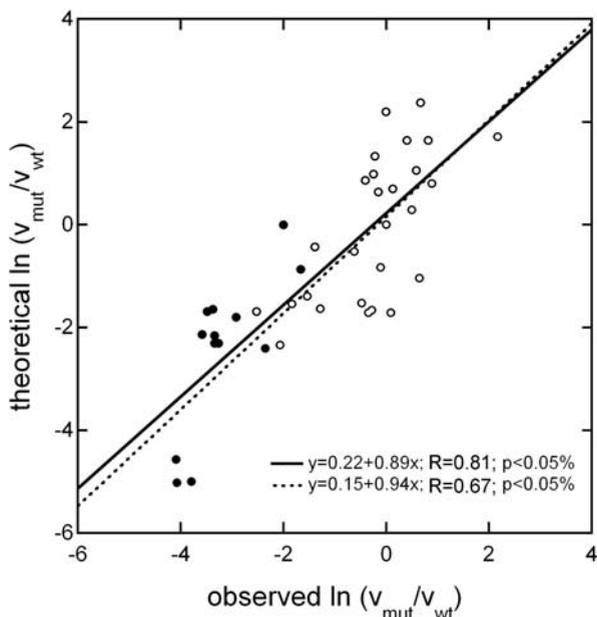


Figure 4.3: Calculated  $V_s$  observed change of aggregation rate upon mutation  $\ln(v_{mut}/v_{wt})$  for a set of mutants of mt AcP involving substitution of aromatic residues ( $\bullet$ ) along with a set of other mutants of mt AcP involving non-aromatic residues ( $\circ$ ) (Chiti et al. 2003). The  $\ln(v_{mut}/v_{wt})$  data have been calculated using equation 4.2. The continuous and dotted lines represent the best linear fits obtained using all and only non-aromatic mutants, respectively.

#### 4.2.5 Tyrosine 98

The aggregation kinetic profiles for mutants Y98Q, Y98A, and Y98R were also recorded (figure 4.2D). As for tyrosine 91, the Y98L variant could not be purified due to aggregation into inclusion bodies after expression in *Escherichia coli*. Although all mutations resulted in slower aggregation, along the lines observed for the other sets of variants, mutation to glutamine results in a less marked deceleration compared to the mutation of Tyr98 to alanine and also of Tyr91 to glutamine (table 4.1). Since tyrosine 98 is the C-terminal residue, it is possible that the C-terminal carboxyl group masks some of the effects observed for mutations at other positions. Apart from this effect, it is still evident that the most anti-amyloidogenic mutation is obtained if a charge residue is added, i.e., in the Y98R variant.

#### 4.2.6 Statistical analysis

The results presented here clearly show that aromatic residues play a fundamental role in promoting aggregation of mt AcP. Indeed, in all cases elimination of an aromatic residue results in a significant, sometimes remarkable, decrease of the aggregation rate.

Nevertheless, it is important to understand if aromatic residues favour amyloid formation because of their high hydrophobicity and propensity to form  $\beta$ -structure, or by forming aromatic-aromatic interactions. To address this issue, we have compared the experimentally obtained aggregation rates with those calculated theoretically by using a previously described algorithm (Chiti et al. 2003). The equation is reported in section 4.5.3 as equation 4.2 and allows the natural logarithm of the ratio of the aggregation rates for the mutant and wild type,  $\ln(v_{mut}/v_{wt})$ , to be calculated. The equation contains terms for hydrophobicity, net charge and free energy variation when a residue changes its conformation from an  $\alpha$ -helix to a  $\beta$ -strand. Moreover, the equation was derived by considering mainly mutations that did not involve aromatic residues (Chiti et al. 2003).

A summary of the experimental and theoretical values of  $\ln(v_{mut}/v_{wt})$  are reported in Table 1 and figure 4.3. In figure 4.3, we compare the data for the mutants which we have examined here (●) with those for other mt AcP mutants involving non-aromatic residues (○). These mutants have been previously used to develop the algorithm mentioned above (Chiti et al. 2003). Thus, if the experimental  $\ln(v_{mut}/v_{wt})$  values for aromatic mutants deviated from the expected behaviour, our conclusion would be that aromatic residues play a role in amyloid formation different from that expected on the basis of their hydrophobicity and ability to form secondary structure. By using equation 4.2 we notice that some experimental values of  $\ln(v_{mut}/v_{wt})$  are slightly lower or slightly higher than those calculated by taking into consideration only hydrophobicity, ability to form secondary structure, and net charge as parameters (figure 4.3). More importantly, the calculated aggregation rates of the mutants are not systematically higher than the observed values (figure 4.3). The data points for aromatic mutants are scattered around the line of best fit to a degree comparable with those of the non-aromatic variants. The p-parameter calculated for the best linear fit carried out with aromatic residues (continuous line) is lower than 0.05% (figure 4.3); the same result is obtained if the data from aromatic mutants are not included in the analysis (dotted line). A  $\chi^2$  test has also been carried out for the data set. The test yields the probability to obtain the distribution if the studied phenomenon is well represented by the tested law. A probability value lower than 5% has been obtained when equation 2 has been used. This suggests that it is not necessary to add an aromaticity term in the equation to improve the correlation between experimental and theoretical values.

## 4.3 Discussion

### 4.3.1 Aromatic residues promote amyloid aggregation of mt AcP due to their hydrophobicity and $\beta$ -sheet propensity

The aim of this work is to characterise the role of aromatic residues in amyloid formation and to understand whether the capability of aromatic residues to promote aggregation arises from their aromaticity or rather from their high hydrophobicity and  $\beta$ -sheet propensity. On the one hand the side chains of phenylalanine and tyrosine contain seven carbons. These residues, particularly phenylalanine, are ranked as highly hydrophobic according to all scales of hydrophobicity so far edited (Creighton 1993). On the other hand, it has been shown that the avoidance of steric clashes between the side chain

and its local backbone causes the  $\beta$ -sheet propensity of aromatic residues to be particularly high (Street and Mayo 1999). The high occurrence of phenylalanine and tyrosine residues in natural  $\beta$ -sheets of proteins also confirms the high propensity of these residues to form  $\beta$ -structure (Chou and Fasman 1974). In addition to having a high hydrophobicity and a high propensity to form  $\beta$ -sheet structure, phenylalanine and tyrosine residues also contain planar rings with six covalently bonded carbon atoms and a delocalised  $\pi$  system. This chemical characteristic, termed aromaticity, has been proposed to be responsible for the ability of these residues to promote amyloid fibril formation.

Our results show that substitution of any of the four aromatic residues present in the aggregation promoting stretches of mt AcP invariably results in a slower aggregation process of the entire molecule (figure 4.2; Table1). These observations confirm previous reports that phenylalanine and tyrosine side chains promote amyloid aggregation very effectively. However, our analysis rules out that aromaticity is responsible per se for the aggregating potential of these residues. An equation derived from mutations involving predominantly non-aromatic residues and considering physicochemical factors other than aromaticity can account for the observed reductions of aggregation rates when the aromatic residues of mt AcP are substituted with others (figure 4.3). The observed decelerations can be entirely attributed to the decrease of hydrophobicity and  $\beta$ -sheet propensity (and increase of net charge if applicable) following substitution. The effectiveness of aromatic residues in promoting amyloid formation has probably to be sought in these, rather than other, chemical characteristics of their side chains. Importantly, the scales of hydrophobicity and  $\beta$ -sheet propensity values for the 20 amino acid residues used in the algorithm were derived from partition coefficients between water and octanol (Creighton 1993) and effective stabilities when adopting a  $\beta$ -sheet structure in monomeric proteins (Street and Mayo 1999), respectively. These factors and the resulting algorithm are not therefore indirectly influenced by  $\pi$ -stacking or other types of aromatic-aromatic interactions.

#### 4.3.2 Aromatic residues are frequent in the cross- $\beta$ core of fibrils but are not necessarily required

Although aromaticity does not seem to be the origin of the efficacy of aromatic residues to promote aggregation, the fact remains that they can establish a network of hydrophobic interactions and can pay the minimal entropic cost in formation of the cross- $\beta$  structure. For these reasons they may have a significant stabilising effect in the resulting fibrils and it is not surprising that they are highly recurrent in amyloidogenic sequences. X-ray studies have led to a detailed characterisation of the crystals formed by assemblies of an amyloidogenic 12-mer peptide (Makin et al. 2005) and a 7-residue peptide derived from the yeast prion Sup35p (Nelson et al. 2005). In the first case stacking between phenylalanine residues is found to stabilise the inter-sheet packing of the structure (Makin et al. 2005), while in the second report tyrosine side chains stack on the solvent-exposed faces of the two sheets, presumably forming stabilising interactions (Nelson et al. 2005).

Several peptides containing residues 16-20 of the amyloid  $\beta$  peptide ( $A\beta$ ) readily form fibrils: The sequence of this segment, KLVFF, presents two aromatic residues (Tjernberg et al. 1999). A search for the residues that promote the aggregation of the entire  $A\beta$  peptide has shown the importance of Phe19 (Wurth et al. 2002). More inter-

estingly, solid state nuclear magnetic resonance (SS-NMR) experiments have led some authors to a model of the A $\beta$  1-40 protofilaments in which two  $\beta$ -strands formed by residues 12-24 and 30-40 give rise to two in register parallel  $\beta$ -sheets which interact through side chain-side chain contacts (Petkova et al. 2002). This structure clearly shows that Phe19 and Phe20 form inter-strand  $\pi$ -stacking (Petkova et al. 2002).

The contacts between phenylalanine residues in fibrils formed from the amylin-derived NFGAIL peptide are thought to be important stabilising interactions (Azriel and Gazit 2001; Gazit 2002; Porat et al. 2003; Porat et al. 2004). In the recently proposed structure of fibrils from full-length amylin any individual peptide molecule contributes to three  $\beta$ -strands, each of which is parallel and in register with the corresponding ones from other molecules (Kajava et al. 2005). In this structure not just Phe23, but also Phe15, His18, and Tyr37 are stacked along the axis of the fibril to form long rows of intermolecular interactions (Kajava et al. 2005).

Phe23 and Tyr37 form additional intra-molecular interactions in each peptide. It was suggested that in the NFGAIL segment the phenylalanine residue directs ordered  $\beta$ -sheet stacking through both specific interactions between aromatic rings and non-specific clustering of phenylalanine with other hydrophobic residues (Wu et al. 2005). This suggests that the hydrophobic properties of this residue are able to favour the aggregation of the entire segment without invoking its ability to form specific interactions.

If many peptides and proteins aggregate via interactions of aromatic residues, many others have been shown to promptly aggregate without any involvement of aromatic residues. A saturation mutagenesis analysis on the de novo designed amyloid peptide STVIIIE has allowed an aggregation prone sequence pattern to be determined (Lopez de la Paz and Serrano 2004). Although each position of this sequence can be mutated to aromatic residues without losing the ability of the hexapeptide to aggregate, the presence of aromatic residues does not seem to be an essential requirement for the aggregation of the hexapeptide (Lopez de la Paz and Serrano 2004). In  $\alpha$ -synuclein, a protein whose aggregation is related to Parkinson's disease, several short segments of 10 residues have been found to aggregate even if separated from the remainder of the sequence, for example the 71-82 (Giasson et al. 2001), 66-74 (Du et al. 2003), and 69-79 sequences (El-Agnaf and Irvine 2002). These three segments, whose sequences largely overlap to each other allowing an aggregation-prone region of the entire protein to be identified, do not contain aromatic residues. It has been proposed that the 31-37 region of the A $\beta$  sequence is important in aggregation of the full-length A $\beta$  peptide as mutations to proline in this region cause a significant decrease in the stability of the resulting fibrils (Williams et al. 2004). Accordingly, other authors have shown that the segment spanning approximately residues 30-38 gives rise to a  $\beta$ -strand in fibrils (Petkova et al. 2002; Torok et al. 2002). This segment does not contain aromatic residues.

Recently, the structure of the amyloid fibrils formed by HET-s from *Podospora anserina* has been investigated with SS-NMR, quenched hydrogen exchange and fluorescence (Ritter et al. 2005). The proposed structure shows two  $\beta$ -strand-turn- $\beta$ -strand motifs interconnected by a long loop (Ritter et al. 2005). Only  $\beta$ -strand 4 contains one aromatic residue, Tyr281. This residue interacts with a non-aromatic residue from  $\beta$ -strand 2 and does not form aromatic-aromatic interactions (Ritter et al. 2005). X-ray studies carried out on the crystals derived from assemblies of the amyloidogenic seg-

ment 1-24 from barnase show that the various phenylalanine residues at position 7 from different molecules do not form intermolecular interactions (Saiki et al. 2005). They seem to participate to a specific pattern in which Phe7 and Val10 residues are alternating on one face of the sheet (Saiki et al. 2005). Similarly, a pattern of alternating Tyr13 and Ile4 is present on the other side of the sheet.

#### 4.4 Conclusions

Our results on mt AcP confirm that aromatic residues have fundamental importance in amyloid assembly. Moreover, a survey of the amyloid fibril structures that have been determined with atomic or nearly atomic resolution in the past three years shows that the intermolecular interactions between the side chains of these residues appear to be highly frequent. However, the establishment of specific contacts between the aromatic rings of these residues is not an essential requirement to initiate aggregation and stabilise the resulting fibrils. When present, the forces that maintain in close contact the side chains of aromatic residues and stabilise the whole fibril do not arise from specific interactions involving the  $\pi$ -electrons or the aromatic nature of these residues but, rather, from their high hydrophobicity and high tendency to form  $\beta$ -sheets. We believe that the elucidation of the precise role played by aromatic moieties in determining the mechanism of amyloid fibril formation, the rate by which this process occurs, and the stability of the resulting fibrils is fundamental to gaining a better understanding of the processes occurring in amyloidogenesis. Such an understanding would also be crucial for improving the accuracy of the existing algorithms in determining aggregation rates and aggregation-promoting regions within polypeptide sequences.

#### 4.5 Materials and methods

##### 4.5.1 Mutagenesis, protein expression and purification

The gene encoding mt AcP was initially inserted in a pGEX-2T plasmid (Amersham, Little Chalfont, England) and the resulting construct used to transform DH5- $\alpha$  *E. coli* cells (Invitrogen, Carlsbad, California). Mutated genes were obtained using the Quick Change site-directed mutagenesis kit<sup>®</sup> from Stratagene (La Jolla, California) (see section 2.4.1). The presence of the desired mutations was assessed by DNA sequencing. Expression and purification of the wild type and mutant proteins were carried out according to the protocol of the pGEX-2T manufacturer. Protein purity was checked by SDS-polyacrylamide gel electrophoresis. The extinction coefficient at 280 nm ( $\epsilon_{280}$ ) for each mutant was calculated as previously described (Gill and von Hippel 1989).

