Oxidative Folding of a Cystine Knot Protein: the Amaranthus -amylase Inhibitor

Thesis

UNRESTRICTED.

Oxidative Folding of a Cystine Knot Protein: the Amaranthus α-amylase Inhibitor

a thesis submitted for
Ph. D. degree in Life Sciences
Open University, U.K.

Maša Čemažar

International Centre for Genetic Engineering and Biotechnology

June 2003

Submission date: 18 June 2003
Award date: 1 September 2003

ABSTRACT

Oxidative Folding of a Cystine Knot Protein: the Amaranthus α-amylase Inhibitor

Maša Čemažar

The work described in this thesis is concerned with the study of the oxidative folding pathway of a cystine knot protein, Amaranthus α-amylase inhibitor (AAI). The data presented in the thesis are obtained with a combination of different biochemical and biophysical techniques, with an emphasis on NMR spectroscopy.

Chapter 1 gives a short introduction to the structural and functional superfamilies of the protein in question. Additionally, it outlines the protein folding problem, with an emphasis on oxidative folding, and stresses the importance of studying such pathways in small proteins. Chapter 2 covers both the basis of experimental techniques and the details of experimental methods used in the study. The remaining four chapters (3, 4, 5 and 6) are results and discussion chapters. Each one deals with a specific aspect of the study of the oxidative folding pathway of Amaranthus α-amylase inhibitor.

Chapter 3 presents the results from an improved solid phase peptide synthesis approach, which increased the overall yield of peptide synthesis and recovered a greater amount of AAI after the refolding step. In Chapter 4 the intermediate species that appear on the oxidative folding pathway are identified and characterised in terms of their disulfide content. An illustration of the structural features of the main folding intermediate (MFI) with NMR, photo-CIDNP and computer modelling techniques is given in Chapter 5. In Chapter 6 the kinetic process of the oxidative folding of Amaranthus α-amylase inhibitor is studied at the level of a single disulfide species and at the level of a single amino acid residue.

The novelty of this thesis is two-fold. First, from the point of view of biological interest, this is the first time that oxidative folding intermediates with non-native disulfide bridges between adjacent cysteine residues are observed. Second, from the point of view of development of biophysical methods, the first application of time-resolved NMR and photo-CIDNP spectroscopy to a complete oxidative folding process is described.

ACKNOWLEDGEMENTS

I would like to express a sincere thanks to my supervisors, Sándor Pongor of ICGEB and Peter Hore of Oxford University. They have both given me never-ending help, advice and understanding throughout the time of my Ph.D.

I owe all my knowledge about peptides to Sotir Zakhariev, whose enthusiastic daily guidance is irreplaceable, as well as the stimulating and challenging discussions.

In Trieste, the members of the Protein Structure and Bioinformatics group have created a friendly, honest and dynamic atmosphere. Corrado Guarnaccia and Francesco Zanuttin have advised me on numerous experiments. Kristian Vlahoviček was always willing to untie any computer knots on my way. Maristella Coglievina, Oliviero Carugo, and Alessandro Pintar have helped me to learn the basics of molecular biology and bioinformatics. I would also like to thank ICGEB for awarding me a pre-doctoral fellowship and numerous possibilities to travel.

The trips to Oxford were filled with people eager to help and offer advice. I could not imagine my times in Oxford without Jakob Lopez and Christiane Timmel. Jonathan Jones and the NMR core group, consisting of Christina Redfield, Nick Soffe and Jonathan Boyd, have helped me on countless occasions with their suggestions and boundless knowledge about NMR. In the Hore group there are many others, who were there to assist me as I passed by, among them Iain Day and Ailsa Curtis. In the Oxford Center for Molecular Sciences, many members gave me a hand, especially Jorn Werner, Charles Blundell, Vicky Higman, Rachel Wain, Mark Krebs and Lesley Greene. I also acknowledge Chris Dobson for his interest in my work and enthusiastic suggestions.

I am indebted to the members of the National Chemistry Institute and Jožef Stefan Institute, especially Jurka Kidrič, Janez Plavec, Janez Mavri and Eva Žerovnik.

A big thanks go to my closest friends, from Kranj, Duino, Oxford and Trieste, who have made me believe that it does not matter how much time we can't see each other and that some things are just not going to change.

I am grateful to Genaro for his comforting and reassuring trust in me over the past three years.

Last, but not most certainly not least, I wish to thank my parents and my grandparents for giving me their unlimited love and for teaching me that everything comes at the right time. I would like to dedicate this thesis to them.

Contents

List of Figures

List of Tables

ABBREVIATIONS

Boc, tert-butyloxycarbonyl

BPTI, bovine pancreatic trypsin inhibitor

But, tert-butyl

Bzl, benzyl

CD, circular dichroism

Da, Dalton

DCM, dichloromethane

DMF, N,N-dimethylformamide

DMSO, dimethyl sulfoxide

DIEA, N,N-diisopropylethylamine

DMF, N,N-dimethylformamide

DMP, 2,6-dimethylpyrdine

3D, 2D, 1D, three-, two-, one-dimensional

EDT, 1,2-ethanedithiol

ESI-MS, electrospray ionization mass spectrometry

EDTA, ethylenediamine-tetraacetic acid

Fmoc, 9-fluorenylmethoxycarbonyl

FMN, Flavin Mononucleotide

NH^4OAc, ammonium acetate

GdnHCl, guanidinium hydrochloride

HBTU, O-benzotriazolyl-N,N,N',N'-tetramethyluronium hexafluorophosphate

HOBt, 1-hydroxybenzotriazole

RP-HPLC, reversed-phase high-performance liquid chromatography

MeCN, acetonitrile

MS, mass spectrometry

NMP, N-methylpyrrolidinone

NOESY, nuclear overhauser enhanced spectroscopy

OPfp, pentafuorophenyl ester

photo-CIDNP, photo-chemically induced dynamic nuclear polarisation

SPPS, solid-phase peptide synthesis

TCEP, tris(carboxyethyl)phosphine

TBTU, 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate

TFA, trifluoroacetic acid

TIS, triisopropylsilane

Trt, trityl

Θ, mean residue ellipticity

Chapter 1

Introduction

1.1 Amaranthus α-amylase inhibitor

The structure and function of a protein are like the genome and personality of a human being. The similarities of the structural and functional features to other proteins produce two fundamental classifications of a protein. The first classification relates the features of the protein's three-dimensional fold to other protein structures, whereas the second classification focusses on the biological activities and their similarities to other activities found in nature. However, neither of these classifications depends on the folding process, which converts the amino acid sequence into a specific fold. This is because protein folding is still relatively poorly understood and can not be a valid criterion in protein classification.

This thesis is concerned with the oxidative folding process of ***Amaranthus α-amylase inhibitor*** (AAI), which belongs to the cystine knot superfamily and to the plant inhibitors of insect α-amylase enzymes. In this chapter, a short introduction is given to the structural and functional superfamilies of the protein. Additionally, the

protein folding problem, with an emphasis on oxidative folding, is outlined with the historical developments that led to its current understanding. The prime reason for wanting to understand the structure and folding of AAI and other inhibitor cystine knots is their potential as scaffolds for bio-engineered pharmaceutical substances. A detailed picture of the oxidative process of AAI can give an example for studies of other pharmaceutically relevant cystine knot proteins.

1.1.1 The cystine knot motif

The cystine knot is an intriguing structural motif with an embedded ring made up of two disulfide bridges and their connecting backbone segments, penetrated by a third disulfide bond [1] (Figure 1.1). The scientific interest in cystine knots is probably owed both to their widespread distribution across organisms in nature (plants, insects, mammals) and to their extreme functional versatility that spreads from enzymes to ion channel inhibitors. The fact that these proteins fulfill a large variety of biological functions could lead to a conclusion that they have developed via a convergent evolution [2]. In other words, the evolutionary acquisition of similar structural features in unrelated proteins is a result of similar adaptations, rather than a common ancestor. The cystine knot is also a motif that provides a very high degree of stability to molecules that contain it and has been therefore considered as a suitable framework for protein engineering applications.

The term "cystine knot" was first introduced in reviews in 1993 [3, 4] comparing the structural similarities between recently determined structures of proteins belonging to the growth hormone family. However, throughout the early 1990's numerous protein structures were being published from other protein families containing the same

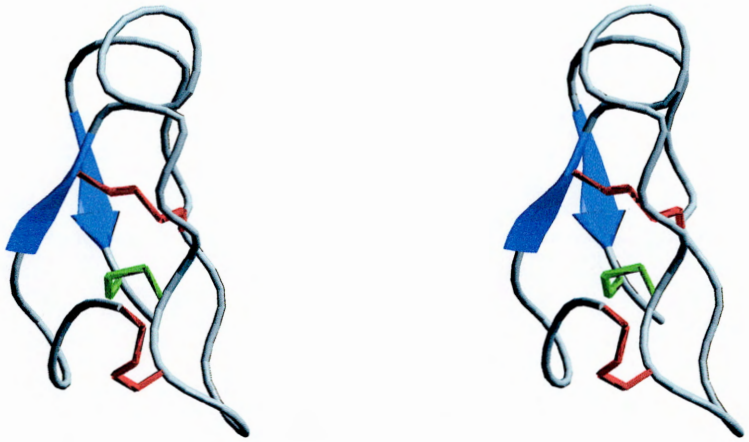

Figure 1.1: **Stereo-view of the cystine knot topology** in the *Amaranthus α-amylase inhibitor* molecule. The two disulfide bridges (I-IV, II-V) building a ring with their connecting backbone segments are coloured in red and the disulfide bridge (III-VI) that is penetrating the ring is shown in green.

structural motif, for example the disulfide rich trypsin inhibitors, cone snail toxins and even cystine knots with a circular protein backbone. This is the reason that the cystine knot superfamily was first divided into inhibitor cystine knots (ICK) and growth factor cystine knots (GFCK) [5, 6] and later an additional subfamily was added featuring proteins with a circular backbone [7], the cyclic cystine knots (CCK), present mainly in plant cyclotides. Numerous reviews have dealt with characterisation and classification of cystine knots [8, 9, 10, 11, 1].

The three cystine knot families share the same disulfide connectivity, which is usually referred to as the *abc abc* topology. This means that the first cysteine in protein sequence forms a disulfide bridge with the fourth (I-IV), the second with the fifth (II-V) and the third with the sixth cysteine (III-VI). The loop regions of the topological knot are remarkably conserved across the three families, which perhaps indicates the significance of the loop size for the stability of the fold. However, GFCKs have two distinguishable features from ICKs and CCKs: their bigger protein size and the disulfide bond which penetrates the embedded ring structure is (I-IV) instead of

(III-VI) in ICKs and CCKs.

The *Amaranthus α-amylase inhibitor* belongs to the inhibitor family of cystine knots along with other enzyme inhibitors, ion channel blockers and various toxin molecules. The inhibitor family, and partially also the other two families of cystine knots, have in addition to the I-IV, II-V, III-VI disulfide connectivity also consensus spacings between cysteine residues as can be seen in Figure 1.2 [1, 5, 11]. Other features of the ICK motif are a triple stranded anti-parallel β-sheet and several hydrogen bonding patterns. A summary of this motif is seen in Figure 1.2.

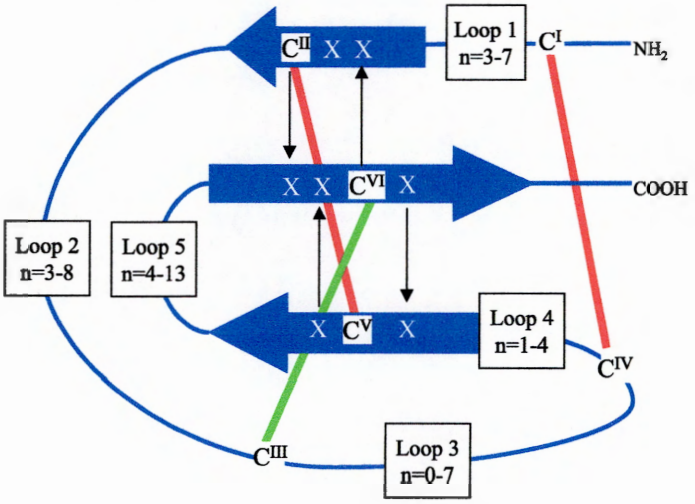

Figure 1.2: **Consensus features of the inhibitor cystine knot motif**: the cysteine residues are labelled I-VI, the intra-cysteine loops 1-5 and the number of intra-cysteine residues is n. The black arrows represent hydrogen bonds from the amide proton to the carbonyl group. As in Figure 1.1 the two disulfide bridges building a ring are shown in red and the third disulfide penetrating the ring is green. Adapted from a recent review by Craik et al [1].