##### 4.5.2 Aggregation kinetics with Thioflavin T fluorescence

Aggregation kinetics of mt AcP and its variants were carried out as previously described (Chiti et al. 2002b). In brief, the reaction was started by diluting the native protein into a solution to reach a final concentration equal to 0.4 mg ml<sup>-1</sup> in 25% (v/v) TFE, 50 mM acetate buffer (pH 5.5), and 25 °C. At several times, 60  $\mu$ l of the sample were added to 440  $\mu$ l of 25 mM ThT, 25 mM phosphate buffer (pH 6.0), at 25 °C. The result-

ing fluorescence was measured with a Perkin-Elmer LS-55 fluorometer and thermostated with a Thermo-HAAKE F8 bath. Excitation and emission wavelengths were 440 nm and 485 nm, respectively. The resulting plot of the ThT fluorescence, expressed as percentage of the maximum value versus time, was fitted to a single exponential equation of the following form:

$$y = A + B \cdot \exp(-v \cdot t) \quad (4.1)$$

where  $A$  is the ThT fluorescence at the apparent equilibrium,  $B$  is the change of ThT fluorescence during the exponential phase,  $v$  is the apparent rate constant, and  $t$  is the time. Before starting the experiment, the sample was centrifuged at  $20,000 \times g$  for 5 minutes and the protein concentration was measured using an extinction coefficient at 280 nm  $\epsilon_{280}$  value calculated as previously described (Gill and von Hippel 1989).

#### 4.5.3 Data calculation

The change of aggregation rate upon mutation is reported as  $\ln(v_{mut}/v_{wt})$ , where  $v_{mut}$  is the experimentally obtained aggregation rate constant of the considered protein variant and  $v_{wt}$  is the corresponding value for the wild type. For the calculation of the theoretical values of  $\ln(v_{mut}/v_{wt})$  the following equation was used (Chiti et al. 2003):

$$\ln(v_{mut}/v_{wt}) = 0.633 \cdot \Delta Hydr + 0.198 \cdot (\Delta \Delta G_{coil-\alpha} + \Delta \Delta G_{\beta-coil}) - 0.491 \cdot \Delta charge \quad (4.2)$$

where  $\Delta Hydr$ ,  $\Delta \Delta G_{coil-\alpha}$ ,  $\Delta \Delta G_{\beta-coil}$ , and  $\Delta charge$  are the change of hydrophobicity,  $\alpha$ -helical propensity,  $\beta$ -sheet propensity, and charge upon mutation, respectively.

## Chapter 5

### Final remarks

#### 5.1 Conformational states distinct from the fully folded structure can present enzymatic activity

As introduced in section 1.4.1, the acylphosphatase from *Sulfolobus solfataricus* (Sso AcP) is an enzyme able to hydrolyse benzoyl-phosphate (BP), with formation of benzoate and phosphate ions (Ramponi et al. 1966; Ramponi 1975; Camici et al. 1976; Cozza et al. 2006). In this thesis we have characterised in detail both folding (chapter 2) and amyloid-like aggregation (chapter 3) processes of this protein. In the first case we have studied a partially folded state that forms in the dead time of the stopped-flow refolding experiments (section 2.2). In the second case we have studied in further detail the role of an unstructured N-terminal segment in the mechanism of aggregation of the protein (section 3.2). Interestingly, in both cases a conformational state different from the native and folded one, yet presenting enzymatic activity comparable to the native state, is transiently populated.

In the case of amyloid-like aggregation of the protein it was previously shown that two distinct phases can be spectroscopically detected during process (Plakoutsi et al. 2005). At the end of the first phase, monitored by far-UV circular dichroism (CD) or light scattering, the protein transiently populates an aggregated state in which the Sso AcP molecules lack regular amyloid-like structure, as inferred from Fourier transform infrared (FTIR) spectroscopy, far-UV CD, Congo red (CR) and Thioflavin T (ThT) binding assays (Plakoutsi et al. 2005). Formation of these early aggregates is induced by an inter-molecular interaction between the unstructured N-terminal segment and the globular part of another molecule (section 3.2). Importantly, such early aggregates possess about 60% of the enzymatic activity of the native and monomeric protein; this enzymatic activity is completely absent after filtration, confirming that catalysis is due to the early aggregates and not to native protein still present in solution at the end of the process (Plakoutsi et al. 2005). The activity then decreases during the second phase of aggregation, when the native-like aggregates convert into protofibrils with regular amyloid-like structure (Plakoutsi et al. 2005).

An enzymatically active conformational state also forms transiently during the folding process. The folding process of Sso AcP in 50 mM acetate buffer at pH 5.5 and 37 °C is characterised by the formation of a partially folded state (Bemporad et al. 2004). This ensemble forms within the dead time of the stopped flow kinetic experiments and converts later into the fully native structure. The partially folded ensemble binds to 8-anilino-1-naphthalenesulfonic acid, suggesting the presence of hydrophobic clusters exposed to the solvent, and presents a far-UV mean residue ellipticity compa-

rable to that of the fully native state, indicating a native-like secondary structure comparable to that of the native state (Bemporad et al. 2004). In section 2.2 we have shown that this partially folded state is characterised by enzymatic activity equal to 80% of the fully folded state value. Further experiments have ruled out the possibility that the measured activity arises from presence of folded or unfolded protein. This activity is highly sensitive to mutations as the partially folded state is highly dynamic. Importantly, the catalytic site is not structured in the absence of substrate in the partially folded intermediate. Although folding and catalysis can be coupled in some cases (Vamvaca et al. 2004) and this possibility cannot be ruled out in our case, Sso AcP folding is not accelerated by the presence of substrate, suggesting that BP does not induce a global collapse of the structure. Finally, molecular dynamics simulations guided by experimentally obtained  $\Phi$  values suggest that this enzymatic activity is possible as in the partially folded ensemble the scaffold of the protein is already formed and this forces catalytic residues to be close to each other and able to carry out the catalytic cycle (section 2.3).

Importantly, there is no experimental evidence of any correlation between these two conformational states. They form in processes as different as folding and amyloid-like aggregation and in different experimental conditions such as in the absence and in the presence of 2,2,2-trifluoroethanol. Nevertheless, the general picture that derives from these experimental observations is that enzymatic activity can be more common than previously thought. Conformational states different from fully folded one can bind to the substrate, catalyse the reaction and release products to an extent comparable to folded structures. Since the first mechanism proposed for catalysis, usually referred to as “lock and key mechanism”, it became clear in the 60s that the substrate induces some changes in the enzyme structure (Koshland 1958). It is now clear the importance of dynamics during catalysis (Kiefhaber et al. 1992; Eisenmesser et al. 2002). The results presented in this thesis may extend this concept as highly dynamic and aggregated states can be active enzymes. Further studies are needed to clarify the biological relevance and distribution of such a feature. Nevertheless, these results are relevant as they extend our knowledge about properties of proteins in solution and give important information about their dynamics.

## 5.2 A comparison between the transition state ensembles of mt AcP and Sso AcP

In chapter 2 we have described an investigation, using  $\Phi$  value analysis, of the transition state ensemble (TSE) for folding of Sso AcP (see section 2.2.3). The results show that residues with  $\Phi_{\ddagger}$  values higher than 0.8 are Val20, Gly52, Ala58 and Arg71. Moreover, the obtained set of  $\Phi$  values (figure 2.5 and table 2.3) has been used in section 2.3 as a restraint in molecular dynamics simulations to obtain an ensemble of possible structures representing the TSE. The results show that the most structured region in TSE of Sso AcP corresponds to the interface between  $\beta$ -strand 1 and  $\alpha$ -helix 2. Moreover, a considerable amount of structure is also present in the  $\beta$ -hairpin between  $\beta$ -strand 2 and  $\beta$ -strand 3 (table 2.3). Structure formation in these regions forces the topology in the remainder of the molecule to be native-like. The regions structured in TSE of Sso AcP are also shown also in figure 5.1A.

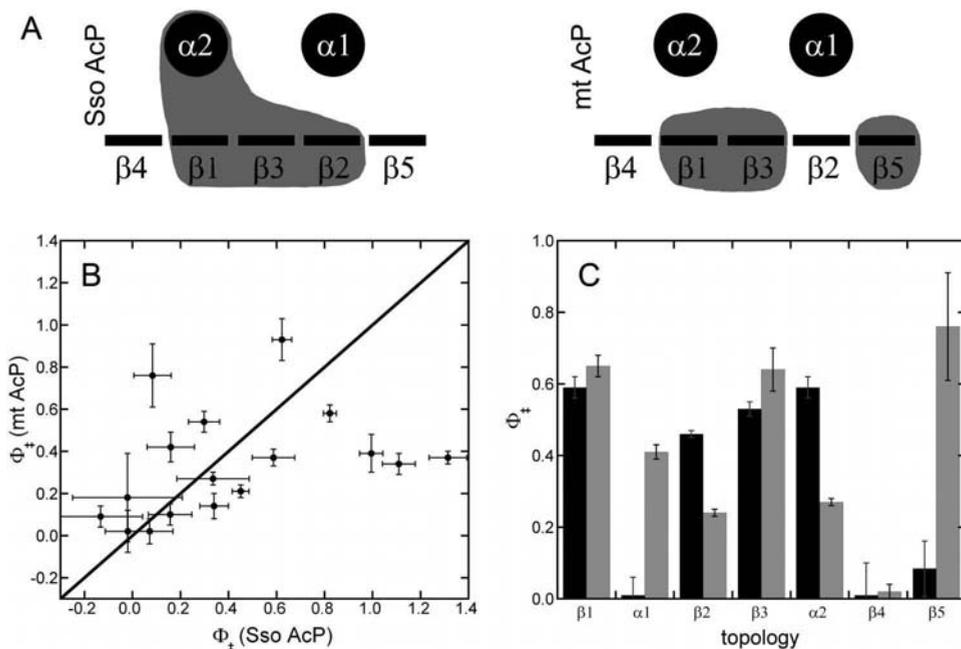


Figure 5.1: A comparison between folding transition states of Sso AcP and mt AcP. (A) a section view of the topology of Sso AcP and mt AcP. Regions structured in TSEs are highlighted in gray. (B) and (C) A comparison between  $\Phi_{\ddagger}$  values of Sso AcP and mt AcP. In (B) each mt AcP  $\Phi_{\ddagger}$  value is plotted versus the  $\Phi_{\ddagger}$  value of the corresponding residue of Sso AcP. The line represents the expected plot for two identical TSEs. In (C) the average values in the secondary structure elements are shown for mt AcP (gray) and Sso AcP (black). Values for mt AcP are reported in (Chiti et al. 1999b).

Many of the positions investigated in this work correspond to positions previously investigated in the TSE of mt AcP (Chiti et al. 1999b). In the case of mt AcP it was shown that the transition state structure is an expanded form of the native conformation, lacking persistent interactions but showing the topology characteristic of the native state (Chiti et al. 1999b). The most structured region in the TSE is the central section of the  $\beta$ -sheet ( $\beta$ -strands 1 and 3). The C-terminal  $\beta$ -strand ( $\beta$ -strand 5) and the loop connecting  $\beta$ -strands 2 and 3 also appear well formed in the transition state ensemble. The positions that present  $\Phi_{\ddagger}$  values higher than 0.7 correspond to Tyr11, Pro54, Phe94 (Chiti et al. 1999b).

The folding TSEs of mt AcP and Sso AcP present some similar features. In both cases the most structured region is the central part of the  $\beta$ -sheet, formed by  $\beta$ -strands 1 and 3 (figure 5.1A and C). Starting from these regions the structure propagates to the remaining secondary structure elements and to the loops. Tyr11 in mt AcP shows a  $\Phi_{\ddagger}$  value equal to  $0.83 \pm 0.07$  (Chiti et al. 1999b). The corresponding Sso AcP residue (Ala18) shows a  $\Phi_{\ddagger}$  value equal to  $0.62 \pm 0.04$  (table 2.3). Importantly, Val20, a residue close to Ala18, displays a  $\Phi_{\ddagger}$  value equal to  $1.31 \pm 0.08$ , suggesting structure formation in the region of Sso AcP that corresponds to the one structured in mt AcP. Moreover, in both transition states the least native-like secondary structure element is  $\beta$ -strand 4 ((Chiti et al. 1999b), table 2.3 and figure 5.1C).

However, some important features of these two TSEs significantly differ. In the case of Sso AcP the second  $\alpha$ -helix is rather structured, with three  $\Phi_{\ddagger}$  values significantly higher than 0 and one  $\Phi_{\ddagger}$  value (Arg71) equal to 1. By contrast, the interactions formed by residues located in the two  $\alpha$ -helices of mt AcP appear to be less consolidated (Chiti et al. 1999b). An important interaction is formed in the TSE of mt AcP by Phe94, which is positioned in  $\beta$ -strand 5. The corresponding position of Sso AcP displays a  $\Phi_{\ddagger}$  value close to 0, showing that this region is not structured in this protein. Finally, Tyr61 is the position of Sso AcP that corresponds to Pro54, the other structured residue in TSE of mt AcP. This residue could not be investigated due to the low destabilisation measured. Nevertheless, the plot representing the two  $\Phi_{\ddagger}$  datasets shows significant deviation from the trend expected for identical TSEs (figure 5.1B).

The TSE of mt AcP was found to be remarkably similar in structure to that of the activation domain of procarboxypeptidase A2 (ADA2h), a protein having the same overall topology but sharing only 13% sequence identity with mt AcP (Villegas et al. 1998; Chiti et al. 1999b). In many other cases similarities were found between folding processes and transition states of protein sharing the same fold (Travaglini-Allocatelli et al. 2004; Zarrine-Afsar et al. 2005; Chi et al. 2007). In the case of Sso AcP, some features make the TSE of the protein similar but not identical to that of the homologous protein mt AcP. These two proteins share 25% sequence identity. Nevertheless, they are evolutionary distant and the need to be active at high temperature results in higher conformational stability for Sso AcP (Corazza et al. 2006). This adaptation could also have implications for the folding mechanism. It is also possible that the partially folded state formed during folding of Sso AcP induces some constraints that affect the folding mechanism and consequently bias the TSE structure. Further characterisation of this state will shed light on the folding mechanism of Sso AcP and of the acylphosphatase-like family.

### 5.3 Different regions of the sequence are involved in protein folding and amyloid-like aggregation

It was previously shown that different regions play major roles in the amyloid-like aggregation and in the folding process of mt AcP (Chiti et al. 2002b).  $\beta$ -strands 1 and 3 and Phe94 are the most structured regions in the folding nucleus of this protein, whereas all the mutations that significantly alter the aggregation rate are located in two regions of the mt AcP primary structure corresponding to residues 16-31 and 87-98 (figure 4.1A). The 16-31 segment spans the loop that follows  $\beta$ -strand 1 and  $\alpha$ -helix 2. The 87-98 segment spans the loop that follows  $\beta$ -strand 4 and  $\beta$ -strand 5. These two segments correspond to two insoluble peptides when dissected from the remainder of the sequence (Chiti et al. 2002b). Moreover, they have values of  $\beta$ -sheet propensity and hydrophobicity above the average values calculated from the entire AcP sequence (Chiti et al. 2002b) and appear to be solvent-exposed and/or flexible in the initial partially unfolded state (Monti et al. 2004). These data led the authors to conclude that there is a kinetic partitioning between folding and aggregation. Starting from the unfolded state, the propensity of the full length protein to aggregate is reduced by intramolecular interactions associated with the formation of the folding nucleus. This suggests that the sequences of natural proteins may have evolved such that partially struc-

tured states populated during folding have the tendency to resist amyloid-like aggregation under normal conditions (Chiti et al. 2002b).