The synthesis and folding of the inhibitor family of cystine knots have been extensively studied. The common features of the *in vitro* folding process that are displayed after successful solid phase peptide synthesis are the heterogeneous folding pathways and the presence of non-native like disulfide intermediates. One extremely important species for drug application in pain therapy, the conotoxin MVIIA [12], has shown that there is little or no preferential formation of native interactions until two

disulfides have formed.

On the other hand it is worth mentioning that there are two properties of cystine knots which make them promising molecules for drug design: exceptional stability and well-defined scaffolds. Most of the exciting applications of these proteins have been noted in pain therapy (MVIIA and related conotoxins specifically block ion channels involved in the transmission of pain), anti-bacterial activity (cyclotide peptides such as kalata B1, circulin A and B) and anti-viral agents (anti-HIV activity of cyclotides). However, there still remains plenty of room for developments in molecular bioengineering approaches.

1.1.2 α-Amylases and their inhibitors

The fact that only a limited number of organisms is able to feed on a given plant, proves that in plants, mechanisms against insect predation have developed through the evolution process. The defence compounds that provide the resistance of plants to insects can be either *aproteic* (antibiotics, alkaloids, terpenes, cyanogenic glucosides) or *proteic* (chitinases, β-1,3-glucanases, lectins, arcelins, vicilins, systemins and enzyme inhibitors) [13]. The enzyme inhibitors group provide a protective action against digestive hydrolases, proteinases and α-amylases.

Enzymes belonging to the α-amylase family are multidomain hydrolases and transferases with a catalytic domain in the form of a $(\beta/\alpha)_8$-barrel [15]. Although they are structurally and functionally related proteins, they have been separated into three families (13, 70, 77) of glycoside hydrolases according to the amino acid sequence similarity. Their function is to catalyse the hydrolysis or transglycosylation of α-glucan linkages in starch components, glycogen and other carbohydrates [13, 15].

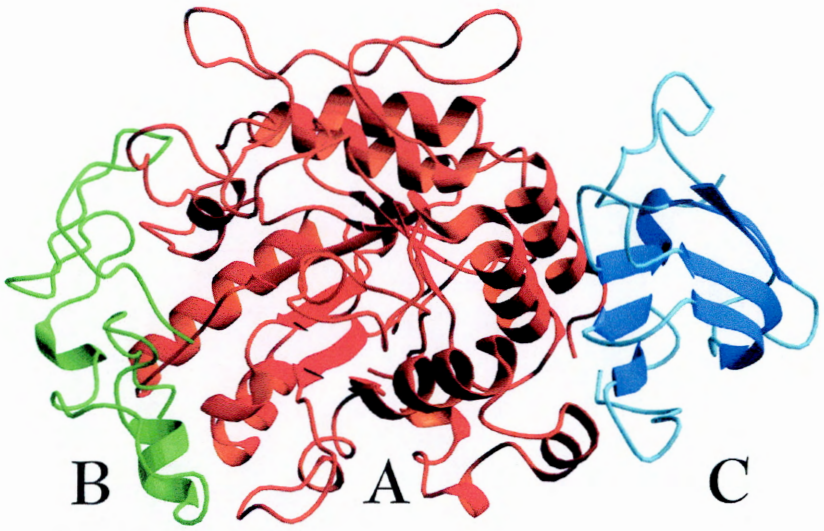

Figure 1.3: ***Tenebrio molitor* α-amylase**: the three distinctive domains A, B and C are coloured in red, green and blue, respectively. The major structural subunit is the A domain, which has the form of a $(\beta/\alpha)_8$-barrel and contains the catalytic site and the ligand binding residues. The figure was made using 1tmq file from the Protein Data Bank with MolMol programme [14].

The true α-amylases, however, act only on α-1,4-glucosidic linkages (α-1,4-glucan-4-glucanohydrolases, Enzyme Commission number 3.2.1.1). Generally they are endo-enzymes, but there exist exceptions with exo-activity, which is likely to be coming from the differences in the amino acid sequence of the glucose binding regions. Contrary to the exo-α-amylases, their endo-equivalents are able to catalyse more than one glucosidic bond breakage during a single enzyme-substrate encounter. They also have a preferential action near the chain ends and hence produce a greater proportion of certain oligosaccharides over others.

α-Amylases regulate the carbohydrate metabolism in microorganisms, plants and animals. In insects, these enzymes are essential for the survival of several species that feed on starchy seeds during their larval and adult stages. Although several sequences have been characterized, the only determined 3D structure of an insect α-amylase is that of the *Tenebrio molitor* enzyme (TMA) [16], represented in Figure 1.3. Domain A of TMA is a characteristic $(\beta/\alpha)_8$-barrel and contains the catalytic site and the ligand

binding residues. Furthermore, a distinctive calcium ion resides in the cavity between the catalytic domain A and the globular domain B. This feature is conserved among many α-amylases.

The proteinaceous inhibitors of α-amylases can be classified according to their tertiary structure into six classes: lectin-like, thaumatin-like, cereal-type, Kunitz-like, γ-purothionin-like and knottin-like [13, 17]. The structures and mechanisms of inhibition of the latter four types of inhibitors have been elucidated by resolving the 3D structures of the complexes between the inhibitors and the enzymes. In the cases of lectin-type, knottin-type and cereal-type inhibitors, the enzyme function is inhibited directly through blockage of the catalytic site, whereas the Kunitz-type inhibitors prevent substrate binding through interaction with both A and B domains of the enzymes near the catalytic site, but not directly with the catalytic residues. The interaction of thaumatin-like and γ-purothionin-like inhibitors with α-amylases has not yet been characterized, although some structures of inhibitors belonging to these two classes have been determined. Different inhibitors, however, have different specificities towards α-amylases from different sources, some for example are insect-specific and inactive towards their mammalian analogues. Including these kind of insect-specific inhibitors in transgenic plant approaches could provide inherent protection against insect pests.

1.1.3 The inhibitor from *Amaranthus hypocondriacus* seeds

The smallest known α-amylase inhibitor is found in the seeds of *Amaranthus hypocondriacus*, a variety of the Mexican crop plant Amaranth or Prince's feather. Although there are other species in Amaranthus plants that confer inhibitory activity, Amaranthus α-amylase inhibitor (AAI) accounts for more than half of the activity in crude seed

extract. The most important feature of its activity is that it is species specific as it inhibits only α-amylase from insect larvae (*Tribolium castaneum, Prostaphanus truncatus, Periplaneta americana* and *Tenebrio molitor*), but not the mammalian equivalents. The structure, function and enzyme binding activity of this inhibitor has been extensively studied.

Figure 1.4: **Amaranthus α-amylase inhibitor (AAI)**: The cysteine rich amino acid-sequence is shown with the six cysteine residues written in red. The disulfide bridge connectivity with the *abc abc* pattern is shown.

AAI was first isolated by Chagolla-Lopez et al [18] in 1994, when they determined its primary structure, disulfide bridges and built a three-dimensional model based on the common structural features of the inhibitor cystine knots. In 1997 a successful solid-phase peptide synthesis was reported [19], together with an *in vitro* oxidative refolding protocol. It was confirmed that the refolded synthetic peptide had identical biological and inhibitory properties as the natural product. The next steps were the three-dimensional structure determinations of both the free inhibitor in solution by NMR and of the complex with the enzyme by X-ray crystallography, which were both published in 1999 [20, 21]. An additional NMR structure of the free inhibitor with a comparison to the already published structure came out in 2001 [22]. A more detailed structural analysis of the three-dimensional structures and a classification of Amaranthus inhibitor within the knottin fold superfamily [2] followed in the same year.

The overall structure of this small 32-amino acid peptide, the three antiparallel β-strands and the typical *abc abc* connectivity of disulfides identify it as a member of

the inhibitor cystine knots (ICK) [20, 21, 2]. Among proteins in the Protein Data Bank that contain *abc abc* disulfide topology, AAI can be clustered together with 15 other proteins that have a high proline content and many of them contain a *cis*-proline [2]. The primary sequence of AAI (Figure 1.4) is strikingly different from other members of the ICK family, yet it maintains an almost identical scaffold. Similarly, the members of this family show few similarities in surface electrostatics and functionally important residues, which strengthens the theory about their convergent evolution [2].

When AAI binds to the enzyme, two important alterations can be seen in its 3D scaffold [2], which can be seen in its unaltered state in Figure 1.5. First, it becomes much more compact and hence decreases the solvent accessible surface area and its volume. Secondly, residue Pro20 undergoes a *trans* to *cis* isomerisation. AAI inserts perfectly into the active site of TMA (*Tenebrio molitor* α-amylase [21], Figure 1.3), which is situated at the interface between domains A and B. The complex sees a high complementarity of the interacting surfaces, since AAI provides a lid-like surface to the deep cleft of the active site in TMA. Upon binding, TMA undergoes a dramatic change in the backbone segment of the active site, shortening the helical segment from His286 to Gly292 by three residues at its C terminus.

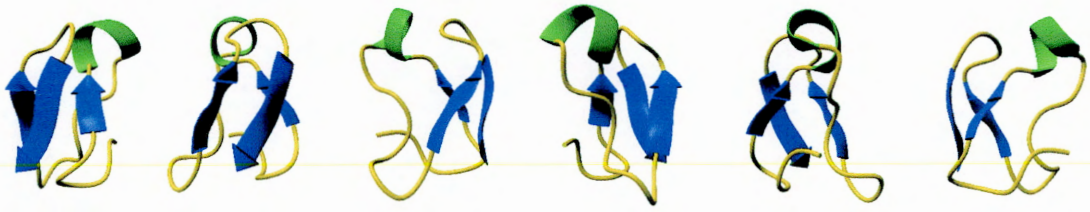

Figure 1.5: **Three-dimensional structure of AAI protein**: these six images are made with MolMol [14] programme, using *1qfd* file from the Protein Data Bank. Each consecutive image of the 3D AAI structure is rotated by 60° anti-clockwise around the vertical axis.

The next question to ask once a 3D structure of a protein has been determined is how the protein folds to this unique, low energy state. Why does this particular

sequence encode the specific structure with a particular function? What kind of intermediate species appear on the folding pathway as the molecule goes from a linear amino acid chain to a globular fold? It is important to determine how separate parts of the molecules are stabilised and when in the folding process the various native-like contacts and native like-interactions are made. The answers to these questions give information that is essential for redesigning molecules with bioengineering approaches.

1.2 Protein folding

1.2.1 The protein-folding problem(s)

After the synthesis of proteins, which come off the ribosome as linear chains of amino acids, they need to acquire the 3D structure of the specific function for which they have evolved. The process of attaining the unique fold is called protein folding. The detailed understanding of this process remains a challenge for two reasons: the secret code entails both the 3D fold and the ability to attain it.

The first problem with protein folding will be solved if we can predict how an amino acid sequence encodes a certain 3D array and hence a specific function. This is the subject of intense theoretical and experimental strategies. It has become clear that the interactions between the side-chains that constitute the 3D fold are not random. The key feature that makes a protein functional, resistant to aggregation and degradation, is the high packing of the side chain atoms. In order to help the advancements for prediction of this kind of packing, various networks have been established, among them probably most prominent the so-called CASP (**C**ritical **A**ssessment of techniques for protein **S**tructure **P**rediction) [23]. The first CASP experiment was run in 1994 and it consisted of three parts: the collection of targets for prediction from the exper-

imental community, the collection of predictions from the modelling community, and the assessment and discussion of the results.

The second problem with protein folding is concerned with the proceedings of the actual kinetics and dynamics of the process. If we assume that the essential concept in protein folding is a random unbiased search, all conformations that a polypetide can assume should be equally probable. Therefore the time taken for the protein to find the native conformation should be the number of conformations available (say, 10^{70} for a 100-residue protein) times the time taken to find one conformation (say, 10^{-11} seconds), which leads to an enormous time of folding: about 10^{52} years. We know that most single domain proteins are able to fold *in vitro* to their native folds within a second, but many lie in the mili- or micro-second range. There seems to be a contradiction, which is known in the literature as the Levinthal's paradox, introduced in 1968 [24]. There is, however, an obvious flaw in Levinthal's analysis, because it assumed that the search is unbiased with no stabilisation of any particular conformation. Today it is widely accepted that the native state of the protein is the most energetically favourable one on the potential energy surface. Each state assumes a certain position on the mentioned surface, which means that not all states are equal in free energy and hence the search for the native fold can not be unbiased. The way this view has evolved from Levinthal's paradox is described in the next section.