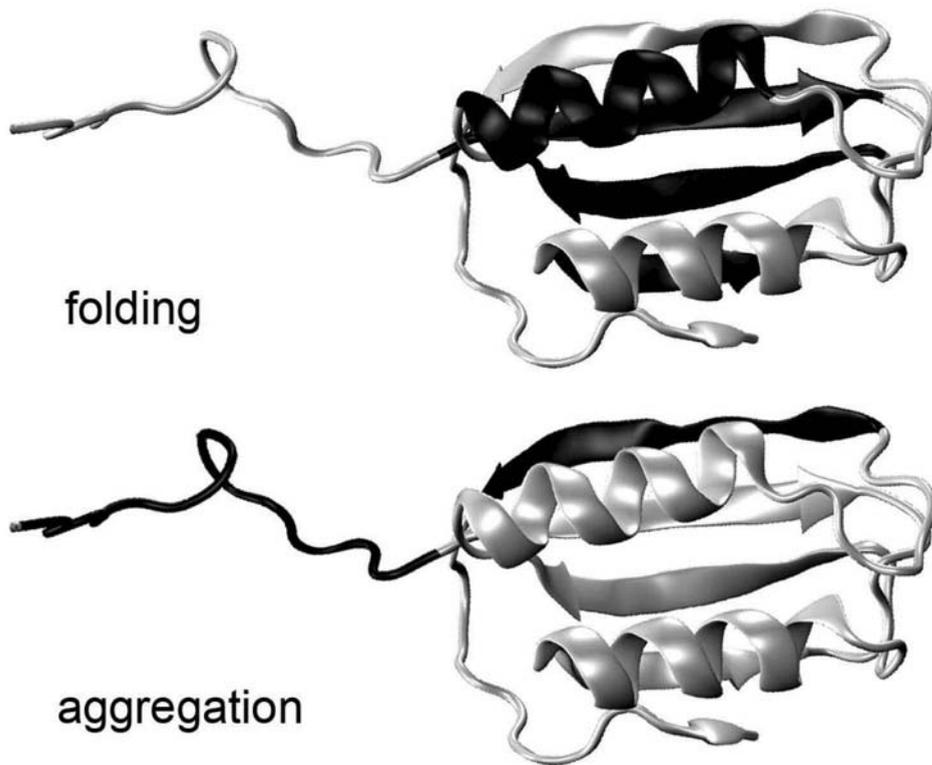


Figure 5.2: A comparison between the regions important for folding and amyloid-like aggregation of Sso AcP. On the top the regions of the sequence structured in the transition state for folding are coloured in black. These span  $\beta$ -strands 1, 2 and 3 and  $\alpha$ -helix 2. Residues in these segments display some  $\Phi_i$  values close to 1 and many  $\Phi_i$  values significantly higher than 0 (table 2.3). On the bottom regions important for amyloid-like aggregation of the protein are coloured in black. These span the fourth  $\beta$ -strand and the unstructured N-terminal segment. These regions were found flexible and exposed to the solvent and important from protein engineering studies ((Plakoutsi et al. 2006) and chapter 3). The figure has been drawn with VMD 1.8.3 for win32 (Humphrey et al. 1996).

Interestingly, similar partitioning between regions important for folding and regions that play major roles in the amyloid-like aggregation process can be done for Sso AcP as well. As shown in figure 5.2 and mentioned above (section 5.2), in the folding transition state structure is present in  $\beta$ -strands 1, 2 and 3 and in  $\alpha$ -helix 2. By contrast, the regions that play a major role in the mechanism of aggregation of the protein are  $\beta$ -strand 4 and the N-terminal unstructured segment (section 3.3 and (Plakoutsi et al. 2006)). Thus, in the case of Sso AcP as well there is a separation between regions that participate to folding and regions that participate to aggregation.

Importantly, Sso AcP aggregates starting from a native-like ensemble. This implies that the biological function supposed for the observed mt AcP partitioning cannot

be extended to Sso AcP. The absence of structure in the N-terminal segment explains the reason why this region does not play any role in protein folding, while its flexibility provides an explanation for its importance in amyloid-like aggregation. In the case of the fourth  $\beta$ -strand, we have shown that mutations in this region strongly accelerate unfolding (table 2.3). This probably implies that in destabilising conditions this strand gets much more flexible and prone to aggregate. Concerning the role of  $\beta$ -strand 4 in folding, one can speculate that evolution acted on this edge  $\beta$ -strand placing in this region some gatekeeper residues that reduce aggregation propensity of the native protein, as it was shown in the case of some unfolded (Otzen et al. 2000) and folded (Richardson and Richardson 2002) proteins. Presence of these residues introduces some constraints that force this region to fold only after topology of the protein is formed.

#### 5.4 Conclusions

In this thesis we have investigated in detail the processes of protein folding and amyloid-like aggregation in the acylphosphatase-like family. These studies are particularly important as they provide clues to better understand the dynamics of proteins in solution. Indeed, the search for the right fold is a crucial event for proteins and any impairment of the mechanisms that regulate equilibria between different conformational states can trigger misfolding events and formation of deleterious aggregates related to a set of human diseases (Chiti and Dobson 2006). Only more detailed knowledge about the mechanisms and the parameters that lead to misfolding will allow the achievement of the level of knowledge required to understand and treat these diseases and to design novel proteins that are perfectly functional in solution.

## Acknowledgements

I want to acknowledge here all the people who were close to me and helped me during these years.

First of all, I want to thank Elisabetta as her love and her constant presence gave me the serenity needed in the many difficult moments and allowed me to face with courage the difficulties of life. A special thank goes to my dad and my brother, for their patience and for their support, not only in the last years. I always think to my mother, to whom this thesis is dedicated and to whom I owe much of what I am now. Finally, I want to thank my friends, in particular Roberto, Ilaria, Margherita, Giacomo and Gemma for the numerous chats and evenings spent together.

Coming to professional acknowledgements, first of all I want to thank Prof. Chiti, my supervisor, for many teachings and for stimulating discussions about our experimental results. I want to thank Prof. Stefani and Accademia dei Lincei for supporting my research. Coming to my colleagues, I want to thank Tommaso for his not only technical contribution, Silvia C. for many useful and productive scientific discussions and Matteo for his informatics support. Finally, I want to thank Alessandro, Alessio, Claudia, Elemir, Elena, Elodie, Francesca, Georgia, Giordana, Giulia, Julia, Lara, Laura, Lorena, Martina, Nadia, Silva, Silvia G. e Valentina (in rigorous alphabetical order) for their suggestions, helpfulness and for the pleasant moments spent together outside the lab.



## Appendix A

### Equations and formulas

In this appendix we shall discuss the most important equations that have been used for data fitting and analysis in this work. Some passages will be omitted for brevity reasons.

#### A.1 Equilibrium unfolding experiments

##### A.1.1 Equilibrium unfolding with guanidine hydrochloride

All equilibrium unfolding experiments carried out in this work using guanidine hydrochloride as a denaturant have shown a reversible and cooperative transition. Thus, they have been analysed using Santoro and Bolen theory (Santoro and Bolen 1988). According to this model a given protein can populate two conformational states, folded ( $F$ ) and unfolded ( $U$ ). These states are in equilibrium and no partially folded conformations form:



folded and unfolded protein are characterised by significantly different spectroscopic signals. In the experiments presented here the most significant difference between  $U$  and  $F$  has been detected using circular dichroism (CD) signal at 222 nm. The CD signals for  $U$  and  $F$  are referred to as  $[\Theta]_{222}^U$  and  $[\Theta]_{222}^F$ , respectively. The amount of folded and unfolded protein at a given denaturant concentration  $[D]$ , referred to as folded fraction  $f_F$  and unfolded fraction  $f_U$ , are given by the following equations:

$$f_F = \frac{[F]}{[F] + [U]} = \frac{[\Theta]_{222} - [\Theta]_{222}^U}{[\Theta]_{222}^F - [\Theta]_{222}^U} \quad (\text{A.1})$$

$$f_U = \frac{[U]}{[F] + [U]} = 1 - f_F = \frac{[\Theta]_{222}^F - [\Theta]_{222}}{[\Theta]_{222}^F - [\Theta]_{222}^U} \quad (\text{A.2})$$

where  $[\Theta]_{222}$  is the CD signal observed at the equilibrium at  $[D]$ ,  $[U]$  and  $[F]$  are the concentrations of  $U$  and  $F$ , respectively. Moreover, an equilibrium constant  $K_U$  between  $U$  and  $F$  can be defined using standard thermodynamics theory:

$$K_U = \frac{[U]}{[F]} = \exp\left(-\frac{\Delta G_{U-F}}{R \cdot T}\right) \quad (\text{A.3})$$

where  $\Delta G_{U-F}$  is the free energy change upon denaturation,  $R$  is ideal gas constant and  $T$  is the temperature. It has been shown that  $\Delta G_{U-F}$  linearly depends on denaturant concentration  $[D]$  (Pace 1986):

$$\Delta G_{U-F} = \Delta G_{U-F}^{H_2O} - m \cdot [D] \quad (\text{A.4})$$

where  $\Delta G_{U-F}^{H_2O}$  is the free energy change upon denaturation in the absence of denaturant and  $m$  is the slope of the straight line. Combining equations from A.1 to A.4 we obtain:

$$\frac{[\Theta]_{222} - [\Theta]_{222}^F}{[\Theta]_{222}^U - [\Theta]_{222}^F} = \exp\left(-\frac{\Delta G_{U-F}^{H_2O} - m \cdot [D]}{R \cdot T}\right). \quad (\text{A.5})$$

Solving equation A.5 in  $[\Theta]_{222}$  we have

$$[\Theta]_{222} = \frac{[\Theta]_{222}^F + [\Theta]_{222}^U \cdot \exp\left(-\frac{\Delta G_{U-F}^{H_2O} - m \cdot [D]}{R \cdot T}\right)}{1 + \exp\left(-\frac{\Delta G_{U-F}^{H_2O} - m \cdot [D]}{R \cdot T}\right)}. \quad (\text{A.6})$$

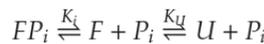
Since the CD signals of folded and unfolded proteins linearly depend on denaturant concentration we obtain

$$[\Theta]_{222} = \frac{([\Theta]_{222}^{F,H_2O} + a_1[D]) + ([\Theta]_{222}^{U,H_2O} + a_2[D]) \cdot \exp\left(-\frac{\Delta G_{U-F}^{H_2O} - m \cdot [D]}{R \cdot T}\right)}{1 + \exp\left(-\frac{\Delta G_{U-F}^{H_2O} - m \cdot [D]}{R \cdot T}\right)}. \quad (\text{A.8})$$

where  $[\Theta]_{222}^{F,H_2O}$  and  $[\Theta]_{222}^{U,H_2O}$  are the CD signals in the absence of denaturant respectively for folded and unfolded protein,  $a_1$  and  $a_2$  are the slopes of the straight lines. Best fits of experimental data to equation A.6 give a quantitative measurement of  $\Delta G_{U-F}^{H_2O}$  and  $m$ . Finally, we can calculate the concentration of middle denaturation  $C_m$  as the denaturant concentration value at which  $K_U = 1$  and  $\Delta G_{U-F} = 0$ . That is, from equation A.4,

### A.1.2 Equilibrium unfolding with guanidine hydrochloride and phosphate

Phosphate ion is a well known competitive inhibitor of acylphosphatase activity (Stefani et al. 1997). Adding phosphate at a concentration  $C_{P_i}$  to a sample containing Sso AcP will change the apparent equilibrium. In particular  $U$  is not able to bind to phosphate  $P$  and unfolding can occur only following phosphate release from native protein  $F$ :



where  $FP_i$  represents the folded protein bound to the phosphate,  $P_i$  represents the unbound phosphate. Moreover, the affinity constant of phosphate  $K_i$  is

$$K_i = \frac{[F] \cdot [P_i]}{[FP_i]} \quad (\text{A.9})$$

where  $[P_i]$  represents the concentration of unbound  $P_i$ . In the presence of  $P_i$  the amount of folded protein will increase and this will result in an apparent protein stabilisation. The stabilisation  $\Delta \Delta G_{U-F}^{P_i}$  due to the presence of  $P_i$  is given by:

$$\Delta\Delta G_{U-F}^{P_i} = \Delta G_{U-F}^{P_i} - \Delta G_{U-F}^{H_2O} \quad (\text{A.10})$$

where  $\Delta G_{U-F}^{P_i}$  and  $\Delta G_{U-F}^{H_2O}$  are the free energy changes upon denaturation in the absence of denaturant, and in the presence and in the absence of  $P_i$ , respectively.  $\Delta\Delta G_{U-F}^{P_i}$  can be calculated as follows:

$$\begin{aligned} \Delta\Delta G_{U-F}^{P_i} &= -RT \ln\left(\frac{[U]}{[F] + [FP_i]}\right) + RT \ln\left(\frac{[U]}{[F]}\right) = \\ &= RT \ln\left(1 + \frac{[F \cdot P_i]}{[F]}\right) = \\ &= RT \ln\left(1 + \frac{[P_i]}{K_i}\right) \end{aligned} \quad (\text{A.11})$$

where  $[P_i]$  can be approximated to  $C_{P_i}$  as phosphate is in large excess relative to the protein. Thus, the phosphate concentration to be used to obtain a certain stabilisation  $\Delta\Delta G_{U-F}^{P_i}$  is given by

$$C_{P_i} = K_i \cdot \left[ \exp\left(\frac{\Delta\Delta G_{U-F}^{P_i}}{RT}\right) - 1 \right]. \quad (\text{A.12})$$

Equation A.12 has been used in section 3.2.2 to calculate the amount of phosphate to be added to obtain a certain stabilisation of Sso AcP.

## A.2 Folding and unfolding kinetics

### A.2.1 Folding kinetics with fluorescence

Folding of wild type Sso AcP was shown to be characterised by three distinct fluorescence phases (Bemporad et al. 2004): **(1)** a rapid increase, escaping the dead time of the stopped flow experiments and corresponding to the formation of an ensemble of partially folded conformations (PFE); **(2)** a rapid decrease, corresponding to the folding of about 90% of the protein; **(3)** a slow decrease, corresponding to the folding of remaining 10% of the protein, which is rate-limited by *cis-trans* isomerism of Leu49-Pro50 peptide bond (Bemporad et al. 2004).