1.2.2 Models of protein folding

There have been many propositions for explanation of how the conformational space is restricted so that the folding time is reduced to the experimental range. The earliest ideas by Levinthal were that there exist specific pathways for folding. By restricting

the molecules to those pathways the polypeptide chain does not need to undergo an extensive search of all the conformational space [24]. In 1973 Anfinsen [25] proposed that the information coded in the amino acid sequence of a protein completely determines its folded structure and that the native state is the global minimum of the free energy. Later, a variety of theories emerged, for example the framework model [26], the diffusion-collision model [27], the nucleation model [28], the hydrophobic-collapse model [29] and the jigsaw model [30]. The hydrophobic-collapse and the framework models were favoured over the nucleation model, because they imply the existence of folding intermediates, which were discovered soon after. All proposed mechanisms and models were able to explain particular pieces of experimental data, but none provided a clear explanation of the folding principles or a solution to Levinthal's paradox.

A real progress has been made recently towards the acceptance of a "new view" of protein folding by the scientific community. The reasons for such a change are partly due to an increasing demand for an explanation of the mechanism of folding, raised by the resolution of numerous protein structures in the post-genomic era. Another reason is that with new biophysical technology, experiments providing information about protein folding on an atomic level and with real-time resolution have been providing more experimental data that can be used to test theoretical predictions. Finally, there have been major advances in theoretical approaches that simulate folding using simplified models in computational analysis [31, 32]. Protein folding is a perfect example of science, where theory and experiment have become crucially inter-dependent for each other's affirmation.

The unified view of protein folding presented in a recent highly cited review by Dobson et al [32], underlies the fact that protein folding is a progression in which both

native and non-native contacts stabilise native-like structural features. The folding either proceeds through a hydrophobic collapse to a compact globule that has stabilising interactions or through a slow formation of a folding core (nucleus), which then rapidly proceeds towards the native state. Folding is seen as a step-wise behaviour, sampling regions of structure that are downhill in energy.

An important element in the "new view" of protein folding is the folding funnel, which was first introduced by Onuchic et al [33]. This is one way of representing the folding landscape with the free energy (enthalpy and entropy) as a function of folding progress variable, also known as the fraction of the native contacts. In the light of this simple surface (see Figure 1.6), it is possible to understand a number of features of the folding process. There are three kinds of states that can be easily distinguished in the folding funnel.

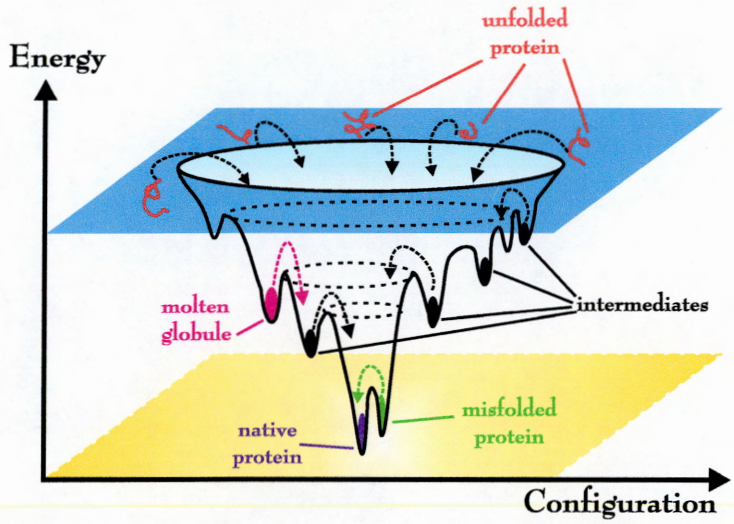

Figure 1.6: **Protein folding funnel**: Schematic of the folding energy landscape, where the energy of the protein is displayed as a function of the topological arrangements of the atoms (adapted from Schultz et at [34]).

The initial state from which the folding proceeds is extremely heterogeneous and encompasses a large conformational space of rapidly inter-converting states. Nowadays,

it is generally accepted that the unfolded or denatured states are not completely random coil as one would expect for a polymer. They have intrinsic propensities for native and non-native like interactions, which funnel the folding process either through global or local conformational preferences [35]. Compact denatured states, commonly known as molten globules, are lower in energy in the folding funnel. These have been in the past defined with a set of well-defined features such as a set of secondary structural elements in the absence of tertiary structure. In contrast, at the bottom of the funnel we find a highly compact state, where the close packing of the side chains is essential for a well-defined conformation. This is the so-called native state. The important lesson that scientists have learned in the past 30 years is that the search for stability and the native fold is far from random. Therefore Levinthal's paradox can be considered resolved.

1.2.3 Protein folding and disease

The understanding of the protein folding process is not only a great intellectual challenge, but also a crucial step in the development of strategies to prevent and treat some debilitating human diseases. Evolution provided proteins with the ability to fold to their unique functional states in the most complex cellular environments, but under some conditions the proteins are unable to fold correctly or to remain correctly folded. This failure can result in a wide range of diseases, some of which are mentioned in Table 1.1.

One group of diseases are amyloid diseases, which include many of the most debilitating conditions in modern society, particulary those associated with ageing, for example type II diabetes and Alzheimer's disease. Some are familial, some associated

Disease	Protein
Hypercholesterolaemia	Low-density lipoprotein receptor
Cystic Fibrosis	Cystic fibrosis transmembrane regulator
Phenylketonuria	Phenylalanine hydroxylase
Huntington's	Huntingtin
Marfan syndrome	Fibrillin
Osteogenesis imperfecta	Procollagen
Sickle cell anaemia	Haemoglobin
α1-Antitryspin deficiency	α1-Antitryspin
Tay-Sachs disease	β-Hexosaminidase
Scurvy	Collagen
Alzheimer's disease	β-Amyloid/presenilin
Parkinson's disease	α-Synuclein
Creutzfeldt-Jakob's disease	Prion protein
Familial amyloidoses	Transthyretin/lysozyme
Retinitis pigmentosa	Rhodopsin
Cataracts	Crystallins
Cancer	p53

Table 1.1: **Representative protein folding diseases**: data taken from Dobson [36].

with medical treatment (e.g. haemodialysis), or infection (e.g. prion diseases), others are sporadic (e.g. most forms of Alzheimer's). Some of these diseases can be even found in both sporadic and familial forms, for example amyloidoses associated with protein transthyretin. Diseases such as Creutzfeldt-Jakob's, Alzheimers's, type II diabetes and systemic amyloidosis are all associated with the deposition of proteinaceous aggregates in a variety of organs, such as the liver, heart and brain [35]. These diseases are described as "amyloidoses" because the aggregated material stains with dyes such as Congo Red in a manner similar to starch (amylose), so the aggregates are termed "amyloids" and the typical fibrous structures as "amyloid fibrils". In addition, some diseases such as Parkinson's and Huntington's disease appear to involve very similar aggregates intracellularly rather than extracellularly, so are not strictly part of amyloidoses.

1.3 Formation of disulfide bonds

1.3.1 Studying oxidative folding

In proteins, which contain cysteine residues, disulfide bonds are a central feature of their native tertiary scaffold. A composite process that includes both the recovery of native disulfide bridges (*disulfide bond regeneration*) and the native three-dimensional structure (*conformational folding*) is called the oxidative folding process [37]. A complete account of an oxidative folding reaction should therefore ideally entail an analysis of the conformational assembly and a study of the disulfide species that appear in the process.

The conformational folding process is most commonly studied by inducing the removal of the native protein structure (either by a high concentration of a denaturant, extreme temperature or pH), which is followed by monitoring the structural changes that take place as the protein is allowed to regain its native fold. The proteins can spontaneously refold *in vitro* to their native structures and this is followed by physicochemical, mostly spectroscopic methods, for example circular dichroism (CD), nuclear magnetic resonance (NMR), fluorescence or UV absorbance. These kinds of studies can give information at the level of a single residue and even a single atom in the polypeptide structure and are hence most versatile for looking at restorations of native secondary and tertiary structures.

On the other hand, the elucidation of the exact role that disulfides play in the process is traditionally studied by a time-resolved RP-HPLC (reversed-phase high performance liquid chromatography) analysis of acid trapped samples from an oxidative folding reaction. The regeneration of native disulfide bonds involves both the oxidation

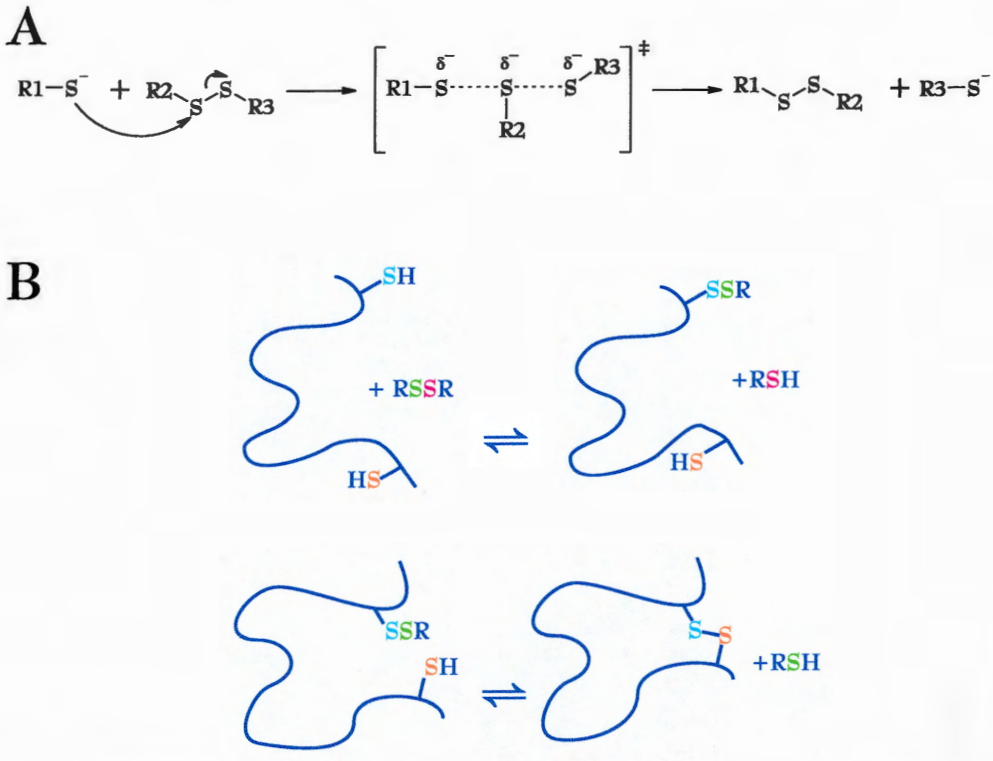

Figure 1.7: **Disulfide formation via thiol-disulfide exchange mechanism**: **A**. Thiol-disulfide exchange mechanism: in the pH range above 8, cysteine thiols are readily converted to thiolate anions (RS^-), which are potent nucleophiles. RS^- anions attack a disulfide bond, displacing one sulfur atom and forming a new bond with the other sulfur atom (nucleophilic substitution). The rate determining step of this concerted process is the formation of a transition state with a partial transfer of the negative charge (δ^-) over the three sulfur atoms [38]. **B**. The formation of a disulfide bond on the polypeptide chain (solid curve) with the help of a small molecule reagent (thiol form: RSH, disulfide form: RSSR). The two steps both proceed via a thiol-disulfide exchange reaction. In order to be able to follow the mechanism of exchange between sulfur atoms, each atom has been given a different colour. The first step shown is intermolecular and the second intramolecular. The rate of the intramolecular step is relevant to protein folding, since it involves also conformational folding process.

of thiols and their isomerisation (reshuffling) from non-native to native disulfide bonds (via thiol-disulfide exchange mechanism, Figure 1.7). The study of oxidative folding *in vitro* can provide a detailed structural description in terms of disulfide intermediate species present along the pathway of the complex folding process. Such pathways depend on the proximity, reactivity and accessibility of cysteine residues in the protein sequence. The real advantage of this approach is the fact that it allows the intermediate species to be isolated and to be studied in detail, because they are stable and covalent.

1.3.2 Model system studies

Several disulfide-rich proteins have been characterised showing a complex diversity among their oxidative folding pathways. These pathways can be distinguished in terms of heterogeneity, abundance and disulfide structure of the intermediate species that accumulate during the oxidative folding process.

BPTI's folding pathway was the subject of an intense dispute in the early 1990's [39], but later resulted in one of the most extensively studied oxidative folding pathways and a major protein folding model. It was much earlier (1977) that Creighton [40] pioneered the technique by doing the analysis of BPTI's folding with iodoacetamide trapping of disulfide intermediate species and a subsequent ion-exchange chromatography separation. A subsequent landmark study by Weissman and Kim in 1991 [41] done with acid trapping and HPLC analysis, revealed a slightly different pathway, because the intermediates with non-native disulfide bridges were detected at much lower concentrations. Two important questions for any kind of oxidative folding analysis were raised from the extensive studies of BPTI. The first one is whether the most abundant intermediates are also the kinetically important species for the pathway and the second one is about the appropriate conditions of study, for example the pH value that can vary from 7.0 to 8.7. With some differences, BPTI's pathway was characterised by both groups with the predominance of only a limited number of folding intermediates, that adopt mainly native disulfide bridges and native-like structures [42, 43, 44, 45, 46]. It is important to remember that 1- and 2-disulfide intermediates were present, but no 3-disulfide species apart from the native protein was detected on this pathway. One of the most abundant intermediates is a two disulfide species with two native disulfide bonds and a native-like structure. Formation of the third disulfide (Cys14-Cys38) is the last step of the

folding process. A prevalence of the native-like structures and native disulfide bridges points to the conclusion that non-covalent interactions that are specific to the amino acid sequence can guide the initial stages of the folding process and hence admit a very limited number of disulfide species on the pathway.

Using the same technique of acid-trapping and HPLC analysis, Chang et al have analysed folding and unfolding of a number of small three-disulfide proteins and found distinctive mechanisms, which reflect distinctive ways the native disulfide bonds in a protein are formed. There are two novel features of the studies of hirudin [47, 48], potato carboxypeptidase inhibitor (PCI) [49, 50], tick anticoagulant peptide (TAP) [51, 52], epidermal growth factor (EGF) [53, 54] and some other proteins [55, 56, 57]. The first one is that some pathways show a greater heterogeneity of one- and two-disulfide intermediates, for example hirudin has about 40 such species. The second one is that some of these proteins have intermediates with three-disulfides, including some non-native bonds.

Various aspects of disulfide formation, folding and unfolding have been studied for ribonuclease A (RNaseA) [58, 59, 60, 37, 61], which has become another well understood model system. The making of the four disulfide bonds of RNaseA proceeds through a pre-equilibrium stage with ensembles of unstructured species, followed by rate-determining steps involving the conversion of two native-like three disulfide species to the native protein. Ribonuclease A is another good example of a protein that develops native disulfide bonds, which cause stable native-like tertiary structures in intermediate species. However, Scheraga et al [37] do not recognise that the oxidative folding mechanism can also produce stable tertiary structures, resulting from non-native disulfide bridges, despite the extensive studies by Chang et al mentioned

above and the published structures of such intermediates [62].

The formation of disulfides in numerous other proteins has been studied, for example hen lysozyme [63, 64, 65, 66], insulin-like growth factor [67, 68, 69], various toxins [70, 71], nerve growth factor [72], cyclotide kalata B1 [73] and others.

1.3.3 Catalysts of oxidative folding *in vivo*

Disulfide bonds are created *in vivo*, along with other post-translational modifications, in the endoplasmic reticulum (ER) of eukaryotes and the periplasm of Gram-negative bacteria. Already Anfinsen et al [25] predicted that proteins require a sufficiently oxidising environment (i.e. in the ER, the reduction potential is negative: E°=-0.18 V) in order to form their disulfide bonds. They had not envisaged, however, the complexity of the oxidative folding reaction and the numerous proteins and small molecules which help in achieving the right disulfide pairing. Furthermore, the long-held hypothesis that glutathione is the primary oxidant in the ER has been disproved despite the millimolar concentrations of both the reduced and the oxidised form found in this compartment [74]. On the other hand, the catalytic activity in many cellular compartments, as well as in the ER, is being attributed to chaperones. They are molecules that bind partially folded or misfolded proteins in a transient or permanent manner to facilitate folding and prevent aggregation [75].

The currently accepted hypothesis is that the major catalysts of disulfide formation are proteins of the thioredoxin family. Thioredoxins are redox proteins that cycle between the reduced (dithiol) and the oxidised (disulfide) form of the sequence Cys^1-X-X-Cys^2 situated at their active site. The thiol group of the accessible Cys^1 reacts with protein disulfide bonds to generate a mixed transient disulfide, which is displaced by

the adjacent group of the second, buried Cys^2 residue to form the thioredoxin disulfide bond. This is a mechanism very similar to the described thiol-disulfide exchange reaction in Figure 1.7. Thioredoxins catalyse three types of reactions: formation, reduction and isomerisation of disulfide bonds and they also display chaperone activity *in vivo* and *in vitro*.

In the major pathway of eukaryotic cells for disulfide bond formation the oxidising equivalents flow from **ER** oxidation 1 protein (Ero1p in yeast, Ero1-L in human) to Protein Disulfide Isomerases (PDI). Ero1p was discovered in 1998 to directly oxidise PDI protein, which in turn oxidises a variety of newly synthesised proteins [76, 77, 78]. Ero1p is an essential membrane associated flavoenzyme and has two pairs of conserved cysteine residues, which are likely to form redox active disulfide bonds. The eukaryotic PDI family of enzymes [79] (i.e. human PDI and yeast PDI1p), on the other hand, are four-domain (*a*, *a'*, *b*, *b'*) enzymes known to have two catalytic sites consisting of the Cys-His-Gly-Cys motif, which are situated in the *a* and *a'* domains. The structures of the *a* and *b* domains were determined separately by NMR and they both belong to the thioredoxin fold [80, 81]. PDI enzyme was first isolated from liver in 1966 by Anfinsen [82], but has since been found in a variety of tissues and its gene is highly conserved across organisms. The crystal structure of a complete bacterial homologue, DsbC (Figure 1.8) has been recently resolved [83] and provides a model for PDI structure.

DsbC contains two C-terminal catalytic domains and two N-terminal dimerisation domains, which is reminiscent of the four thioredoxin-like domains of PDI. It is one of the seven proteins, which are involved in the formation of disulfide bonds in the bacterial periplasm (DsbA to DsbG). The V-shaped dimer structure has the catalytic

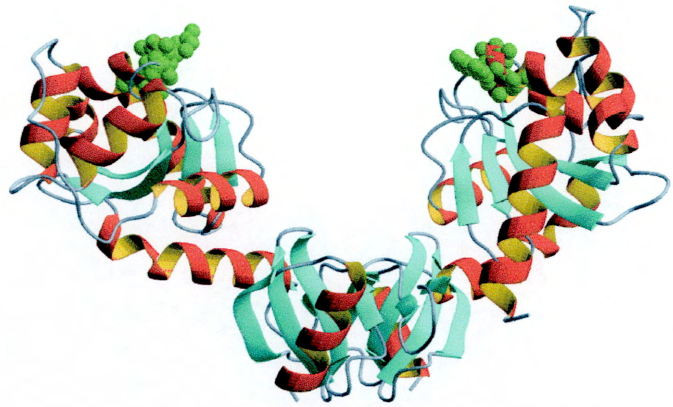

Figure 1.8: **Protein disulfide isomerase, DsbC, structure**: ribbon diagram of the homodimer with the atoms of two active sites shown in green spheres (Cys98-Pro99-Tyr100-Cys101). The N-terminal residues (1-61) of each monomer form the dimerisation domains, whereas the C-terminal domains (residues 78-216) have each a thioredoxin and a helical subdomain. The broad hydrophobic cleft in the middle is large enough to bind small protein domains (dimensions: 40x40x25 Å) .

sites pointing towards each other across the broad hydrophobic cleft, which is important for binding of unfolded proteins and may be involved in chaperone and disulfide bond isomerase activity.

Another protein that can oxidise PDI1p in yeast independently of Ero1p is the flavoenzyme Erv2p (**E**ssential for **r**espiration and **v**iability 2 **p**rotein). Its function in the ER and three-dimensional structure have been recently determined [84, 85]. Unlike Ero1p, Erv2p binds the luminal flavin adenine dinucleotide (FAD) tightly and is oxidised directly by O_2 in order to transfer disulfide bonds to PDI1p. The strained loop created by a disulfide bond between the residues Cys176 and Cys178 is a good oxidising agent and its active site, composed of Cys-Gly-Glu-Cys residues, is located on the N-terminal α_3 helix. The Erv2p pathway for biosynthetic disulfide bond formation is the first to identify the ultimate electron acceptor, O_2 [84].

1.4 Aim and scope of the thesis

This introductory chapter provides scientific background for the thesis, which elucidates the oxidative folding process of the α-amylase inhibitor from Amaranthus plant seeds. The analysis of this process is aimed to describe it both at the level of the disulfide species and the three-dimensional conformations, which are formed along the pathway. The scope of such a study is two-fold. On one hand it contributes to the solution of the protein folding problem(s) and at the same time it focuses at discoveries of new biological phenomena and at developing novel biophysical techniques for the future studies of protein folding.

Chapter 2

Experimental techniques

2.1 Where theory meets the experiment

The biophysical and biochemical basis of protein folding has been a challenging and fascinating topic in theoretical and experimental studies for more than three decades. A major goal for such studies is the development of a model that can make predictions about the effect of various factors on the kinetics and thermodynamics of the folding process [86]. The development of such a model requires an interdisciplinary approach, involving a range of biochemical and biophysical techniques that study structure, dynamics, energetics and mechanism. Further, these techniques need to be applied to numerous proteins in order to be able to draw trends and patterns between various systems and make a successful hypothesis of a protein folding model.

The present chapter introduces the techniques used to elucidate the oxidative folding of *Amaranthus α-amylase inhibitor* (AAI). Each section includes a historical and theoretical background of the technique, followed by the experimental details used in the study of AAI protein. An emphasis is made on nuclear magnetic resonance

(NMR) spectroscopy and its application in photo-chemically induced dynamic nuclear polarisation (photo-CIDNP), because a substantial proportion of the data was obtained with these techniques.

2.2 Solid-phase peptide synthesis

2.2.1 The Merrifield technique

The solid phase peptide synthesis technique was pioneered by Bruce Merrifield in 1963 [87], when he published a single author paper describing the synthesis of the tetrapeptide Leu-Ala-Gly-Val by successive addition of benzyloxycarbonylamino acids to a polystyrene resin. The principle of his approach was that any growing chain, be it a peptide, oligonucleotide or other oligomer is elaborated while being attached to a stable solid support. There are numerous advantages of this approach over the traditional solution-phase synthesis:

- all reactions are carried out in a single vessel
- excess reagents drive reactions swiftly to completion and are easily filtered from the system, eliminating the need for purification of intermediates after each step
- many repetitive steps enable the automation of the whole process

On the other hand, the solid phase peptide synthesis has the disadvantage of being insensitive to many analytical techniques for the characterisation of intermediates of the reaction and also by causing insoluble by-products by incomplete reactions. At the beginning some peptide chemists were very reluctant and critical of the approach, but time has shown that its advantages have clearly outweighed its disadvantages.

2.2.2 Protecting groups and formation of a peptide bond

Peptide synthesis requires selective protection (blockade) of the reactive amino and carboxy groups in amino acids to obtain the desired peptides and form correctly the peptide bond between the amino acid building blocks. Any coupling of amino acids in peptide synthesis therefore requires coupling of protected amino acids, followed a by deprotection step.