All mutants investigated in this study showed similar fluorescence traces. To explain such behaviour we have supposed the following equilibrium:



In this equilibrium  $I_{cis}$  and  $I_{trans}$  represent the PFE with the proline in *cis* and *trans*,  $F$  is the native state,  $k_{12}$ ,  $k_{21}$  and  $k_{I \rightarrow F}$  are kinetic constants. By applying the law of mass action to this equilibrium we obtain the following system:

$$\begin{cases} \frac{d}{dt} [I_{cis}]_{(t)} &= -k_{12} [I_{cis}]_{(t)} + k_{21} [I_{trans}]_{(t)} \\ \frac{d}{dt} [I_{trans}]_{(t)} &= k_{12} [I_{cis}]_{(t)} - (k_{21} + k_{I \rightarrow F}) [I_{trans}]_{(t)} \\ \frac{d}{dt} [F]_{(t)} &= k_{I \rightarrow F} [I_{cis}]_{(t)} \end{cases} \quad (\text{A.13})$$

This system can be solved with boundary conditions. In particular

- At a given time  $t$ , the total concentration of protein remains constant:

$$[F]_{(t)} + [I_{cis}]_{(t)} + [I_{trans}]_{(t)} = C_{tot}$$

where  $C_{tot}$  is the total concentration of protein.

- At the beginning of the process no molecules in the folded conformation exist; the relative amount of the two PFEs is given by kinetic constants:

$$[F]_{(0)} = 0; \quad \frac{[I_{cis}]_{(0)}}{[I_{trans}]_{(0)}} = \frac{k_{21}}{k_{12}}; \quad [I_{cis}]_{(0)} + [I_{trans}]_{(0)} = C_{tot}.$$

- At the equilibrium only the native state is populated:

$$[F]_{(\infty)} = C_{tot}; \quad [I_{cis}]_{(\infty)} = 0; \quad [I_{trans}]_{(\infty)} = 0.$$

With these conditions system A.13 has the following solution:

$$\begin{cases} [I_{cis}]_{(t)} &= [I_{cis}]_{(0)} \cdot \exp(-k_{12} \cdot t) \\ [I_{trans}]_{(t)} &= [I_{trans}]_{(0)} \cdot \exp(-k_{I \rightarrow F} \cdot t) \\ [F]_{(t)} &= [I_{trans}]_{(0)} [1 - \exp(-k_{I \rightarrow F} \cdot t)] + [I_{cis}]_{(0)} [1 - \exp(-k_{12} \cdot t)] \end{cases} \quad (\text{A.14})$$

Thus, the fluorescence signal at a given time  $f_{(t)}$  is a linear combination of the signals of the three considered states:

$$f_{(t)} = a_1 \cdot [I_{cis}]_{(t)} + a_2 \cdot [I_{trans}]_{(t)} + a_3 \cdot [F]_{(t)} \quad (\text{A.15})$$

where  $a_1$ ,  $a_2$  and  $a_3$  are proportional to the fluorescence emission intensity of  $I_{cis}$ ,  $I_{trans}$  and  $F$ , respectively. Substituting equations A.14 in equation A.15 we have

$$\begin{aligned} f_{(t)} &= [I_{cis}]_{(0)} (a_1 - a_3) \cdot e^{(-k_{12} \cdot t)} + [I_{trans}]_{(0)} (a_2 - a_3) \cdot e^{(-k_{I \rightarrow F} \cdot t)} + \\ &+ a_3 ([I_{cis}]_{(0)} + [I_{trans}]_{(0)}). \end{aligned} \quad (\text{A.16})$$

Hence, by substituting

$$[I_{cis}]_{(0)} (a_1 - a_3) = A_1; \quad [I_{trans}]_{(0)} (a_2 - a_3) = A_2; \quad a_3 ([I_{cis}]_{(0)} + [I_{trans}]_{(0)}) = q$$

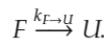
in equation A.16 we obtain

$$f_{(t)} = A_1 \cdot e^{(-k_{12} \cdot t)} + A_2 \cdot e^{(-k_{I \rightarrow F} \cdot t)} + q. \quad (\text{A.17})$$

Equation A.17 is able to reproduce experimental results for both wild type Sso AcP and the set of mutants studied. Best fits of experimental data to equation A.17 have been used in section 2.2.3 to obtain quantitative measurements of folding rates of several Sso AcP variants.

## A.2.2 Unfolding kinetics with circular dichroism

Unfolding kinetics of Sso AcP were followed by means of circular dichroism signal at 230 nm. In all cases experimental results showed a single exponential behaviour. This can be explained supposing a two state equilibrium between the unfolded state  $U$  and the folded state  $F$ :



We suppose that folding reaction is much slower in the experimental conditions used for unfolding. Thus, applying the law of mass action we obtain the following system:

$$\begin{cases} \frac{d}{dt} [F]_{(t)} &= -k_{F \rightarrow U} [F]_{(t)} \\ \frac{d}{dt} [U]_{(t)} &= k_{F \rightarrow U} [F]_{(t)} \end{cases} \quad (\text{A.18})$$

Imposing boundary conditions about the amount of F and U at the beginning of unfolding one can get the following solutions:

$$[F]_{(t)} = C_{tot} \cdot e^{(-k_{F \rightarrow U}t)}; \quad [U]_{(t)} = C_{tot} \cdot [1 - e^{(-k_{F \rightarrow U}t)}] \quad (\text{A.19})$$

where  $C_{tot}$  is the total protein concentration. As reported above for fluorescence during folding (see section A.2.1), we can now calculate the change in signal over time  $[\Theta]_{(t)}$  as:

$$\begin{aligned} [\Theta]_{(t)} &= a_1 \cdot [F]_{(t)} + a_2 \cdot [U]_{(t)} = \\ &= a_1 \cdot C_{tot} \cdot e^{(-k_{F \rightarrow U}t)} + a_2 \cdot C_{tot} \cdot [1 - e^{(-k_{F \rightarrow U}t)}] \end{aligned} \quad (\text{A.20})$$

where  $a_1$  and  $a_2$  are proportional to the microscence of F and U, respectively. By substituting

$$C_{tot} \cdot (a_1 - a_2) = A; \quad a_2 \cdot C_{tot} = q \quad (\text{A.21})$$

we obtain

$$[\Theta]_{(t)} = A \cdot e^{(-k_{F \rightarrow U}t)} + q. \quad (\text{A.22})$$

Equation A.22 is able to reproduce experimental results for both wild type Sso AcP and the set of mutants studied. Best fits of experimental data to equation A.17 have been used in section 2.2.3 to obtain quantitative measurements of unfolding rates of several Sso AcP variants.

### A.2.3 Chevron plots

The plot reporting natural logarithm of apparent folding and unfolding rate constants  $\ln k_{app}$  versus denaturant concentration  $[D]$  is usually referred to as chevron plot (Jackson and Fersht 1991). In a two state folder the obtained behaviour can be explained supposing that folding and unfolding rate constants linearly depend on denaturant concentration:

$$\ln k_{U \rightarrow F} = \ln k_{U \rightarrow F}^{H_2O} + m_u \cdot [D] \quad (\text{A.23})$$

$$\ln k_{F \rightarrow U} = \ln k_{F \rightarrow U}^{H_2O} + m_f \cdot [D] \quad (\text{A.24})$$

where  $\ln k_{U \rightarrow F}$  and  $\ln k_{F \rightarrow U}$  are respectively natural logarithm of folding and unfolding rate constants,  $\ln k_{F \rightarrow U}^{H_2O}$  and  $\ln k_{U \rightarrow F}^{H_2O}$  are respectively natural logarithm of folding and unfolding rate constants in the absence of denaturant,  $m_u$  and  $m_f$  are the slopes of the straight lines. The apparent constant  $k_{app}$  is the sum of folding and unfolding constants:

$$k_{app} = k_{F \rightarrow U} + k_{U \rightarrow F}. \quad (\text{A.25})$$

Thus, we can calculate  $\ln k_{app}$  substituting equations A.23 and A.24 in A.25:

$$\begin{aligned} \ln k_{app} &= \ln (k_{F \rightarrow U} + k_{U \rightarrow F}) = \\ &= \ln \left[ k_{F \rightarrow U}^{H_2O} \cdot \exp(m_u \cdot [D]) + k_{U \rightarrow F}^{H_2O} \cdot \exp(m_f \cdot [D]) \right] \end{aligned} \quad (\text{A.26})$$

Equation A.26 reproduces chevron plots for two state folders. However, the chevron plots for the Sso AcP variants studied in this work showed a down-ward curvature at low denaturant concentrations. This was attributed to the formation of an ensemble of partially folded conformations in the dead time of the stopped flow experiments (Bemporad et al. 2004). Thus, a modified version of equation A.26 was used:

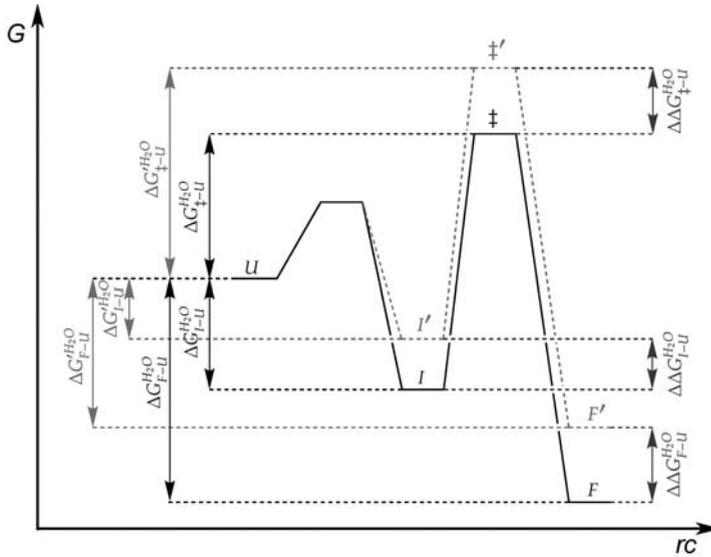


Figure A.1: Schematic representation of free energy  $G$  versus reaction coordinate  $rc$  during folding of Sso AcP in the absence of denaturant. Symbols are defined in the text. The partially folded ensemble  $I$  is supposed to be on-pathway. The effect of a generic mutation is represented in red. For simplicity reasons, no effect has been drawn for the transition state between unfolded state  $U$  and  $I$ . Blue colour refers to parameters used in the text to define  $\Phi_I^{H_2O}$  and  $\Phi_{\ddagger}^{H_2O}$  (equation A.28).

$$\ln k_{app} = \ln \left[ k_{F \rightarrow U}^{H_2O} \cdot \exp(m_u \cdot [D]) + k_{I \rightarrow F}^{H_2O} \cdot \exp(m_1 \cdot [D]^2 + m_2 \cdot [D] + m_3) \right] \quad (\text{A.27})$$

where  $k_{I \rightarrow F}^{H_2O}$  is the folding rate constant starting from the partially folded ensemble in the absence of denaturant. A second order polynomial function has been introduced in the folding arm of the plot. Equation A.27 is able to reproduce chevron plots for the Sso AcP protein variants studied in section 2.2.3.

### A.3 Calculation of $\Phi$ values

$\Phi$  value analysis has been carried out in this study on the partially folded ensemble (PFE) and transition state ensemble (TSE) populated during the folding of Sso AcP (figure A.1).  $\Phi$  values for PFE,  $\Phi_I^{H_2O}$ , and for TSE,  $\Phi_{\ddagger}^{H_2O}$ , are defined as (Matouschek and Fersht 1991; Matouschek et al. 1992):

$$\Phi_I^{H_2O} = \frac{\Delta\Delta G_{I-U}^{H_2O}}{\Delta\Delta G_{F-U}^{H_2O}}, \quad \Phi_{\ddagger}^{H_2O} = \frac{\Delta\Delta G_{\ddagger-U}^{H_2O}}{\Delta\Delta G_{F-U}^{H_2O}} \quad (\text{A.28})$$

where  $\Delta\Delta G_{I-U}^{H_2O}$ ,  $\Delta\Delta G_{\ddagger-U}^{H_2O}$  and  $\Delta\Delta G_{F-U}^{H_2O}$  are the change in conformational stability in the absence of denaturant for PFE, TSE and native state, respectively.

Calculation of  $\Delta\Delta G_{F-U}^{H_2O}$  has been carried out as reported in A.1.1. In particular

$$\Delta\Delta G_{F-U}^{H_2O} = \Delta G_{F-U}^{H_2O} - \Delta G_{F-U}^{H_2O} \quad (\text{A.29})$$

where the prime refers to the mutant. To calculate  $\Delta G_{F-U}$  and  $\Delta G'_{F-U}$  the average  $\underline{m}$  value over the mutants was used in equation A.8 as the  $m$  value that one can obtain

by the best fit of a single mutant to the Santoro & Bolen model arises from the few points in the transition zone of the plot and is therefore highly sensitive to the experimental error (Matouschek and Fersht 1991). Thus,

$$\Delta\Delta G_{F-U}^{H_2O} = (C'_m - C_m) \cdot \bar{m}. \quad (\text{A.30})$$

To calculate  $\Delta\Delta G_{I-U}^{H_2O}$ , with reference to the figure, one has to notice that

$$\begin{aligned} \Delta\Delta G_{I-U}^{H_2O} &= \Delta G'_{I-U}{}^{H_2O} - \Delta G_{I-U}^{H_2O} = \\ &= \Delta G'_{I-U}{}^{H_2O} - \Delta G_{I-U}^{H_2O} + G'_F{}^{H_2O} - G_F{}^{H_2O} + G_F{}^{H_2O} - G_F{}^{H_2O} = \\ &= \Delta\Delta G_{F-U}^{H_2O} - \Delta\Delta G_{F-I}^{H_2O}. \end{aligned} \quad (\text{A.31})$$

Similar calculations allow us to find  $\Delta\Delta G_{F-I}^{H_2O}$ :

$$\begin{aligned} \Delta\Delta G_{F-I}^{H_2O} &= \Delta G'_{F-I}{}^{H_2O} - \Delta G_{F-I}^{H_2O} = \\ &= \Delta G'_{F-I}{}^{H_2O} - \Delta G_{F-I}^{H_2O} + G'_\ddagger{}^{H_2O} - G_\ddagger{}^{H_2O} + G_\ddagger{}^{H_2O} - G_\ddagger{}^{H_2O} = \\ &= \Delta\Delta G_{\ddagger-I}^{H_2O} - \Delta\Delta G_{\ddagger-F}^{H_2O}. \end{aligned} \quad (\text{A.32})$$

$\Delta\Delta G_{\ddagger-I}^{H_2O}$  and  $\Delta\Delta G_{\ddagger-F}^{H_2O}$  can now be calculated from kinetic experiments. In fact, using Arrhenius equation we can calculate folding rates as follows:

$$k_{I\rightarrow F}^{H_2O} = \frac{\kappa \cdot k_B T}{h} \cdot \exp\left(\frac{\Delta G_{\ddagger-I}^{H_2O}}{RT}\right); \quad k'_{I\rightarrow F}{}^{H_2O} = \frac{\kappa \cdot k_B T}{h} \cdot \exp\left(\frac{\Delta G'_{\ddagger-I}{}^{H_2O}}{RT}\right) \quad (\text{A.33})$$

where  $k_{I\rightarrow F}^{H_2O}$  and  $k'_{I\rightarrow F}{}^{H_2O}$  are the folding rate constants in the absence of denaturant for the wild type and mutant,  $\kappa$  is a frequency factor,  $k_B$  and  $h$  are Boltzmann and Planck constants, respectively. From equation A.33 we obtain

$$\Delta\Delta G_{\ddagger-I}^{H_2O} = -RT \ln\left(\frac{k'_{I\rightarrow F}{}^{H_2O}}{k_{I\rightarrow F}^{H_2O}}\right). \quad (\text{A.34})$$

Unfolding rate constants can be calculated as follows:

$$k_{F\rightarrow U}^{H_2O} = \frac{\kappa \cdot k_B T}{h} \cdot \exp\left(\frac{\Delta G_{\ddagger-F}^{H_2O}}{RT}\right); \quad k'_{F\rightarrow U}{}^{H_2O} = \frac{\kappa \cdot k_B T}{h} \cdot \exp\left(\frac{\Delta G'_{\ddagger-F}{}^{H_2O}}{RT}\right) \quad (\text{A.35})$$

where  $k_{F\rightarrow U}^{H_2O}$  and  $k'_{F\rightarrow U}{}^{H_2O}$  are the unfolding rate constants in the absence of denaturant for wild type and mutant, respectively. From equation A.35 we obtain

$$\Delta\Delta G_{\ddagger-F}^{H_2O} = -RT \ln\left(\frac{k'_{F\rightarrow U}{}^{H_2O}}{k_{F\rightarrow U}^{H_2O}}\right). \quad (\text{A.36})$$

Combining equations from A.28 to A.36 we experimentally find  $\Phi$  values:

$$\Phi_I^{H_2O} = 1 - \left[ \frac{-RT \ln\left(\frac{k'_{I\rightarrow F}{}^{H_2O} \cdot k_{F\rightarrow U}^{H_2O}}{k_{I\rightarrow F}^{H_2O} \cdot k'_{F\rightarrow U}{}^{H_2O}}\right)}{(C'_m - C_m) \cdot \bar{m}} \right]; \quad (\text{A.37})$$

$$\Phi_\ddagger^{H_2O} = 1 - \left[ \frac{-RT \ln\left(\frac{k_{F\rightarrow U}^{H_2O}}{k'_{F\rightarrow U}{}^{H_2O}}\right)}{(C'_m - C_m) \cdot \bar{m}} \right]. \quad (\text{A.38})$$

Equations A.37 and A.38 have been used in chapter 2 to calculate  $\Phi$  values of a set of Sso AcP variants (see section 2.4.8).