The criteria for a successful protection of the α-amino group during peptide synthesis are demanding. It needs to suppress its nucleophilic activity either by draining the electron density into an appropriate substituent or by concealing it with a screen of steric hindrance. The most important protecting groups for both historic and experimental reasons are Benzyloxycarbonyl (*Z*), *t*-Butoxycarbonyl (*Boc*) and 9-Fluorenylmethoxycarbonyl (*Fmoc*) protecting groups, which are shown in Figure 2.1. The discovery of the *Z* group (Bergmann and Zervas in 1932 [88]) is regarded a turn-

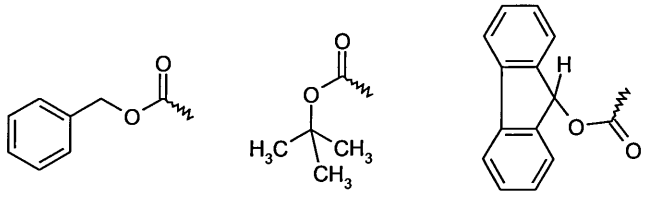

Figure 2.1: **α-amino protecting groups**: Benzyloxycarbonyl (*Z*-), *t*-Butoxycarbonyl (*Boc*-) and 9-Fluorenylmethoxycarbonyl (*Fmoc*-).

ing point in the origin of modern peptide synthesis. In the 1960s, as the solid state approach was appearing, *Boc* and *Fmoc* groups were becoming more popular.

The side chain protecting groups are attached either to the amino acid side-chains or carboxy terminal resin linkage. They can all be cleaved off (deprotected) simultaneously in a single step after the synthesis is complete. The choice of benzyl

ester for the peptide linkage to the resin strongly favors similar protection of the amino acid ends: benzyl ether, thioether, ester and urethane.

As far as solid supports are concerned, the widely used polyamide supports were introduced by Sheppard and co-workers [89, 90]. It was desirable to have a polymeric carrier which would be readily solvated in dipolar organic solvents together with the peptide chain. This system would be similar to an open gel system in which there is an easy access for reagents to the active sites.

Today, there are mainly two well described and known approaches: the Merrifield and the Sheppard approach. The Merrifield technique is synonymous to the Boc-polystyrene implementation, whereas the Sheppard technique uses Fmoc-protection on polyamide supports. Most steps of both approaches are very similar and are outlined in Figure 2.2. Solid phase peptide syntheses start with an addition of protected amino acid in an excess amount to the amino resin in a non-polar solvent such as dichloromethane. After the first amino acid has been reversibly linked to the solid support, a series of repeated reaction cycles follow for each residue added to the support. The three most important steps for each residue are deprotection in mildly acidic conditions, neutralisation and acylation.

2.2.3 Synthesis of AAI peptide

The protected peptide chain was assembled on a 2-chlorotrityl resin (Alexis corporation) to which the first amino acid Fmoc-Ser(tBu)-OH was manually introduced. The chain assembly was carried out manually on a 0.750 mmol scale and the initial substitution rate was 0.283 mmol/g. A protection scheme was used, which could be cleaved in a single step after the synthesis was completed (triphenylmethyl, Trt:

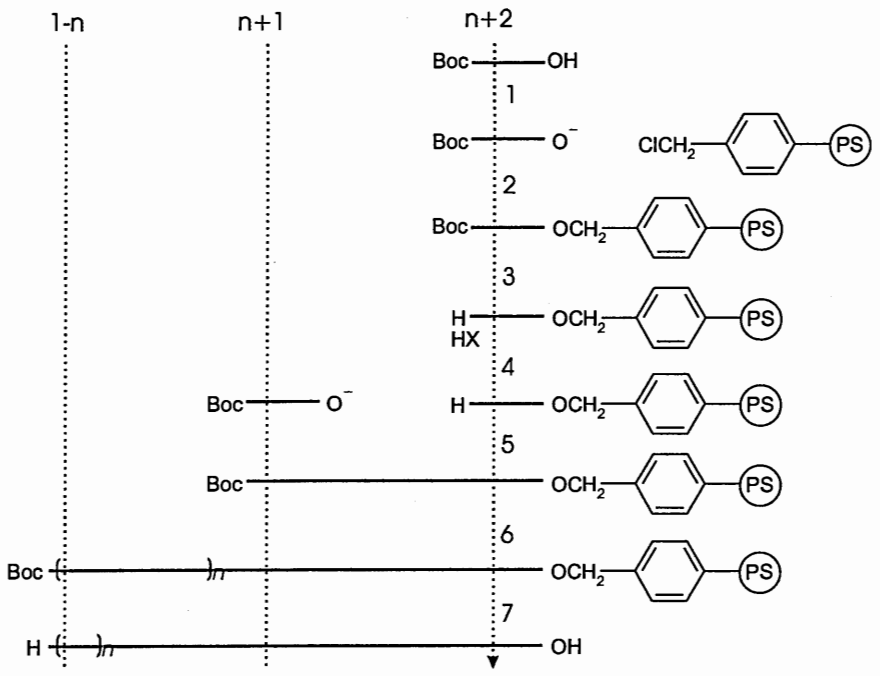

Figure 2.2: **Solid Phase Peptide Synthesis steps with Boc-strategy**: 1. salt formation, 2. esterification by nucleophilic displacement, 3. deprotection in mildly acidic conditions, 4. neutralisation, 5. coupling, 6. repetition of steps 3-5, 7.vigorous acidolysis. Adapted from Jones [91].

Cys, Asn; tert-butyl, tBu: Tyr, Ser, Thr; tert-butyl ester, O-tBu: Asp, Glu; tert-butoxycarbonyl, Boc: Lys, Trp; Pbf: Arg). Single couplings were carried out for 60-120 min using in situ Fmoc-amino acid with HOBt/TBTU/DIEA, apart from cysteine residues, which were introduced as Fmoc-Cys(Trt)-OPfp with HOBt/TBTU/DMP. Double couplings were necessary at the following positions: 1, 2, 3, 4, 5, 6, 7, 8, 11, 23. Fmoc-removal was with 20% piperidine in N-methylpyrrolidinone (NMP). The synthesis was monitored at each coupling step. Kaiser test [92] was performed to test for unreacted primary amines left on the resin after the coupling of every amino acid residue apart from proline, after which chloranil test [93] is more suitable and tests for secondary amines. Cleavage/deprotection mixture (20 ml/g peptide resin) containing TFA:H_2O:phenol:EDT:TIS = 81.5:5:5:2.5:1 was prepared fresh prior to use. The peptide resin was left in the cleavage mixture for 3h, after which the molecular weight was

checked with ESI-MS. The peptide mixture was filtered separately from the resins and washed with TFA, then evaporated at room temperature in vacuo. The peptide solution was precipitated with 20 volumes of cold diethylether, the pellets were separated, washed with ether (3x 30 ml) and dried.

2.3 Chromatographical techniques

The International Union of Pure and Applied Chemistry (IUPAC) states that chromatography is a physical method of separation in which the components to be separated are distributed between two phases, one of which is stationary while the other is mobile (either gas or fluid) and moves in a definite direction.

Tswett is considered the inventor of column chromatography and was the first to publish a paper in 1906 about its use [94]. He described the separation and isolation of intensely colored green and yellow chloroplast pigments and therefore named the technique after *chromatos*, the Greek word for color. However, the foundation of modern liquid chromatography is considered to be laid by Martin and Synge in 1941 [95], when they published a paper on separation of amino acid mixtures. In the following post-war years until the 1960s, many different methods were developed such as thin layer chromatography (TLC), gas liquid chromatography (GLC), ion-exchange chromatography (IEC) etc. The development of HPLC was much slower. It is the most widely used technique nowadays and can be extended to any form of column chromatography with a liquid mobile phase. At first, in the 1970s, long-chain hydrocarbons were bonded to the surface of silica beads and used as a stable stationary liquid phase, but this produced high back pressure, which is why the name high pressure liquid chromatography (HPLC) was associated with it. In the late 1970s, short columns with small particles

were introduced to achieve faster separations with better efficiencies, which renamed HPLC into high performance liquid chromatography.

2.3.1 HPLC techniques

Chromatographic processes in HPLC can be classified according to the type of stationary phase, which refers to the solid support contained within the column over which the mobile phase continuously flows. Columns containing various types of stationary phases are commercially available. Some of the more common types include: adsorption, size exclusion, normal phase, reversed-phase, ion exchange and affinity.

One of the most common HPLC techniques, which was used for the analysis of protein samples in this thesis is the so called Reverse Phase-HPLC (RP). It operates on the basis of hydrophilicity and lipophilicity. The stationary phase consists of silica based packing with covalently bound n-alkyl chains. For example, C-8 signifies an octyl chain and C-18 an octadecyl ligand in the matrix. The more hydrophobic the matrix on each ligand, the greater is the tendency of the column to retain hydrophobic moieties. By means of using gradients between hydrophilic (mostly water) and hydrophobic (usually acetonitrile) solvents, mixtures are separated on the basis of their hydrophobicity. Separation occurs as the mobile phase with the sample passes over the stationary phase and the compounds of interest are retained. Two other widely used techniques are Ion-Exchange (IE) and Affinity Chromatography. IE is performed with columns containing charge-bearing functional groups attached to a polymer matrix. Affinity, on the other hand, operates by using immobilized biochemicals that have a specific affinity to the compound of interest.

2.3.2 RP-HPLC chromatography with AAI protein

Analytical and preparative reversed-phase HPLC were carried out on a GILSON system. It is configured with an 805 manometric module, an 811C dynamic mixer, two model 306 pumps, 118 UV/VIS detector, computer controlled GILSON 506 C system module and Unipoint system software. Samples were eluted with linear gradients of solvents A (H_2O / 0.1% TFA) and B (CH_3CN / 0.1% TFA), unless specified otherwise. Analytical separation was done on a ZORBAX 300 SB-5C18 (150 x 4.6 mm, ID, Hewlett Packard) column at flow rates of 0.8-1.0 ml min^{-1}. Preparative RP-HPLC was done on a Waters Prep LC Universal Base Module with a Prep-Pac cartridge Delta Pac 300 15RC18 (100 x 40 mm, ID, Waters) column at a flow rate of 15 ml min^{-1}.

2.4 ESI-MS and protein structure

Mass spectrometry detects the mass to charge (m/z) ratio of molecular ions and their fragments that are produced after ionisation, for example after an electron impact:

$$M + e^- \longrightarrow M^{\bullet +} + 2e^- \tag{2.1}$$

Molecular ions reveal information about the molecular weight and structure of their precursor molecule. A mass spectrum is a plot of the relative abundance of ions vs. their m/z ratio. Electrospray ionisation (ESI) is one of the most widely used ionisation techniques in mass spectrometry of biomolecules. It is produced by applying a strong electric field under atmospheric pressure, to a liquid with a weak flux, passing through a capillary tube. The electric field is obtained by applying a strong potential difference of 3-6 kV between the capillary and the counter electrode, separated by 0.3-2 cm.

Highly charged droplets are formed, which shrink as the neutral solvent evaporates. At a certain point the charge repulsion forces overcome the cohesion forces, which causes a Coulombic explosion. The mechanism of the field evaporation of charged surfaces is still unclear, but the important fact is that molecular ions are produced, which can be detected in a mass spectrometer.

The atmospheric pressure ionisation (API) is 10^3-10^4 times more efficient than at a reduced pressure of an electrospray ionisation source. The problem remains in coupling the atmospheric source compartment with an analyser compartment, which needs to be kept at very low pressure (10^{-5} Torr). The problem is solved by introducing a very small opening between the compartments and by using high-capacity vacuum pumps. The needle which is maintained at high voltage encounters a counterflow of warm gas (typically N_2 at 45-75°C), which sweeps away the uncharged material and solvent vapor. Ions enter as a free jet expansion into a skimmer and then into the inlet optics of the mass analyser. This procedure is shown schematically in Figure 2.3. This approach allows a liquid chromatography apparatus to deliver the sample to the ion spray, which has been used for applications in analytical chemistry and biochemistry.