#### A.4 Development of enzymatic activity during folding

Folding of Sso AcP has been followed in chapter 2 by means of enzymatic activity. Sso AcP is able to hydrolyse benzoyl-phosphate (BP) producing phosphate and benzoate ions (Stefani et al. 1997; Corazza et al. 2006). Unlike the hydrolysis products, the reactant has a significant extinction coefficient at 283 nm (Ramponi et al. 1966). Thus, enzymatic activity is proportional to the opposite of the first order derivative of the absorbance at 283 nm  $A_{283(t)}$ .

The behaviour shown by the Sso AcP variants investigated when unfolded protein is diluted into refolding buffer can be explained by supposing that no differences exist between the two intermediates analysed in section A.2.1. Thus, a single intermediate converts into the native state. In these conditions concentrations of partially folded state  $[I]_{(t)}$  and native state  $[F]_{(t)}$  are given by the following equation:

$$[I]_{(t)} = C_{tot} \cdot e^{(-k_{I \rightarrow F} \cdot t)}; \quad [F]_{(t)} = C_{tot} \cdot [1 - e^{(-k_{I \rightarrow F} \cdot t)}] \quad (\text{A.39})$$

where  $k_{I \rightarrow F}$  is the folding rate constant and  $C_{tot}$  is the protein concentration (see section A.2.1). Both states are characterised by a catalytic constant  $k_{CAT}$ . The enzymatic activity will be a linear combination of the activities of the partially folded state  $I$  and fully folded state  $F$ :

$$-\frac{d}{dt}A_{283(t)} = k_{CAT}^I \cdot [I]_{(t)} + k_{CAT}^F \cdot [F]_{(t)} \quad (\text{A.40})$$

where  $k_{CAT}^I$  and  $k_{CAT}^F$  are catalytic constants for partially folded and fully folded states, respectively. Equation A.39 allows us to calculate the expected behaviour for  $A_{283}$  as a function of time,  $A_{283(t)}$ . Indeed, from equation A.39 we obtain

$$-\int_{A_{283(0)}}^{A_{283(t)}} dA_{283(t)} = \int_0^t (k_{CAT}^I \cdot [I]_{(t)} + k_{CAT}^F \cdot [F]_{(t)}) dt. \quad (\text{A.41})$$

Equation A.41 can be solved to obtain  $A_{283(t)}$ :

$$A_{283(t)} = A_{283(0)} + \frac{C_{tot}}{k_{I \rightarrow F}} (k_{CAT}^F - k_{CAT}^I) + \frac{C_{tot}}{k_{I \rightarrow F}} (k_{CAT}^I - k_{CAT}^F) \cdot e^{-k_{I \rightarrow F} \cdot t} - k_{CAT}^F C_{tot} t. \quad (\text{A.42})$$

Equation A.42 reproduces experimental traces obtained in section 2.2.1. Moreover, substituting equation A.39 in equation A.40 we obtain the enzymatic activity as a function of time:

$$-\frac{d}{dt}A_{283(t)} = C_{tot} [k_{CAT}^F - (k_{CAT}^F - k_{CAT}^I) \cdot e^{-k_{I \rightarrow F} \cdot t}]. \quad (\text{A.43})$$

Equation A.43 has been used in section 2.2.1 to analyse data and obtain quantitative measurements of the catalytic constants of partially folded and fully folded states of the Sso AcP variants analysed in the text.

## Appendix B

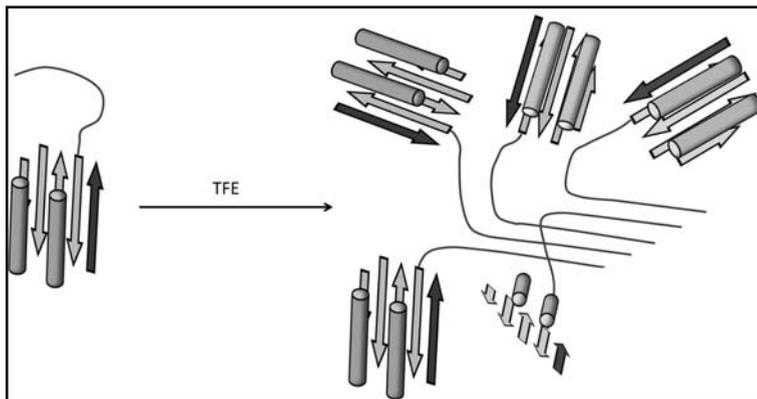
### Models for Sso AcP aggregation

#### B.1 Introduction

In this appendix the possible models for the Sso AcP aggregation mechanism based on previously obtained experiments (section 3.1.2) are discussed on the basis of the experiments presented in this thesis 3.2. The fourth  $\beta$ -strand is depicted in dark gray. The N-terminal unstructured segment is depicted as a filament.

#### B.2 Description of models

##### B.2.1 Interaction between N-termini

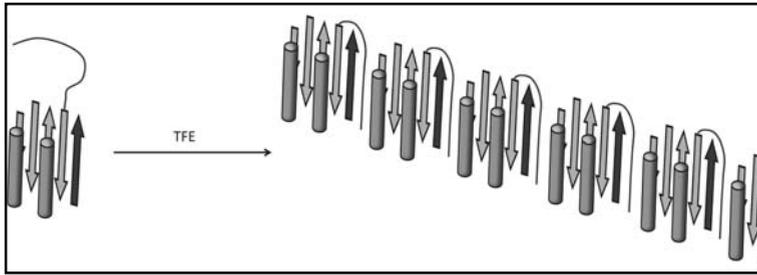


In this model the early aggregates are stabilised by intermolecular interactions between N-terminal tails of two molecules. The globular part of Sso AcP participates only to the second phase of the process.

This model does not reproduce the experimental results as

1. The peptide alone should aggregate in the aggregation promoting conditions ((Plakoutsi et al. 2006) and section 3.2.3).
2.  $\beta$ -strand 4 should not appear important as it is actually detected by site directed mutagenesis and limited proteolysis (Plakoutsi et al. 2006).

### B.2.2 Interaction between N-terminus and $\beta$ -strand 4

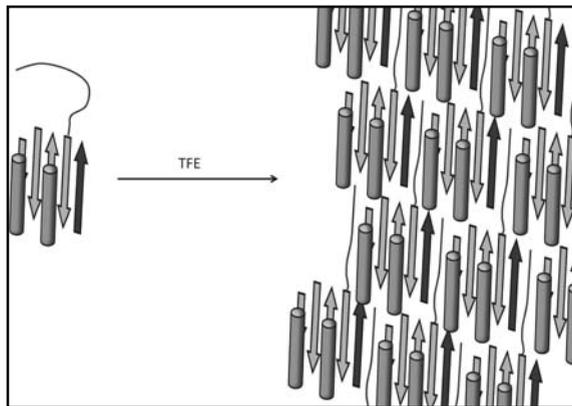


In this model the N-terminal tail acts as a bridge between the fourth  $\beta$ -strand and another region of the molecule (the fifth  $\beta$ -strand is shown for instance in the figure). This induces formation of early aggregates.

This model does not reproduce the experimental results as:

1. At least one more region should play a key role in the aggregation in addition to N-terminus and  $\beta$ -strand 4 (Plakoutsi et al. 2006).
2.  $\Delta$ N11 Sso AcP should aggregate in the presence of peptides (section 3.2.3).
3.  $\Delta$ N11 Sso AcP should co-aggregate in the presence of wild type protein (section 3.2.3).

### B.2.3 Interaction between N-terminus and $\beta$ -strand 4; interaction between globular parts

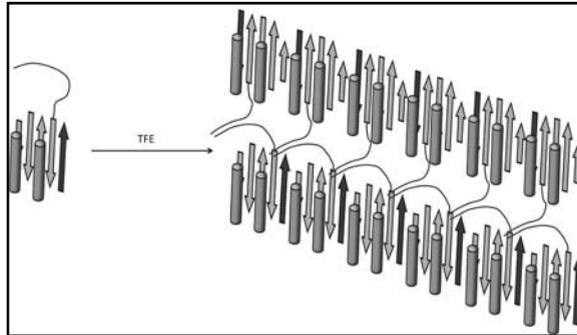


This is a variant of the previous model. In this case (see the figure) bridging of the N-terminal tail does not give rise to a filamentous aggregate in which Sso AcP molecules interact with two partners. By contrast, a larger assembly forms that can expand in three dimensions as N-terminal segment interacts with other regions of other Sso AcP molecules as well (in the figure a two dimension aggregate is shown for simplicity reasons).

This model does not reproduce the experimental results as:

1. At least one more region should play a key role in the aggregation in addition to N-terminus and  $\beta$ -strand 4 (Plakoutsi et al. 2006).

2.  $\Delta$ N11 Sso AcP should aggregate in the presence of peptides (section 3.2.3).
3.  $\Delta$ N11 Sso AcP should co-aggregate in the presence of wild type protein (section 3.2.3).



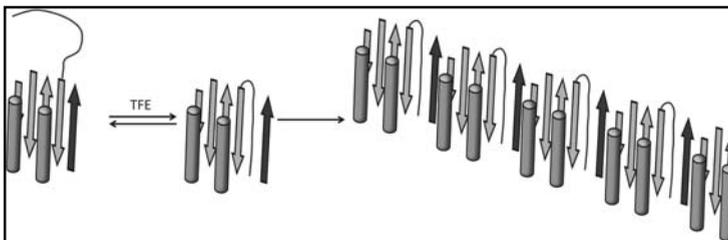
#### B.2.4 Separate interactions between N-termini and globular parts

Both N-terminus and fourth  $\beta$ -strand play a major role in the process. Aggregation of Sso AcP is mediated by interactions between the fourth  $\beta$ -strand and the globular part of another molecule and by intermolecular interactions of the N-terminal tails.

This model does not reproduce the experimental results as:

1.  $\Delta$ N11 Sso AcP should give rise to oligomers in the aggregation promoting conditions (section 3.2.3).
2. The peptide alone should aggregate in the aggregation promoting conditions ((Plakoutsi et al. 2006) and section 3.2.3).
3. Wild type Sso AcP should hijack  $\Delta$ N11 Sso AcP into early aggregates (section 3.2.3).
4. At least one more region should play a key role in the aggregation in addition to N-terminus and  $\beta$ -strand 4 (Plakoutsi et al. 2006).

#### B.2.5 Intra-molecular interaction between N-terminus and $\beta$ -strand 4



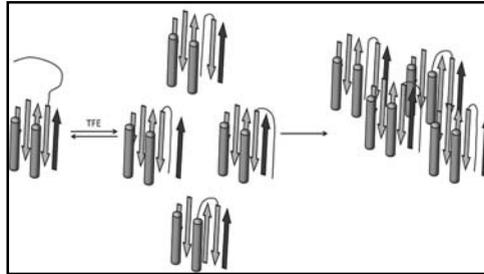
In the presence of 2,2,2-Trifluoroethanol (TFE) the N-terminal peptide gives rise to an intra-molecular interaction with  $\beta$ -strand 4. Thus, an aggregation nucleus forms that polymerises generating the early aggregates.

This model does not reproduce the experimental results as:

1. The mutant with the tail at C-terminus should not aggregate because the unstructured tail cannot reach  $\beta$ -strand 4 in the mutant (section 3.2.1).

2. Since the interaction is an intra-molecular one, presence of peptides should speed up aggregation of wild type Sso AcP and induce aggregation of  $\Delta N11$  Sso AcP (section 3.2.3).
3. At least one more aggregation promoting region should be found (Plakoutsi et al. 2006).

### B.2.6 Non-specific intra-molecular interaction between N-terminal segment and globular part of the Sso AcP molecule

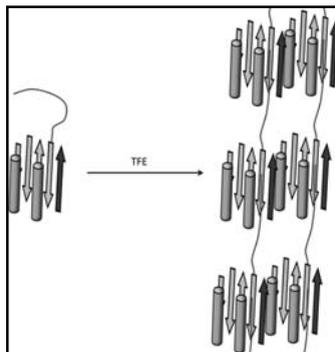


In the presence of TFE the N-terminal peptide gives rise to a non-specific intra-molecular interaction with different regions of the globular part of Sso AcP as the molecule somehow opens due to the destabilisation in the TFE solvent (see the figure). This leads to the formation of an aggregation prone state that is able to initiate the amyloid-like aggregation process.

This model does not reproduce the experimental results as:

1. Since the interaction is an intra-molecular one, presence of peptides should speed up aggregation of wild type Sso AcP and induce aggregation of  $\Delta N11$  Sso AcP (section 3.2.3).
2. Since the interaction is non-specific, no aggregation promoting regions should be found (Plakoutsi et al. 2006).

### B.2.7 N-terminus and $\beta$ -strand 4 interact with other regions

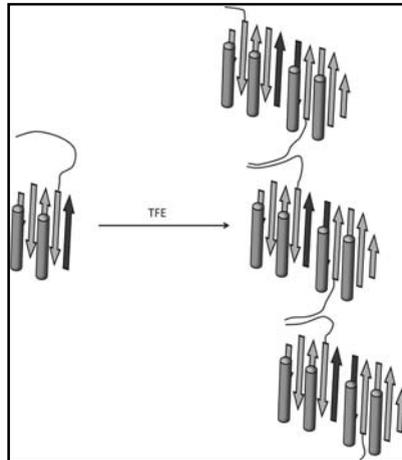


In this model both N-terminal tail and fourth  $\beta$ -strand play a major role in the process. However, they do not interact with each other. They interact specifically with different regions of the molecule.

This model does not reproduce the experimental results as:

1.  $\Delta N11$  Sso AcP should give rise to dimers in the aggregation promoting conditions (section 3.2.3).
2. At least two more aggregation promoting regions should be found with mutagenesis and limited proteolysis (Plakoutsi et al. 2006).

#### B.2.8 Interactions N-terminus-N-terminus and $\beta$ -strand 4- $\beta$ -strand 4

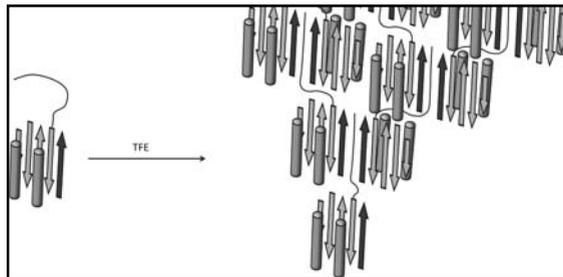


This model is built on the idea that early aggregates are stabilised by tail-to tail interactions and strand-to-strand interactions. In the figure above the strands stack to each other giving rise to an antiparallel sheet but one can imagine a parallel sheet as well.

This model does not reproduce the experimental results as:

1.  $\Delta N11$  Sso AcP should give rise to dimers in the aggregation promoting conditions (section 3.2.3).
2. The peptide alone should aggregate in the aggregation promoting conditions ((Plakoutsi et al. 2006) and section 3.2.3).

#### B.2.9 N-terminus acts as a bridge between two $\beta$ -strands 4

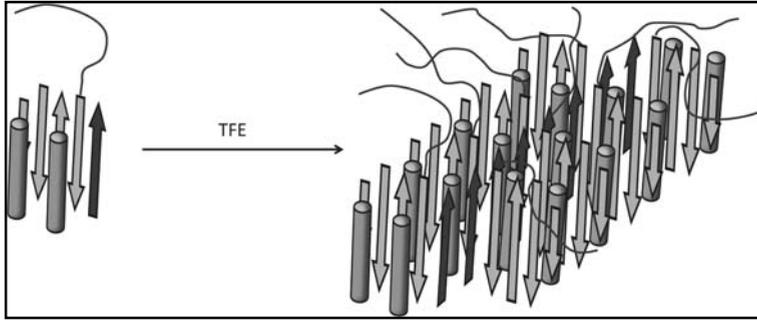


This is another possible model built using only the N-terminus and the fourth  $\beta$ -strand only. In this case an interaction between the strands is mediated by the terminus. A pyramid-like aggregate forms.

This model does not reproduce the experimental results as:

1.  $\Delta N11$  Sso AcP should dimerise in conditions that promote aggregation of wild type protein (section 3.2.3).

#### B.2.10 Destabilisation induced by N-terminus

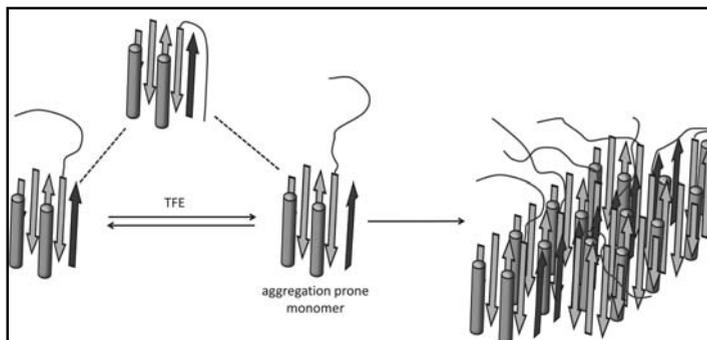


In this model N-terminal tail and fourth  $\beta$ -strand play different roles in the process. Since a stabilisation was shown in  $\Delta N11$  Sso AcP (Plakoutsi et al. 2006), it is possible that in the presence of TFE the tail destabilises the globular part of Sso AcP. This facilitates fluctuations within the native-like state populated in TFE. Thus, the fourth  $\beta$ -strand leads to the formation of native-like assemblies.

This model does not reproduce the experimental results as:

1. The I72V mutation on  $\Delta N11$  Sso AcP should induce aggregation on the same time scale as the wild type protein (section 3.2.2).
2. Peptides should not affect the wild type aggregation (section 3.2.3).

#### B.2.11 N-terminus induces formation of an aggregation prone monomer through specific interactions



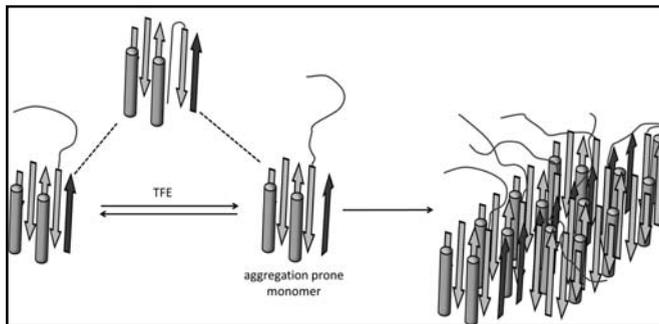
In the presence of TFE the native monomer and an aggregation prone monomer in which the fourth  $\beta$ -strand is in an amyloidogenic conformation are in equilibrium. The tail plays a major role in the transition state of this equilibrium by interacting specifically with the fourth  $\beta$ -strand. This interaction is a fundamental step in the

formation of the aggregation prone monomer. Starting from such a state the fourth  $\beta$ -strand leads to the formation of native-like assemblies.

This model does not reproduce the experimental results as:

1. Peptides should promote aggregation of  $\Delta$ N11 Sso AcP (section 3.2.3).
2. Peptides should increase the aggregation rate of wild type Sso AcP (section 3.2.3).
3. C-tail Sso AcP should not aggregate (section 3.2.1).

#### B.2.12 N-terminus induces formation of an aggregation prone monomer through non-specific interactions

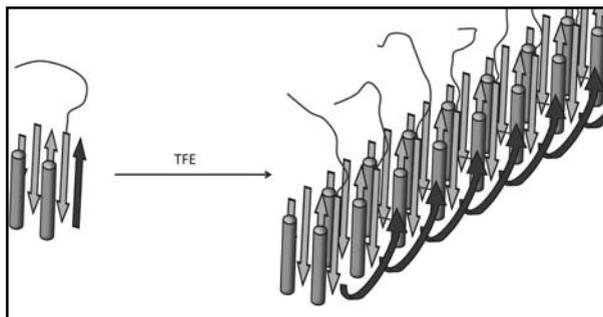


In the presence of TFE the native monomer and an aggregation prone monomer in which the fourth  $\beta$ -strand is in an amyloidogenic conformation are in equilibrium. The tail plays a major role in the transition state of this equilibrium by interacting with the globular unit in a non-specific way. Starting from this aggregation prone monomer, the fourth  $\beta$ -strand leads to the formation of native-like assemblies.

This model does not reproduce the experimental results as:

1. Peptides should promote aggregation of  $\Delta$ N11 Sso AcP (section 3.2.3).
2. Peptides should increase the aggregation rate of wild type Sso AcP (section 3.2.3).

#### B.2.13 A domain swapping based model



This model is based on domain swapping. In the presence of TFE  $\beta$ -strand 4 of a molecule replaces  $\beta$ -strand 4 of the following molecule. This leads to the formation of early aggregates.

This model does not reproduce the experimental results as:

1.  $\Delta N11$  Sso AcP should aggregate and co-aggregate with wild type protein (section 3.2.3).
2. I72V- $\Delta N11$  Sso AcP should aggregate (section 3.2.2).
3. Peptides should have no effect on the aggregation of both wild type Sso AcP and  $\Delta N11$  Sso AcP (section 3.2.3).

## Bibliography

- Anfinsen C.B. (1973) Principles that govern the folding of protein chains. *Science* 181, 223-230.
- Anfinsen C.B., Haver E., Sela M., and White F.H.J. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* 47, 1309-1314.
- Azriel R., and Gazit E. (2001) Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. *J Biol Chem* 276, 34156-34161.
- Bai Y. (1999) Kinetic evidence for an on-pathway intermediate in the folding of cytochrome c. *Proc Natl Acad Sci U S A* 96, 477-480.
- Baldwin R.L. (1989) How does protein folding get started? *Trends Biochem Sci* 14, 291-294.
- Bemporad F., Calloni G., Campioni S., Plakoutsi G., Taddei N., and Chiti F. (2006) Sequence and structural determinants of amyloid fibril formation. *Acc Chem Res* 39, 620-627.
- Bemporad F., Capanni C., Calamai M., Tutino M.L., Stefani M., and Chiti F. (2004) Studying the folding process of the acylphosphatase from *Sulfolobus solfataricus*. A comparative analysis with other proteins from the same superfamily. *Biochemistry* 43, 9116-9126.
- Benkovic S.J., and Hammes-Schiffer S. (2003) A perspective on enzyme catalysis. *Science* 301, 1196-1202.
- Bouchard M., Zurdo J., Nettleton E.J., Dobson C.M., and Robinson C.V. (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci* 9, 1960-1967.
- Bousset L., Briki F., Doucet J., and Melki R. (2004a) The native-like conformation of Ure2p in fibrils assembled under physiologically relevant conditions switches to an amyloid-like conformation upon heat-treatment of the fibrils. *J Struct Biol* 141, 132-142.
- Bousset L., Redeker V., Decottignies P., Dubois S., Le Marechal P., and Melki R. (2004b) Structural characterization of the fibrillar form of the yeast *Saccharomyces cerevisiae* prion Ure2p. *Biochemistry* 43, 5022-5032.
- Bousset L., Thomson N.H., Radford S.E., and Melki R. (2002) The yeast prion Ure2p retains its native  $\alpha$ -helical conformation upon assembly into protein fibrils in vitro. *EMBO J* 17, 2903-2911.
- Bradley P., Misura K.M., and Baker D. (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* 309, 1868-1871.

- Bucciantini M., Giannoni E., Chiti F., Baroni F., Formigli L., Zurdo J., Taddei N., Ramponi G., Dobson C.M., and Stefani M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* 416, 507-511.
- Bukau B., Weissman J., and Horwich A. (2006) Molecular chaperones and protein quality control. *Cell* 125, 443-451.
- Calamai M., Taddei N., Stefani M., Ramponi G., and 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 K.W., Ramponi G., Stefani M., and Chiti F. (2003) Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. *J Mol Biol* 330, 577-591.
- Camici G., Manao G., Cappugi G., and Ramponi G. (1976) A new synthesis of benzoyl phosphate: a substrate for acyl phosphatase assay. *Experientia* 32, 535-536.
- Capaldi A.P., Kleantous C., and Radford S.E. (2002) Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* 9, 209-216.
- Chelli R., Gervasio F.L., Procacci P., and Schettino V. (2002) Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *J Am Chem Soc* 124, 6133-6143.
- Chi C.N., Gianni S., Calosci N., Travaglini-Allocatelli C., Engstrom K., and Jemth P. (2007) A conserved folding mechanism for PDZ domains. *FEBS Lett* 581, 1109-1113.
- Chiarugi P., Degl'Innocenti D., Raugei G., Fiaschi T., and Ramponi G. (1997) Differential migration of acylphosphatase isoenzymes from cytoplasm to nucleus during apoptotic cell death. *Biochem Biophys Res Commun* 231, 717-721.
- Chiti F., Calamai M., Taddei N., Stefani M., Ramponi G., and Dobson C.M. (2002a) Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc Natl Acad Sci US A* 99, 16419-16426.
- Chiti F., and Dobson C.M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75, 333-366.
- Chiti F., Stefani M., Taddei N., Ramponi G., and Dobson C.M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805-808.
- Chiti F., Taddei N., Baroni F., Capanni C., Stefani M., Ramponi G., and Dobson C.M. (2002b) Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol* 9, 137-143.
- Chiti F., Taddei N., Bucciantini M., White P., Ramponi G., and Dobson C.M. (2000) Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J* 19, 1441-1449.
- Chiti F., Taddei N., Giannoni E., van Nuland N.A.J., Ramponi G., and Dobson C.M. (1999a) Development of enzymatic activity during protein folding. Detection of a spectroscopically silent native-like intermediate of muscle acylphosphatase. *J Biol Chem* 274, 20151-20158.
- Chiti F., Taddei N., White P.M., Bucciantini M., Magherini F., Stefani M., and Dobson C.M. (1999b) Mutational analysis of acylphosphatase suggests the im-

- portance of topology and contact order in protein folding. *Nat Struct Biol* 6, 1005-1009.
- Chiti F., van Nuland N.A.J., Taddei N., Magherini F., Stefani M., Ramponi G., and Dobson C.M. (1998) Conformational stability of muscle acylphosphatase: the role of temperature, denaturant concentration, and pH. *Biochemistry* 37, 1447-1455.
- Chiti F., Webster P., Taddei N., Clark A., Stefani M., Ramponi G., and Dobson C.M. (1999c) Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci* 96, 3590-3594.
- Chou P.Y., and Fasman G.D. (1974) Conformational parameters for amino acids in helical,  $\beta$ -sheet, and random coil regions calculated from proteins. *Biochemistry* 13, 211-222.
- Chow M.K., Ellisdon A.M., Cabrita L.D., and Bottomley S.P. (2004) Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. *J Biol Chem* 279, 47643-47651.
- Ciani B., Hutchinson E.G., Sessions R.B., and Woolfson D.N. (2002) A designed system for assessing how sequence affects  $\alpha$  to  $\beta$  conformational transitions in proteins. *J Biol Chem* 277, 10150-10155.
- Clarke J., Hamill S.J., and Johnson C.M. (1997) Folding and stability of a fibronectin type III domain of human tenascin. *J Mol Biol* 270, 771-778.
- Conway K.A., Harper J.D., and Lansbury P., T. Jr. (2000) Fibrils formed in vitro from  $\alpha$ -synuclein and two mutant forms linked to Parkinson's disease are typical amyloid. *Biochemistry* 39, 2552-2563.
- Corazza A., Rosano C., Pagano K., Alverdi V., Esposito G., Capanni C., Bemporad F., Plakoutsi G., Stefani M., Chiti F., Zuccotti S., Bolognesi M., and Viglino P. (2006) Structure, Conformational Stability, and Enzymatic Properties of Acylphosphatase From the Hyperthermophile *Sulfolobus solfataricus*. *Proteins* 62, 64-79.
- Creighton T.E. (1993) Physical Interactions That Determine the Properties of Proteins. In: *Proteins: Structure and molecular properties*, Freeman W. H. and Co., New York, pp. 139-170.
- Daggett V., and Fersht A.R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem Sci* 28, 18-25.
- Dill K.A. (1990) Dominant forces in protein folding. *Biochemistry* 29, 7133-7155.
- Dill K.A. (1999) Polymer principles and protein folding. *Protein Sci* 8, 1166-1180.
- Dill K.A., Ozkan S.B., Weikl T.R., Chodera J.D., and Voelz V.A. (2007) The protein folding problem: when will it be solved? *Curr Opin Struct Biol* 17, 342-346.
- Dinner A.R., Sali A., Smith L.J., Dobson C.M., and Karplus M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* 25, 331-339.
- Dobson C.M. (2003) Protein folding and misfolding. *Nature* 426, 884-890.
- Dobson C.M. (2004) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* 15, 3-16.
- Du H.N., Tang L., Luo X.Y., Li H.T., Hu J., Zhou J.W., and Hu H.Y. (2003) A peptide motif consisting of glycine, alanine, and valine is required for the fibrillization and cytotoxicity of human  $\alpha$ -synuclein. *Biochemistry* 42, 8870-8878.