2.4.1 Mass spectrometry of proteins

The ESI-MS technique was primarily developed by Fenn and co-workers, who have pioneered the sampling of ions at atmospheric pressure coupled with quadrupolar detection. Apart from molecular weight determination of biomolecules, the other application of this technique is its use as an interface for coupling of mass spectrometry with liquid chromatography. The multiple protonation of basic residues in the ESI mass spectra of proteins extended the mass range to more than 1,000,000 Da with quadrupole detection

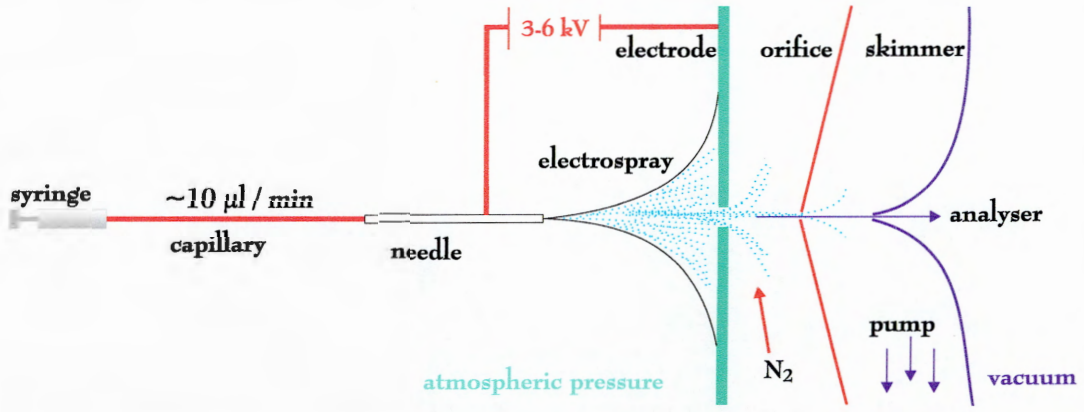

Figure 2.3: **Electrospray ionisation (ESI) mass spectrometer**: Schematic illustration of atmospheric pressure ionisation (API) electrospray technique. ESI operates by infusing a liquid containing the sample of interest through a capillary, whose outlet is in an electric field. This causes the liquid to disperse into fine droplets which are eventually desolvated to release ionized species into the gas-phase.

and a precision of 0.01% in the 5-40 kDa mass range.

The molecular mass can be calculated by deconvolution of multiply charged peaks into the singly charged peaks. For positive and negative ions, respectively, the following equations are given for calculation of the molecular mass:

$$M = z_1(r_1 - m_p) \tag{2.2}$$

$$M = z_1(r_1 + m_p) \tag{2.3}$$

where z_1 is the charge and r_1 is the m/z ratio of a molecular ion and m_p is the proton mass, which is included because the API occurs by loss or gain of a number of proton(s). Calculation of M for each of the observed m/z values enhances the precision of the estimation with an experimental error $< 0.01\%$.

The ESI-MS spectra normally contain a series of peaks which have a statistical distribution of intensities. They represent multiply charged molecular ions obtained through protonation or deprotonation. The number of peaks in the spectrum is usually

about half the number of the most abundant charge state. Monitoring the behaviour of the series of peaks produced in the electrospray ionisation process, the charge state distribution (CSD), was one of the first application of ESI-MS to protein structural studies. It has been shown that proteins in their native conformation, for example, give narrow CSD, whereas under denaturing conditions proteins give spectra with a broad series of lines, centered around a higher charge [96]. Another application of ESI-MS which has been widely studied is the exchange of labile protons with solvent deuterons [97]. This method provides global and local information on the dynamic properties of proteins on the basis that protons in the hydrophobic core or those involved in hydrogen bonding will have exchange rates several orders of magnitude slower that the solvent exposed ones. Problems such as folding, stability and protein-DNA interactions have been successfully studied with ESI-MS [98, 99].

In the last ten years, we have seen an increasing emergence of a variety of systems combining liquid chromatography and mass spectrometry (LC-MS). This is a very powerful analytical tool for the study of biological phenomena at the molecular level, which synergistically combines the advantages of both techniques to give better information. The mass spectrometer has an increased ionisation efficiency due to a pre-separation of the compounds in the sample. Essentially, it allows to determine molecular mass and structural characteristics of all the peaks in a chromatogram. This technique incorporates steps such as desalting and pre-concentration and prevents sample handling loss. The disadvantages of this system include the relatively high costs of equipment and operation and the limited column and buffer types of the LC system.

2.4.2 ESI-MS experiments with AAI

Electrospray ionisation mass spectrometry experiments were performed on a Perkin Elmer API 150 EX electrospray, single quadrupole mass spectrometer, with a resolution of 3000 Da. The samples were analysed in positive mode either by direct syringe infusion or via a LC-MS system. The direct syringe infusion was done by a Harvard Apparatus (pump type 11) at flow rates 20-50 μl/min. The protein samples were either dissolved in various proportions of water and acetonitrile or analysed immediately after the elution from a RP-HPLC coloumn. To keep the samples at pH 2 for successful analysis in positive mode for ESI, a total of 0.1% TFA was present. In the case of the LC-MS system, a Gilson micro-flow system was configured with an 805 manometric module, an 811C dynamic mixer and two pumps (models 305 and 306). A micro bore rapid resolution column (Agilent Zorbax 300 SB-C18, 1.0 x 150 mm, 3.5 μm) was connected to the API electrospray to run water/actetonitrile gradients at flow rates 40-50 μl/min. The molecular weights of the peptides and the masses for theoretical digestion of the intermediates were calculated using the Applied Biosystems software (programmes BioMultiview and ProMac).

2.5 Circular Dichroism

2.5.1 Circular dichroism of proteins

Circular dichroism (CD) spectroscopy is a form of light absorption spectroscopy that measures the difference in absorbance of right- and left-circularly polarized light. A circularly polarised beam is the result of an addition of two plane polarized light beams perpendicular to each other with a quarter wave phase offset. Circularly polarized

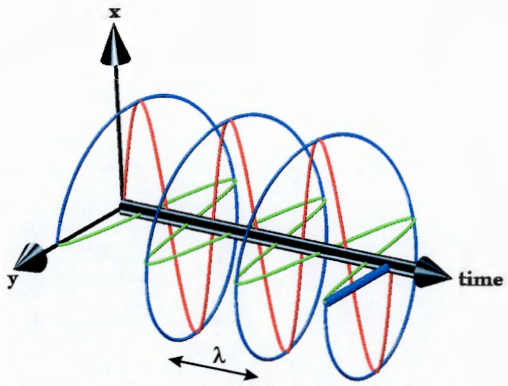

Figure 2.4: **Circularly polarised light** is composed of two plane waves (denoted with red and green) of equal amplitude by differing in phase by 90°. The tip of the field vector appears to be moving in a circle as it travels through time. If it appears to be rotating clockwise, the light is said to be right-circularly polarized, otherwise (counter-clockwise rotation) it is left-circularly polarized light. The electric field vector makes one complete revolution as the beam advances one wavelength, as denoted by the distance λ on the time axis. Adapted from O.Smart with permission (*http : //bach.bip.bham.ac.uk/osmart*).

light has constant magnitude of the oscillation and oscillating direction, as shown in Figure 2.4. Right- and left-circularly polarized light will also be absorbed to different extents at some wavelengths due to differences in extinction coefficients for the two polarized rays. The differential absorption of radiation polarized in two directions as a function of frequency is called circular dichroism (CD).

There exists a number of ways the CD of a sample is reported in the literature. The most commonly used units are mean residue elipticity (degree cm^2 $dmol^{-1}$). Equation 2.4 gives the relationship between elipticity, Θ, and mean residue elipticity, $[\Theta]_{MRE}$.

$$mean\ residue\ elipticity \quad = \quad [\Theta]_{MRE} \quad = \quad \frac{\Theta}{10\ n\ c\ l} \quad \Longrightarrow \quad \left[\frac{deg\ cm^2}{dmol}\right] \tag{2.4}$$

Circular dichroism of proteins is mainly concerned with absorption in the ultraviolet region of the spectrum where signals from the peptide bonds and amino acid side-chains can give various information about protein structure [100]. In the far-UV

region (240 nm to 190 or 180 nm), the principal absorbing group is the peptide bond. These studies can be used to quantitatively evaluate the secondary structure of the protein, because regular secondary structure gives distinct spectra. In the near-UV region (320 to 240 nm) the aromatic side chains (Phe, Tyr, Trp) absorb in the range 250 to 290 nm. The chiral environments of these side chains, created by the tertiary structure of the protein, give rise to CD spectra, which can be used as fingerprints of the specific structure. Additionally, broad signals at around 260 nm belong to the absorbance by disulfide bonds. Their intensity depends on different bond angles and the effects of the neighboring groups.

The principal advantages of CD derive from its speed and convenience as good quality spectra can be obtained within 30 min [101]. Further, low amounts of material (compared to for example NMR) are required and the technique is non-destructive, samples being re-usable for further analysis. It is easy to study a wide range of protein concentrations. The main limitation of CD is the limited structural information available giving only an estimate of how much of each kind of secondary structure is present, but not where exactly it is situated. It also gives little information on the quaternary structure of proteins.

2.5.2 Circular dichroism of equilibrium states of AAI

Circular dichroism experiments were performed on a Jasco J-810 spectropolarimeter, which has a wavelength range from 163 nm to 900 nm and is equipped with Spectra Manager software. The spectra both in the far-UV and in the near-UV region were acquired with 1 nm step resolution, 50 nm min^{-1} speed of acquisition, 1 second response time, 1 nm band width, averaged over 16 acquisitions and 20 milli-degree sensitivity.

In the far-UV region, the cuvette with 0.2 mm path-length was used and the spectra were acquired from 190 nm to 260 nm. In the near-UV region, however, a path-length of 1mm was used to acquire spectra from 240 nm to 320 nm. Reference spectra of the buffer in which the protein samples were dissolved, was recorded for both the far-UV and the near-UV experiments and they were subtracted from the spectra to obtain the difference spectra. The protein concentrations for recording the spectra were 0.1 mM and 2 mM in the far-UV and the near-UV regions, respectively.

2.6 NMR and photo-CIDNP

The scientific groups of Bloch and Purcell simultaneously published the results of the first successful NMR experiments in 1946 [102, 103]. In 1954, only one year after Watson and Crick's historic discovery, the first application of NMR to a biological sample appeared in *Nature* featuring the hydration of deoxyribonucleic acid [104]. This was followed by the first ^{1}H NMR spectrum of a protein (ribonuclease) at 40 MHz by Saunders et al in 1957 [105]. Almost 50 years afterwards, NMR is an extremely widespread technique used in a wide range of scientific disciplines, ranging from physics, chemistry, and material science to medicine and recently to the upcoming disciplines such as structural genomics and proteomics. The recognition of the NMR field can be seen by three Nobel Prizes being awarded to scientists for its discovery, development and applications to biomolecular problems: Bloch and Purcell in 1952, Ernst in 1991 and Wüthrich in 2002.

Because of the insensitivity of the NMR technique, attempts to enhance the signals in NMR were pursued thoroughly. An enhancement in emissive and absorptive NMR signals was observed in free radical reactions first by Bargon et al [106] and Ward

and Lawler [107]. Although this phenomenon was first called dynamic nuclear polarisation (DNP), it could not be explained solely on the basis of Overhauser effects. A complete explanation of the mechanism was given by the so-called radical pair mechanism independently by Closs [108] and Kaptein and Oosterhoff [109]. The effect was called chemically induced dynamic nuclear polarisation (CIDNP) and was explained to arise from a spin sorting process arising during the recombination of radical pairs, which leads to greatly perturbed populations of nuclear spin state levels. A laser induced method named photo-CIDNP was devised by Kaptein to enhance specific signals in a protein NMR spectrum [110, 111, 112]. Photo-CIDNP is well understood by the RPM and has been applied to many different problems of structural biology.

2.6.1 NMR theory

Nuclear spin and magnetism

Spin is the intrinsic angular momentum of a nucleus. This vector quantity assumes a value of magnitude $(I(I+1))^{\frac{1}{2}}\ \hbar$ and a quantised number (m_I=I, I-1, I-2, ... -I+1, -I) of orientations, both depending on I, the nuclear spin quantum number. In absence of a magnetic field the orientation onto which the spin is projected is arbitrary, so all the spin levels are degenerate in energy. However, when an external field B_0 is applied to the sample, this degeneracy is removed by the Zeeman effect. The energy levels are consequently split by ΔE (Equation 2.5, where $\hbar$ is 1.05 x 10^{-34}), which depends both on the static magnetic field, B_0, and the gyromagnetic ratio of the nucleus, γ.

$$\Delta E = \hbar \gamma B_0 \tag{2.5}$$

Nuclear Magnetic Resonance (NMR) is the phenomenon of induced transitions between the allowed orientations of the nuclear spin by application of a radio frequency field. Transitions between the states are only observed when the resonance condition is satisfied:

$$h\nu = \hbar\gamma B_0 \tag{2.6}$$

When subjected to a magnetic field, the nuclei spread themselves amongst the $2I + 1$ available energy levels according to the Boltzmann distribution (Equation 2.7). The transitions follow the NMR selection rule $\Delta m_I = \pm 1$ (where m_I is the magnetic quantum number and assumes $2I+1$ values: $-I, -I+1, ..., I-1, I$) to undergo an excitation one level higher or a stimulated emission one level lower in energy.

$$\frac{N_{upper}}{N_{lower}} = \exp(-\frac{\Delta E}{k_B T}) \tag{2.7}$$

The intensity of the NMR signal reflects the net amount of energy absorbed (ΔE) and depends on the difference in population of the two levels between which the transition occurs. In NMR, the transitions for which $\Delta m_I = -1$ outnumber those, where $\Delta m = +1$, just by about 1 in $10^4 - 10^6$, so it is as if the difference in populations is only one nucleus in every $10^4 - 10^6$. It is therefore crucial to maximize the intensity of NMR signals by using strong magnetic fields and to make use of nuclei with high natural abundance and large gyromagnetic ratio, γ. These considerations make ^{1}H the most popular nucleus for NMR studies and currently magnetic fields up to 21.1 T (900 MHz ^{1}H frequency) are commercially available.