- DuBay K.F., Pawar A.P., Chiti F., Zurdo J., Dobson C.M., and Vendruscolo M. (2004) Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J mol Biol* 341, 1317-1326.
- Dunker A.K., Cortese M.S., Romero P., Iakoucheva L.M., and Uversky V.N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* 272, 5129-5148.
- Dunker A.K., Lawson J.D., Brown C.J., Williams R.M., Romero P., Oh J.S., Oldfield C.J., Campen A.M., Ratliff C.M., Hipps K.W., Ausio J., Nissen M.S., Reeves R., Kang C., Kissinger C.R., Bailey R.W., Griswold M.D., Chiu W., Garner E.C., and Obradovic Z. (2001) Intrinsically disordered protein. *J Mol Graph Model* 19, 26-59.
- Dwyer M.A., Looger L.L., and Hellinga H.W. (2004) Computational design of a biologically active enzyme. *Science* 304, 1967-1971.
- Editorial. (2005) So much more to know. *Science* 309, 78-102.
- Eisenmesser E.Z., Bosco D.A., Akke M., and Kern D. (2002) Enzyme dynamics during catalysis. *Science* 295, 1520-1523.
- El-Agnaf O.M., and Irvine G.B. (2002) Aggregation and neurotoxicity of  $\alpha$ -synuclein and related peptides. *Biochem Soc Trans* 30, 559-565.
- Fernandez-Escamilla A.M., Rousseau F., Schymkowitz J., and Serrano L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 22, 1302-1306.
- Fersht A.R. (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci* 92, 10869-10873.
- Fersht A.R. (1997) Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* 7, 3-9.
- Fink A.L. (2005) Natively unfolded proteins. *Curr Opin Struct Biol* 15, 35-41.
- Garcia-Viloca M., Gao J., Karplus M., and Truhlar D.G. (2004) How enzymes work: analysis by modern rate theory and computer simulations. *Science* 303, 186-195.
- Gazit E. (2002) A possible role for  $\pi$ -stacking in the self-assembly of amyloid fibrils. *FASEB J* 16, 77-83.
- Gianni S., Geierhaas C.D., Calosci N., Jemth P., Vuister G.W., Travaglini-Allocatelli C., Vendruscolo M., and Brunori M. (2007a) A PDZ domain recapitulates a unifying mechanism for protein folding. *Proc Natl Acad Sci U S A* 104, 128-133.
- Gianni S., Guydosh N.R., Khan F., Caldas T.D., Mayor U., White G.W., DeMarco M.L., Daggett V., and Fersht A.R. (2003) Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A* 100, 13286-13291.
- Gianni S., Ivarsson Y., Jemth P., Brunori M., and Travaglini-Allocatelli C. (2007b) Identification and characterization of protein folding intermediates. *Proc Natl Acad Sci U S A* 128, 105-113.
- Giannoni E., Cirri P., Paoli P., Fiaschi T., Camici G., Manao G., Raugei G., and Ramponi G. (2000) Acylphosphatase is a strong apoptosis inducer in HeLa cell line. *Mol Cell Biol Res Commun* 3, 264-270.

- Giasson B.I., Murray I.V.J., Trojanovski J.Q., and Lee V.M.Y. (2001) A hydrophobic stretch of 12 residues in the middle of  $\alpha$ -synuclein is essential for filament assembly. *J Biol Chem* 276, 2380-2386.
- Gill S.C., and von Hippel P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182, 319-326.
- Gosal W.S., Morten I.J., Hewitt E.W., Smith D.A., Thomson N.H., and Radford S.E. (2005) Competing pathways determine fibril morphology in the self-assembly of  $\beta$ 2-microglobulin into amyloid. *J Mol Biol* 351, 850-864.
- Gsponer J., Hopearuoho H., Whittaker S.B., Spence G.R., Moore G.R., Paci E., Radford S.E., and Vendruscolo M. (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc Natl Acad Sci U S A* 103, 99-104.
- Guijarro J.I., Sunde M., Jones J.A., Campbell I.D., and Dobson C.M. (1998) Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* 95, 4224-4228.
- Humphrey W., Dalke A., and Schulten K. (1996) VMD - Visual Molecular Dynamics. *J Molec Graphics* 14, 33-38.
- Ivanova M.I., Gingery M., Whitson L.J., and Eisenberg D. (2003) Role of the C-terminal 28 residues of  $\beta$ 2-microglobulin in amyloid fibril formation. *Biochemistry* 25, 13536-13540.
- Jackson S.E., and Fersht A.R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30, 10428-10435.
- Jahn T.R., and Radford S.E. (2005) The Yin and Yang of protein folding. *FEBS J* 272, 5962-5970.
- Jimenez J.L., Guijarro J.I., Orlova E., Zurdo J., Dobson C.M., Sunde M., and Saibil H.R. (1999) Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J* 18, 815-821.
- Kajava A.V., Aebi U., and Steven A.C. (2005) The parallel superpleated  $\beta$ -structure as a model for amyloid fibrils of human amylin. *J Mol Biol* 348, 247-252.
- Kallberg Y., Gustafsson M., Persson B., Thyberg J., and Johansson J. (2001) Prediction of amyloid fibril-forming proteins. *J Biol Chem* 276, 12945-12950.
- Kaplan J., and DeGrado W.F. (2004) De novo design of catalytic proteins. *Proc Natl Acad Sci U S A* 101, 11566-11570.
- Karplus M., and Weaver D.L. (1994) Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* 3, 650-668.
- Kayed R., Sokolov Y., Edmonds B., McIntire T.M., Milton S.C., Hall J.E., and Glabe C.G. (2004) Permeabilization of lipid bilayers is a common conformation-dependent activity of soluble amyloid oligomers in protein misfolding diseases. *J Biol Chem* 279, 46363-46366.
- Khurana M., Uversky V.N., Nielsen L., and Fink A.L. (2001) Is Congo red an amyloid-specific dye? *J Biol Chem* 276, 22715-22721.
- Kiefhaber T., Schmid F.X., Willaert K., Engelborghs Y., and Chaffotte A. (1992) Structure of a rapidly intermediate in ribonuclease T1 folding. *Protein Sci* 1, 1162-1172.
- Klunk W.E., Pettegrew J.W., and Abraham D.J. (1989) Quantitative evaluation of congo red binding to amyloid-like proteins with a  $\beta$ -pleated sheet conformation. *J Histochem Cytochem* 37, 1273-1281.

- Koshland D.E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* 44, 98-104.
- Krebs M.R.H., Bromley E.H.C., and Donald A.M. (2005) The binding of thioflavin-T to amyloid fibrils: localisation and implications. *J Struct Biol* 149, 30-37.
- Krishnan R., and Lindquist S.L. (2005) Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* 435, 765-772.
- LeVine III H. (1995) Thioflavin T interaction with amyloid  $\beta$ -sheet structures. *Amyloid: Int J Exp Clin Invest* 2, 1-6.
- Levinthal C. 1969. How to fold graciously, pp. 22-24. University of Illinois press.
- Lindberg M.O., and Oliveberg M. (2007) Malleability of protein folding pathways: a simple reason for complex behaviour. *Curr Opin Struct Biol* 17, 21-29.
- Lindorff-Larsen K., B. B., R., DePristo M.A., Dobson C.M., and Vendruscolo M. (2005a) Simultaneous determination of protein structure and dynamics. *Nature* 433, 128-132.
- Lindorff-Larsen K., R\o gen P., Paci E., Vendruscolo M., and Dobson C.M. (2005b) Protein folding and the organization of the protein topology universe. *Trends Biochem Sci* 30, 13-19.
- Lopez de la Paz M., and Serrano L. (2004) Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci U S A* 101, 87-92.
- MacKerell A.D., Bashford D., Bellott M., Dunbrack R.L., Evanseck J.D., Field M.J., Fischer S., Gao J., Guo H., Ha S., Joseph-McCarthy D., Kuchnir L., Kuczera K., Lau F.T.K., Mattos C., Michnick S., Ngo T., Nguyen D.T., Prodhom B., Reiher W.E., Roux B., Schlenkrich M., Smith J.C., Stote R., Straub J., Watanabe M., Wiorkiewicz-Kuczera J., Yin D., and Karplus M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102, 3586-3616.
- Magg C., Kubelka J., Holtermann G., Haas E., and Schmid F.X. (2006) Specificity of the initial collapse in the folding of the cold shock protein. *J Mol Biol* 360, 1067-1080.
- Makin O.S., Atkins E., Sikorski P., Johansson J., and Serpell L.C. (2005) Molecular basis for amyloid fibril formation and stability. *Proc Natl Acad Sci U S A* 102, 315-320.
- Marcon G., Plakoutsi G., Canale C., Relini A., Taddei N., Dobson C.M., Ramponi G., and 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.
- Matagne A., and Dobson C.M. (1998) The folding process of hen lysozyme: a perspective from the 'new view'. *Cell Mol Life Sci* 54, 363-371.
- Matouschek A., and Fersht A.R. (1991) Protein engineering in analysis of protein folding pathways and stability. *Methods Enzymol* 202, 82-112.
- Matouschek A., Kellis Jr J.T., Serrano L., Bycroft M., and Fersht A.R. (1989a) Transient folding intermediates characterized by protein engineering. *Nature* 346, 440-445.
- Matouschek A., Kellis Jr J.T., Serrano L., and Fersht A.R. (1989b) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340, 122-126.

- Matouschek A., Serrano L., and Fersht A.R. (1992) The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 224, 819-835.
- Maxwell K.L., Wildes D., Zarrine-Afsar A., De Los Rios M.A., Brown A.G., Friel C.T., Hedberg L., Horng J.C., Bona D., Miller E.J., Vallee-Belisle A., Main E.R., Bemporad F., Qiu L., Teilum K., Vu N.D., Edwards A.M., Ruczinski I., Poulsen F.M., Kragelund B.B., Michnick S.W., Chiti F., Bai Y., Hagen J., Serrano L., Oliveberg M., Raleigh D.P., Wittung-Stafshede P., Radford S.E., Jackson S.E., Sosnick T.R., Marqusee S., Davidson A.R., and Plaxco K.W. (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* 14, 602-616.
- Mayor U., Guydosh N.R., Johnson C.M., Grossmann J.G.S., S., Jas G.S., Freund S.M., Alonso D.O., Daggett V., and Fersht A.R. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* 421, 863-867.
- Mayor U., Johnson C.M., Daggett V., and Fersht A.R. (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci U S A* 97, 13518-13522.
- Miyazono K., Sawano Y., and Tanokura M. (2005) Crystal structure and structural stability of acylphosphatase from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. *Proteins* 61, 196-205.
- Modesti A., Taddei N., Bucciantini M., Stefani M., Colombini B., Raugei G., and Ramponi G. (1995) Expression, purification, and characterization of acylphosphatase muscular isoenzyme as fusion protein with glutathione S-transferase. *Protein Expr Purif* 6, 799-805.
- Monsellier E., and Chiti F. (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* 8, 737-742.
- Monsellier E., Ramazzotti M., de Laureto P.P., Tartaglia G.G., Taddei N., Fontana G., Vendruscolo M., and Chiti F. (2007) The distribution of residues in a polypeptide sequence is a determinant of aggregation optimized by evolution. *Biophys J* 93, 4382-4391.
- Monti M., Garolla di Bard B.L., Calloni G., Chiti F., Amoresano A., Ramponi G., and 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.
- Moult J. (2006) Rigorous performance evaluation in protein structure modelling and implications for computational biology. *Philos Trans R Soc Lond B Biol Sci* 361, 453-458.
- Myers J.K., and Oas T.G. (2001) Preorganized secondary structure as an important determinant of fast protein folding. *Nat Struct Biol* 8, 552-558.
- Nagao C., Terada T.P., Yomo T., and Sasai M. (2005) Correlation between evolutionary structural development and protein folding. *Proc Natl Acad Sci U S A* 102, 18950-18955.
- Nakayama K.I., Hatakeyama S., and Nakayama K. (2001) Regulation of the cell cycle at the G1-S transition by proteolysis of cyclin E and p27Kip1. *Biochem Biophys Res Commun* 282, 853-860.

- Nediani C., Fiorillo C., Marchetti E., Pacini A., Liguri G., and Nassi P. (1996) Stimulation of cardiac sarcoplasmic reticulum calcium pump by acylphosphatase. Relationship to phospholamban phosphorylation. *J Biol Chem* 271, 19066-19073.
- Nelson R., Sawaya M.R., Balbirnie M., Madsen A.O., Riek C., Grothe R., and Eisenberg D. (2005) Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature* 435, 773-778.
- Nilsson M.R. (2004) Techniques to study amyloid fibril formation in vitro. *Methods* 34, 151-160.
- Nolting B., and Andert K. (2000) Mechanism of protein folding. *Proteins* 41, 288-298.
- Osborne M.J., Schnell J., Benkovic S.J., Dyson H.J., and Wright P.E. (2001) Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism. *Biochemistry* 40, 9846-9859.
- Otzen D.E., Itzhaki L.S., elMasry N.F., and Jackson S.E.F., A. R. (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc Natl Acad Sci U S A* 91, 10422-10425.
- Otzen D.E., Kristensen O., and Oliveberg M. (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. *Proc Natl Acad Sci U S A* 97, 9907-9912.
- Pace C.N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131, 266-280.
- Pagano K., Ramazzotti M., Viglino P., Esposito G., Degl'Innocenti D., Taddei N., and Corazza A. (2006) NMR solution structure of the acylphosphatase from *Escherichia coli*. *J Biomol NMR* 36, 199-204.
- Paoli P., Camici G., Manao G., Giannoni E., and Ramponi G. (2000) Acylphosphatase possesses nucleoside triphosphatase and nucleoside diphosphatase activities. *Biochem J* 349, 43-49.
- Park H.S., Nam S.H., Lee J.K., Yoon C.N., Mannervik B., Benkovic S.J., and Kim H.S. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 311, 535-538.
- Pastore A., Saudek V., Ramponi G., and Williams R.J.P. (1992) Three-dimensional structure of acylphosphatase. Refinement and structural analysis. *J Mol Biol* 224, 427-440.
- Pawar A.P., Dubay K.F., Zurdo J., Chiti F., Vendruscolo M., and Dobson C.M. (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol* 350, 379-392.
- Pedersen J.S., Christensen G., and Otzen D.E. (2004) Modulation of S6 fibrillation by unfolding rates and gatekeeper residues. *J Mol Biol* 341, 575-588.
- Pervushin K., Vamvaca K., Vogeli B., and Hilvert D. (2007) Structure and dynamics of a molten globular enzyme. *Nat Struct Mol Biol* 14, 1202-1206.
- Petkova A.T., Ishii Y., Balbach J.J., Antzoukin O.N., Leapman R.D., and Delaglio F. (2002) A structural model for Alzheimer's  $\beta$ -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* 99, 16742-16747.

- Petkova A.T., Yau W.M., and Tycko R. (2006) Experimental constraints on quaternary structure in Alzheimer's  $\beta$ -amyloid fibrils. *Biochemistry* 45, 498-512.
- Plakoutsi G., Bemporad F., Calamai M., Taddei N., Dobson C.M., and Chiti F. (2005) Evidence for a mechanism of amyloid formation involving molecular reorganisation within native-like precursor aggregates. *J Mol Biol* 351, 910-922.
- Plakoutsi G., Bemporad F., Monti M., Pagnozzi D., Pucci P., and Chiti F. (2006) Exploring the mechanism of formation of native-like and precursor amyloid oligomers for the native acylphosphatase from *Sulfolobus solfataricus*. *Structure* 14, 993-1001.
- Plakoutsi G., Taddei N., Stefani M., and Chiti F. (2004) Aggregation of the Acylphosphatase from *Sulfolobus solfataricus*: the folded and partially unfolded states can both be precursors for amyloid formation. *J Biol Chem* 279, 14111-14119.
- Plaxco K.W., Simons K.T., and Baker D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277, 985-994.
- Porat Y., Mazor Y., Efrat S., and Gazit E. (2004) Inhibition of islet amyloid polypeptide fibril formation: A potential role for heteroaromatic interactions. *Biochemistry* 43, 14454-14462.
- Porat Y., Stepensky A., Ding F., Naider F., and Gazit E. (2003) Completely different amyloidogenic potential of nearly identical peptide fragments. *Biopolymers* 69, 161-164.
- Poulsen T.D., Garcia-Viloca M., Gao J., and Truhlar D.G. (2003) Free Energy Surface, Reaction Paths, and Kinetic Isotope Effect of Short-Chain Acyl-CoA Dehydrogenase. *J Phys Chem B* 107, 9567-9578.
- Quintas A., Vaz D.C., Cardoso I., Saraiva M.J., and Brito R.M. (2001) Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J Biol Chem* 276, 27207-27213.
- Ramponi G. (1975) 1,3-diphosphoglycerate phosphatase. *Methods Enzymol* 42, 409-426.
- Ramponi G., Treves C., and Guerritore A. (1966) Aromatic acyl phosphates as substrates of acyl phosphatase. *Arch Biochem Biophys* 115, 129-135.
- Religa T.L., Markson J.S., Mayor U., Freund S.M., and Fersht A.R. (2005) Solution structure of a protein denatured state and folding intermediate. *Nature* 437, 1053-1056.
- Richardson J.S., and Richardson D.C. (2002) Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci U S A* 99, 2754-2759.
- Ritter C., Maddelein M., Siemer A.B., Luhrs T., Ernst M., Meier B.H., Saupe S.J., and Riek R. (2005) Correlation of structural elements and infectivity of the HET-s prion. *Nature* 435, 844-848.
- Roder H., Elove G.A., and Englander S.W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335, 700-704.
- Saiki M., Honda S., Kawasaki K., Zhou D., Kaito A., Konakahara T., and Morii H. (2005) Higher-order molecular packing in amyloid-like fibrils constructed

- with linear arrangements of hydrophobic and hydrogen-bonding side-chains. *J Mol Biol* 348, 983-998.
- Salvatella X., Dobson C.M., Fersht A.R., and Vendruscolo M. (2005) Determination of the folding transition states of barnase by using PhiI-value-restrained simulations validated by double mutant PhiIJ-values. *Proc Natl Acad Sci U S A* 102, 12389-12394.
- Santoro M.M., and Bolen D.W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl  $\alpha$ -chymotrypsin using different denaturants. *Biochemistry* 27, 8063-8068.
- Sawaya M.R., Sambashivan S., Nelson R., Ivanova M.I., Sievers S.A., Apostol M.I., Thompson M.J., Balbirnie M., Wiltzius J.J., McFarlane H.T., Ø. M., A., Riek C., and Eisenberg D. (2007) Atomic structures of amyloid cross- $\beta$  spines reveal varied steric zippers. *Nature* 447, 453-457.
- Schuler B., Lipman E.A., and Eaton W.A. (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419, 743-747.
- Serio T.R., Cashikar A.G., Kowal A.S., Sawicki G.J., Moslehi J.J., Serpell L., Arnsdorf M.F., and Lindquist S.L. (2000) Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* 289, 1317-1321.
- Serpell L.C. (2000) Alzheimer's amyloid fibrils: structure and assembly. *Biochim Biophys Acta* 1502, 16-30.
- Serpell L.C., Sunde M., Benson M.D., Tennent G.A., Pepys M.B., and P.E. F. (2000) The protofilament substructure of amyloid fibrils. *J Mol Biol* 300, 1033-1039.
- Shastry M.C., Luck S.D., and Roder H. (1998) A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. *Biophys J* 74, 2714-2721.
- Soldi G., Bemporad F., Torrasa S., Relini A., Ramazzotti M., Taddei N., and Chiti F. (2006a) Amyloid formation of a protein in the absence of initial unfolding and destabilization of the native state. *Biophys J* 89, 4234-4244.
- Soldi G., Plakoutsi G., Taddei N., and Chiti F. (2006b) Stabilization of a native protein mediated by ligand binding inhibits amyloid formation independently of the aggregation pathway. *J Med Chem* 27, 6057-6064.
- Stefani M., and Dobson C.M. (2003) Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* 81, 678-699.
- Stefani M., Taddei N., and Ramponi G. (1997) Insights into acylphosphatase structure and catalytic mechanism. *Cell Mol Life Sci* 53, 141-151.
- Street A.G., and Mayo S.L. (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.
- Sun Y., Breydo L., Makarava N., Yang Q., Bocharova O.V., and Baskakov I.V. (2007) Site-specific conformational studies of prion protein (PrP) amyloid fibrils revealed two cooperative folding domains within amyloid structure. *J Biol Chem* 282, 9090-9097.
- Sunde M., and Blake C.C.F. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem* 50, 123-159.

- Swindells M.B.O., C. A., Jones D.T., Pearl L.H., and Thornton J.M. (1993) Recurrence of a binding motif? *Nature* 362, 299.
- Taddei N., Chiti F., Fiaschi T., Bucciantini M., Capanni C., Stefani M., Serrano L., Dobson C.M., and Ramponi G. (2000) Stabilisation of  $\alpha$ -helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J Mol Biol* 14, 633-647.
- Taddei N., Chiti F., Magherini F., Stefani M., Thunnissen M.M., Nordlund P., and Ramponi G. (1997) Structural and kinetic investigations on the 15-21 and 42-45 loops of muscle acylphosphatase: evidence for their involvement in enzyme catalysis and conformational stabilization. *Biochemistry* 36, 7217-7224.
- Taddei N., Chiti F., Paoli P., Fiaschi T., Bucciantini M., Stefani M., Dobson C.M., and Ramponi G. (1999) Thermodynamics and kinetics of folding of common-type acylphosphatase: comparison to the highly homologous muscle isoenzyme. *Biochemistry* 38, 2135-2142.
- Taddei N., Stefani M., Magherini F., Chiti F., Modesti A., Raugei G., and Ramponi G. (1996) Looking for residues involved in the muscle acylphosphatase catalytic mechanism and structural stabilization: role of Asn41, Thr42, and Thr46. *Biochemistry* 35, 7077-7083.
- Taddei N., Stefani M., Vecchi M., Modesti A., Raugei G., Bucciantini M., Magherini F., and Ramponi G. (1994) Arginine-23 is involved in the catalytic site of muscle acylphosphatase. *Biochim Bioophys Acta* 1208, 75-80.
- Tartaglia G., Cavalli A., Pellarin P., and Caflich A. (2004) The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci* 13, 1939-1941.
- Thomas P.J., Qu B.H., and Pedersen P.L. (1995) Defective protein folding as a basis of human disease. *Trends Biochem Sci* 20, 456-459.
- Thunnissen M.M., Taddei N., Liguri G., Ramponi G., and Nordlund P. (1997) Crystal structure of common type acylphosphatase from bovine testis. *Structure* 5, 69-79.
- Tjernberg L., Hsiao W., Bark N., Thyberg J., and Johansson J. (2002) Charge attraction and  $\beta$  propensity are necessary for amyloid fibril formation from tetrapeptides. *J Biol Chem* 277, 43243-43246.
- Tjernberg L.O., Callaway D.J.E., Tjernberg A., Hahne S., Lilliehook C., Terenius L., Thyberg J., and Nordstedt C. (1999) A molecular model of Alzheimer amyloid  $\beta$ -peptide fibril formation. *J Biol Chem* 274, 12619-12625.
- Torok M., Milton S., Kaye R., Wu P., McIntire T., Glabe C.G., and Langen R. (2002) Structural and dynamic features of Alzheimer's A $\beta$  peptide in amyloid fibrils studied by site-directed spin labeling. *J Biol Chem* 277, 40810-40815.
- Tracz S.M., Abedini A., Driscoll M., and Raleigh D.P. (2004) Role of aromatic interactions in amyloid formation by peptides derived from human amylin. *Biochemistry* 43, 15901-15908.
- Travaglini-Allocatelli C., Gianni S., and Brunori M. (2004) A common folding mechanism in the cytochrome c family. *Trends Biochem Sci* 29, 535-541.
- Tycko R. (2004) Progress towards a molecular-level structural understanding of amyloid fibrils. *Curr Opin Struct Biol* 14, 96-103.

- Uversky V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11, 739-756.
- Uversky V.N., Kutysenko V.P., Protasova N., Rogov V.V., Vassilenko K.S., and Gudkov A.T. (1996) Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. *Protein Sci* 5, 1844-1851.
- Vamvaca K., Vogeli B., Kast P., Pervushin P., and Hilvert D. (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proc Natl Acad Sci U S A* 101, 12860-12864.
- van Nuland N.A., Chiti F., Taddei N., Raugei G., Ramponi G., and Dobson C.M. (1998) Slow folding of muscle acylphosphatase in the absence of intermediates. *J Mol Biol* 283, 883-891.
- Vendruscolo M., Paci E., Dobson C.M., and Karplus M. (2001) Three key residues form a critical contact network in a protein folding transition state. *Nature* 409, 641-645.
- Villegas M., Martinez J.C., Aviles F.X., and Serrano L. (1998) Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J Mol Biol* 283, 1027-1036.
- Villegas V., Zurdo J., Filimonov V.V., Aviles F.X., Dobson C.M., and Serrano L. (2000) Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci* 9, 1700-1708.
- Walsh D.M., Hartley D.M., Kusumoto Y., Fezoui Y., Condron M.M., Lomakin A., Benedek G.B., Selkoe D.J., and Teplow D.B. (1999) Amyloid  $\beta$ -protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *J Biol Chem* 274, 25945-25952.
- West M.W., Wang W., Patterson J., Mancias J.D., Beasley J.R., and Hecht M. (1999) De novo amyloid proteins from designed combinatorial libraries. *Proc Natl Acad Sci U S A* 96, 11211-11216.
- Wetlaufer D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* 70, 697-701.
- White G.W., Gianni S., Grossmann J.G., Jemth P., Fersht A.R., and Daggett V. (2005) Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J Mol Biol* 350, 757-775.
- Williams A.D., Portelius E., Kheterpal I., Guo J.T., Cook K.D., Xu Y., and Wetzel R. (2004) Mapping A $\beta$  amyloid fibril secondary structure using scanning proline mutagenesis. *J Mol Biol* 335, 833-842.
- Wu C., Lei H., and Duan Y. (2005) The role of Phe in the formation of well-ordered oligomers of amyloidogenic hexapeptide (NFGAIL) observed in molecular dynamics simulations with explicit solvent. *Biophys J* 88, 2897-2906.
- Wurth C., Guimard N.K., and Hecht M.H. (2002) Mutations that reduce aggregation of the Alzheimer's A $\beta$ 42 peptide: An unbiased search for the sequence determinants of A $\beta$  amyloidogenesis. *J Mol Biol* 319, 1279-1290.
- Zagrovic B., Snow C.D., Shirts M.R., and Pande V.S. (2002) Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J Mol Biol* 323, 927-937.

- Zanuy D., Porat Y., Gazit E., and Nussinov R. (2004) Peptide sequence and amyloid formation; molecular simulations and experimental study of a human islet amyloid polypeptide fragment and its analogs. *Structure* 12, 439-455.
- Zarrine-Afsar A., and Davidson A.R. (2004) The analysis of protein folding kinetic data produced in protein engineering experiments. *Methods* 34, 41-50.
- Zarrine-Afsar A., Larson S.M., and Davidson A.R. (2005) The family feud: do proteins with similar structures fold via the same pathway? *Curr Opin Struct Biol* 15:42-9, 42-49.
- Zhang Y., Arakaki A.K., and Skolnick, J. (2005) TASSER: an automated method for the prediction of protein tertiary structures in CASP6. *Proteins* 61 (suppl 7), 91-98.
- Zuccotti S., Rosano C., Ramazzotti M., Degl'Innocenti D., Stefani M., Manao G., and Bolognesi M. (2004) Three-dimensional structural characterization of a novel *Drosophila melanogaster* acylphosphatase. *Acta Crystallogr D Biol Crystallogr* 60, 1177-1179.



PREMIO FIRENZE UNIVERSITY PRESS  
TESI DI DOTTORATO

- Coppi E., *Purines as Transmitter Molecules. Electrophysiological Studies on Purinergic Signalling in Different Cell Systems*, 2007
- Natali I., *The Ur-Portrait. Stephen Hero ed il processo di creazione artistica in A Portrait of the Artist as a Young Man*, 2007
- Petretto L., *Imprenditore ed Università nello start-up di impresa. Ruoli e relazioni critiche*, 2007
- Mannini M., *Molecular Magnetic Materials on Solid Surfaces*, 2007
- Bracardi M., *La Materia e lo Spirito. Mario Ridolfi nel paesaggio umbro*, 2007
- Bemporad F., *Folding and aggregation studies in the acylphosphatase-like family*, 2008
- Buono A., *Esercito, istituzioni, territorio. Alloggiamenti militari e «case Herme» nello Stato di Milano (secoli XVI e XVII)*, 2008
- Castenasi S., *La finanza di progetto tra interesse pubblico e interessi privati*, 2008
- Gabbiani C., *Proteins as possible targets for antitumor metal complexes: biophysical studies of their interactions*, 2008
- Colica G., *Use of microorganisms in the removal of pollutants from the wastewater*, 2008

Finito di stampare presso  
Grafiche Cappelli Srl – Osmannoro (FI)