NMR experiment

When matter is placed into a static magnetic field, B_0, the large number of spins precess around the B_0 axis at a speed called the Larmor frequency, or the NMR absorption frequency, ω_L. The Larmor frequency, which can be deduced from Equation 2.6 is defined as:

$$\omega_L = \gamma B_0 \tag{2.8}$$

In a macroscopic sample at thermal equilibrium the slight excess of spins in the lower energy level gives a *bulk magnetization*, M_0, which is stationary and aligned parallel to B_0. According to the familiar vector model, the direction of magnetic field B_0 defines the z' axis of the static coordinate system (x',y',z'), known as the *laboratory frame.* NMR experiments consist of a series of pulses, which are linearly oscillating magnetic fields in the radio-frequency range. Say a pulse is applied around, for example the x' axis, with a radio-frequency, ω_{RF} (Figure 2.5). It causes the bulk magnetization vector to tilt away from the z' axis, precessing about the z' axis at the Larmor frequency. In order to view the experiment more easily, we define the *rotating frame*, which is a coordinate system that rotates around the z' axis with the frequency of ω_{RF}. A pulse is now simply a temporary application of a static magnetic field B_1, orthogonal to B_0. After the application of for example a 90° pulse, the magnetization M remains in the xy plane generating time dependent radio frequency signals, described as a *free induction decay* (FID) in Figure 2.6B. The more familiar frequency dependent spectra which are used in NMR representation are related to the FID through Fourier Transformation (Equation 2.10).

$$M(t) = M_0 \exp(i\Delta\Omega t) \exp(-t/T_2) \tag{2.9}$$

$$S(\omega) = \int_0^\infty M_0 \exp(i\Delta\Omega t) \exp(-t/T_2)\, dt \tag{2.10}$$

$\Delta\Omega$ is the difference between the Larmor frequency, ω_L and the spectrometer carrier frequency, ω_0. T_2 denotes the time constant of the exponential decay with which the FID decays (explained below). An alternative way of representation how the pulses transfer magnetization is the product operator formalism, which is inherently quantum mechanical in nature. This formalism describes an NMR experiment in terms of the vectors that represent the magnetization in various direction or as a product of vectors of various spin systems. They can be used for an exact analysis of modern NMR experiments.

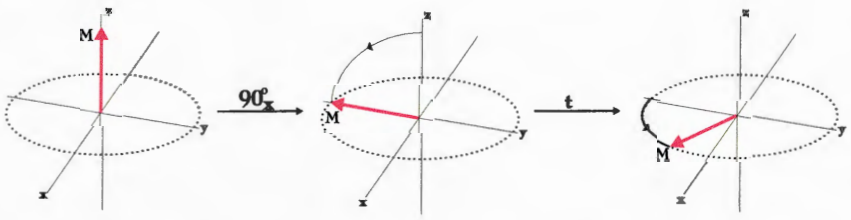

Figure 2.5: 90°_x **pulse in the rotating frame**: Magnetization is first aligned along the z axis and precesses about the x axis during the 90°-pulse to reach the xy-plane. Magnetization will start to precess at frequency $\Delta\Omega$ around the z axis as soon as the pulse is switched off.

Relaxation

Relaxation makes a sample re-usable for multiple pulses and can be divided into two different processes (Figure 2.6A). The populations of the two energy levels return to the equilibrium (Boltzmann) distribution due to a process called spin-lattice or longitudinal relaxation. This means that the z magnetization recovers its original value. This process is approximately exponential and can be described in terms of a single time constant. The second process, called spin-spin or transverse relaxation, is due to the decay of the xy magnetization back to zero by the randomization of phases of the

individual spins.

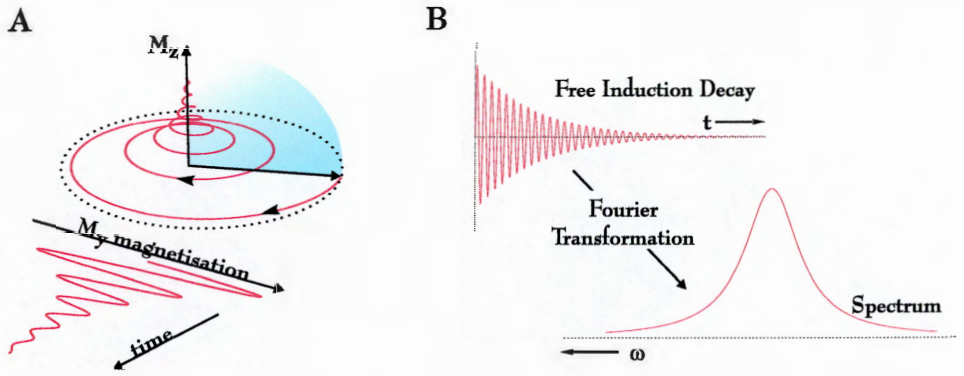

Figure 2.6: **Relaxation and FID**: A. Spin-lattice and spin-spin relaxation: the former restores the z magnetization to its original value, whereas the latter returns xy magnetization to its equilibrium value of zero. B. The basis of the Fourier Transform technique, that transforms the FID signal from the time to the frequency domain.

2.6.2 Photo-CIDNP theory

The chemically induced dynamic nuclear polarisation (CIDNP) phenomenon is observed as greatly enhanced emission or absorption in NMR signals of products in reactions where radical pairs act as intermediates. It originates from the ability of the nuclei to alter the electronic spin state of a radical pair so as to alter its reactivity. The NMR intensity enhancements obtained using the CIDNP method can be used to study radical reaction mechanisms (including determination of spin multiplicity, g-factor and hyperfine coupling) and in conjunction with a suitable light source and a sensitiser, a photo-CIDNP method was developed, which proved a very powerful tool for the study of protein structure and dynamics [110, 111, 112, 113, 114, 115, 116, 117].

Radical pair mechanism and biological photo-CIDNP

The radical pair mechanism scheme [108, 109] may be applied to all CIDNP reactions, but the reaction of interest here is a reversible photo-reaction of a dye with an amino-

acid residue. When an amino-acid residue such as tyrosine, tryptophan, histidine or methionine is exposed to a photo-excited dye such as a flavin, nuclear spin polarization can be generated in the side-chains resulting in selective enhancement of NMR lines. In Figure 2.7A, a scheme for radical pair mechanism is shown, where a molecule (P) is first photo-excited to its singlet state. It then undergoes inter-system crossing to its triplet state, after which it abstracts an electron from a donor molecule (Q) to form a triplet caged-radical pair (primary encounter pair). There are two ways to proceed from here. Either through recombination to starting materials through triplet-singlet mixing or through separation and formation of escape products with other species.

The reaction pathway in biological systems differs from the usual radical pair mechanism in that it is a cyclic reaction (Figure 2.7B). After the dye is excited to its triplet state, both escape and recombination pathways will produce the same products, because there are no other molecules available to scavenge the escaped radicals. This is essential as CIDNP observed is then of the substance itself and not of a chemically modified form.

It may seem, however, that no CIDNP should be observed as there is only one reaction product (native protein) and the equal and opposite polarisations in escape and recombination products would cancel each other out exactly. Fortunately, dipolar coupling with the unpaired electrons in the escaped radicals makes nuclear spin relaxation very efficient, providing that the life-time of the radicals is not much shorter than the nuclear spin-lattice relaxation time, T_1. Escape polarisation is relaxed before the radicals recombine, so only recombination polarisation is observed as a CIDNP peak.

A small amount of dye, a photo-sensitiser, is mixed in with the protein solution in a NMR tube. The most commonly used dyes are flavins, which have a visible ab-

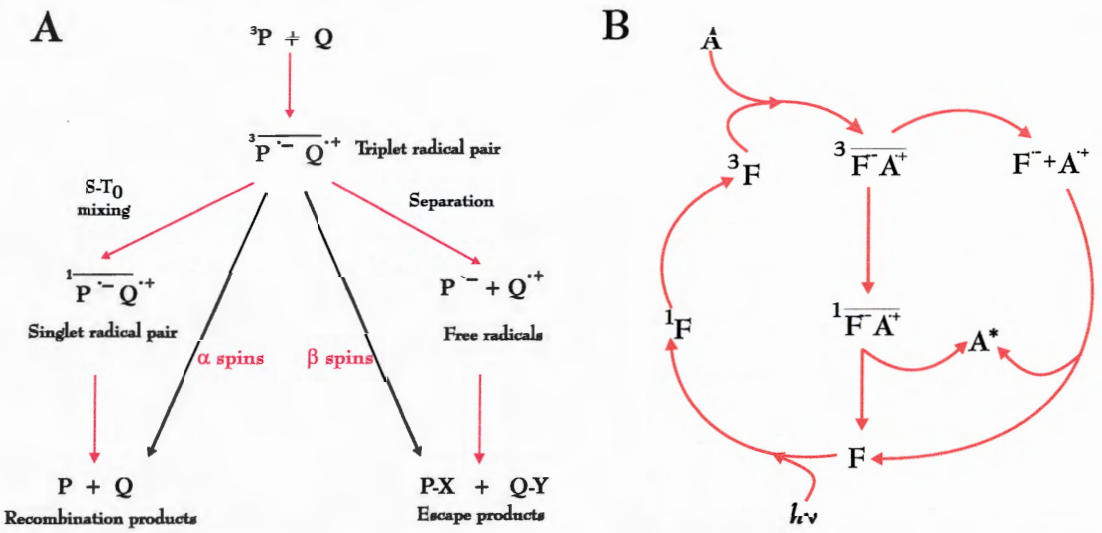

Figure 2.7: **Radical pair mechanism (RPM) and cyclic reaction pathway**: A. RPM guides the reaction of a radical pair by using the nuclear spin state (α or β) as a label. In this example the former (α) leads to the formation of the recombination products. The radical pair first undergoes singlet-triplet inter-conversion during diffusion and a reencounter leads to recombination. For the latter (β), it leads to the formation of escape products. The radicals ***P*** and ***Q*** diffuse apart and react with other molecules. B. Cyclic reaction pathway of photo-CIDNP for proteins: *F* represents the dye, *A* stands for the reacting amino acid side chain. The asterisk denotes nuclear polarization. Only recombination polarisation is observed as a CIDNP peak, because of efficient nuclear spin relaxation in escape products.

sorption band near λ_{max}=450nm. One of the most widely used, Flavin Mononucleotide (FMN), gives excellent signal to noise ratio and amino acid selectivity. One problem with flavin dyes is their photo-degradation, which allows typically only 16 to 64 acquisitions per sample. The main source for signal decay is the progressive loss of the flavin dye due to its reduction. Several solutions have been proposed to resolve this problem, which include the use of a replenishment or mixing device or adding an oxidising agent directly to the protein solution, such as H_2O_2 [113]. In order to resolve this problem of degradation, other dyes have been explored, which could substitute for FMN. Broadhurst [116] and Lopez [115] have separately reviewed a wide a range of different dyes, including xanthenes, thiazines, porphyrins and quinones.

Photo-CIDNP of Amino Acids and Proteins

Only three amino acid residues are readily polarisable: tyrosine (Tyr, Y), tryptophan (Trp, W) and histidine (His, H). Photo-CIDNP spectra of these residues are shown in Figure 2.8. Tyrosine shows a strong emission for the H(3,5) protons and a weak absorption for the H(2,6) and β-CH_2 protons. The photo-CIDNP spectrum of tryptophan shows absorption for H(2,4,6) protons on the aromatic ring and emission for β-CH_2 protons. Similarly, for histidine we observe absorption for the ring protons H(2,4) and emission for β-CH_2. The phases of enhancements are consistent with the prediction of Kaptein's rule for the net CIDNP effect [110].

There are numerous factors that influence the intensity of the CIDNP signal in an amino acid. Most importantly, photo-CIDNP gives us residue specific information on surface accessibility of the polarisable residues, since radical pair formation requires the contact of an amino acid side-chain with the photo-excited dye (Figure 2.7B). A failure to detect a residue polarisation, however, may not always be due to its inaccessibility on the protein surface. Hence, care should be taken in interpreting CIDNP intensities. The cancellation process is for example responsible for the dependence of CIDNP intensities on both the pH and the concentration of the amino acid. The photo-CIDNP signal intensity dependence on pH has been optimised for N-acetyl tyrosine, N-acetyl tryptophan and N-acetyl histidine at pH 4, pH 9 and pH 7, respectively. Intensities can depend also on kinetic competition for the photo-excited dyes between different amino acid side chains present in a protein. It turns out that tyrosine and tryptophan compete much more effectively for flavin triplets than histidine.

The polarisations of amino acid side-chains are expected to be qualitatively the same in a protein environment as those described above for free amino acids, since *g*-

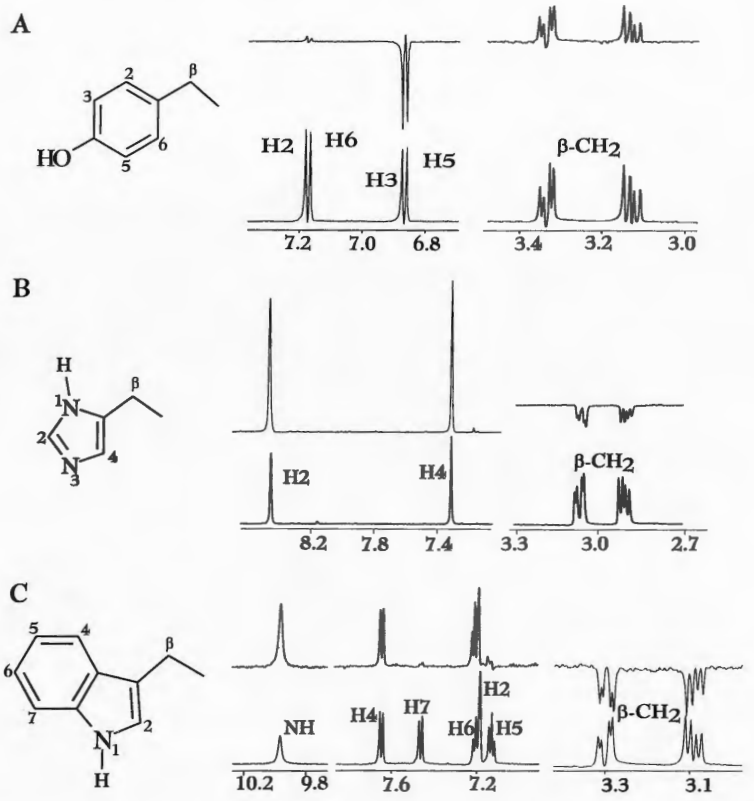

Figure 2.8: **Photo-CIDNP of Amino Acids**: 600 MHz 1D ^{1}H photo-CIDNP (top) and NMR spectra (bottom) of the amino acid residues. The aromatic regions are on the left and aliphatic part on the right hand side. A. Tyr (4mM, pH 4), B. His (10mM, pH 7) and C. Trp (3mM, pH 9). FMN concentration for all samples was 0.2 mM. All spectra have 16 scans with 100ms light flashes.

values and hyperfine coupling constants do not change significantly. However, the large size of proteins and diverse accessibility of different residues causes the magnitude of polarisation intensities in the spectra to differ. The slower rotational diffusion in proteins does not only cause cross-polarisation effects (i.e. transfer of polarisation between nuclei by cross-relaxation processes) to change sign, but can affect NMR linewidths.

Photo-CIDNP proved to be a very successful technique for probing protein surface and has three important advantages over conventional NMR spectroscopy: better resolution of the residue signals, enhancement of signal intensity and information about surface accessible aromatic residues. An example of bovine α-lactalbumin protein (BLA) photo-CIDNP is shown in Figure 2.9. The monitoring of surface accessibility has

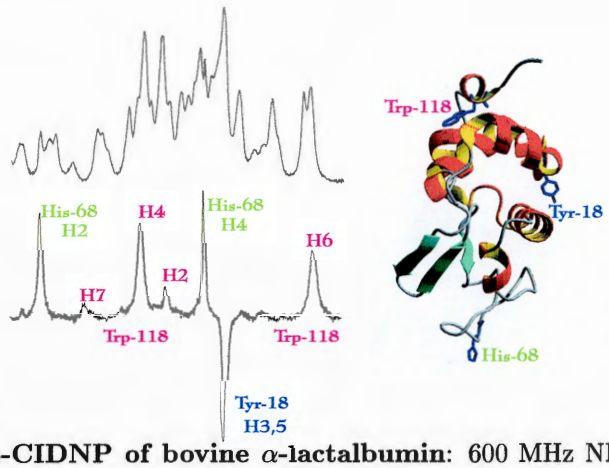

Figure 2.9: **Photo-CIDNP of bovine α-lactalbumin**: 600 MHz NMR (top) and photo-CIDNP (bottom) spectra of bovine α-lactalbumin (BLA) (1.5 mM, pH 7.40, 40°C, 0.2 mM FMN, 16 scans). The aromatic residues, which are solvent accessible and give the photo-CIDNP signals, are denoted on the three-dimensional structure of BLA, made with MolMol [14].

been applied through CIDNP in various contexts and examples include protein-ligand, protein-micelle, protein-nucleic acid, protein-oligosaccharide and protein-protein interactions. Another alternative application that has received surprisingly little attention outside the Oxford CIDNP group (Hore et al), is its application to protein folding. Combining the surface accessibility information with the advantages of using photo-CIDNP as a real-time technique have recently been exploited and several theses and articles are starting to appear [113, 114, 115, 117, 118, 119].

2.6.3 Experimental methods for NMR and photo-CIDNP

For a photo-CIDNP experiment, two spectra are recorded, one after the irradiation with a light source (light spectrum) and another one without irradiation under the same set of conditions (dark spectrum). The difference between the light and the dark spectra, called the CIDNP spectrum, shows only those resonances which are perturbed by the irradiation, i.e. only protons with appreciable hyperfine couplings in the corresponding radicals are polarised. The schematic of the set-up is shown in Figure 2.10.

All NMR spectra were recorded on home-built 600 and 750 MHz NMR spectrom-

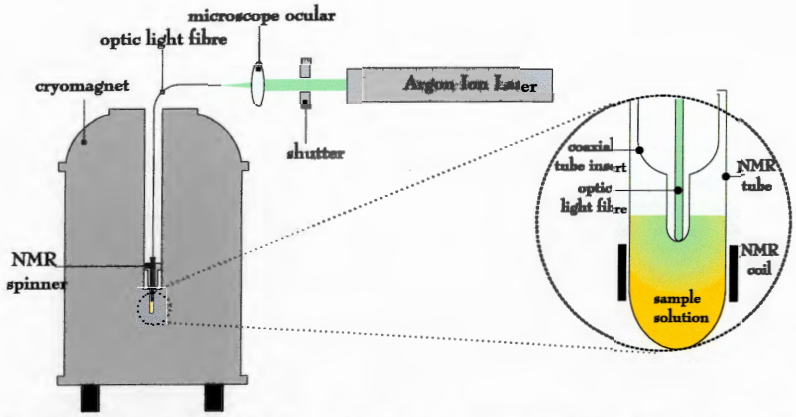

Figure 2.10: **Experimental NMR and photo-CIDNP set-up**: Argon ion laser, magnet immersed in liquid He and N_2 and NMR tube setup with coaxial insert and optical fibre.

eters (Oxford Centre for Molecular Sciences) with operating frequencies of 600.20 and 750.02 MHz, respectively. The spectrometers are equipped with Oxford Instruments magnets, Omega software and digital control equipment (Bruker Instruments), home-built triple resonance probe heads and home-built linear amplifiers. For photo-CIDNP experiments, the light of a 4 W continuous-wave argon ion laser (Spectra-Physics Stabilite 2016-05) with light pulses of 100 ms was introduced into the NMR tube via an optical fibre (diameter 1 mm), inside a coaxial Pyrex insert (Wilmad WGS 5BL). All NMR data were processed using Felix 2.3 software from Biosym Technologies. The experimental details of simple 1D NMR and photo-CIDNP experiments are included in the results and discussion chapters with their corresponding figures.

Two-dimensional NMR spectroscopy

In order to characterise the tertiary structure of some conformations on the folding pathway, 2D COSY and NOESY experiments were done. Here an introduction of both techniques is given with schematic drawings of typical pulse sequences for such experiments.

The basis of the 2D COSY experiment is the process of coherence transfer, in

which magnetisation is transferred through bonds between coupled spins (Figure 2.11).

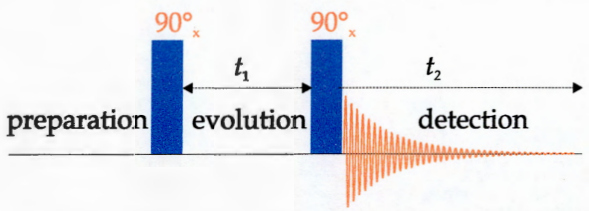

Figure 2.11: **COSY**: Homonuclear COrrelated SpectroscopY pulse sequence, 90°_x indicates the 90° pulses around the x axis, whereas t_1 and t_2 are the evolution and the observation period, respectively. The FID is detected during t_2.

Two radio-frequency nonselective pulses are applied, of which the second is the observation pulse. The FID recorded during the detection period t_2 is influenced by the length of the evolution period t_1, between the two pulses. A second time dimension can be thus created by incrementation of t_1. For each value of t_1 a FID recorded as a function of t_2 is stored so that a data matrix of $s(t_1, t_2)$ is obtained, which after 2D Fourier Transformation produces a 2D frequency spectrum $S(\omega_1, \omega_2)$. The chemical shift information is contained within the diagonal peaks, whereas the off-diagonal cross peaks delineate scalar coupling connectivities. A single 2D COSY experiment can in theory resolve all scalar spin-spin coupling connectivities between protons in a macromolecular structure.

The 2D COSY experiment performed on MFI of AAI protein, was done at 750.01690 MHz spectrometer frequency, with spectral width in both dimensions, ω_1 and ω_2, of 6666.67Hz. The number of points was 1024 in each dimension. The spectra in ω_1 were averaged over 48 acquisitions and 1024 in ω_2. The delay time between the acquisitions was 1.45 s. The protein solution was prepared in a Shigemi tube, in 99.9% D_2O at pH 7.0. The protein concentration was 1.0 mM.

A typical 2D NOESY pulse sequence with three radio-frequency pulses is shown in Figure 2.12. The effect of the first two pulses is the same as for a COSY pulse sequence and the evolution period t_1 is similarly incremented to create the second dimension. The mixing time, τ_m, which separates the second and the third 90°_x pulses is of key

importance, because it is the time in which the NOE's build up. For large biologically relevant molecules in aqueous solution, such as proteins and nucleic acids, the typical mixing times range from 50 to 300 ms. The mixing time, τ_m, has to be selected so that a maximum NOE builds up. There are two competing processes occurring during the delay τ_m, cross-relaxation from the perturbed spins which causes NOE enhancements and spin-lattice relaxation which restores all intensities to their equilibrium values (thus destroying NOEs).

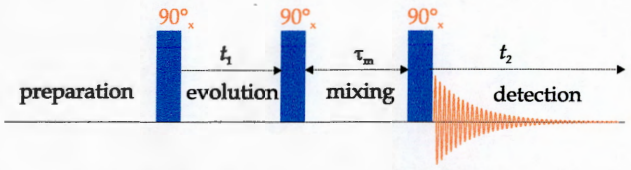

Figure 2.12: **NOESY**: Nuclear Overhauser SpectroscopY pulse sequence. 90°_x indicate the 90° pulses around the x axis, τ_m is the mixing period, when the NOE's build up, whereas t_1 and t_2 are the evolution and the observation period, respectively. The FID is detected during t_2.

However, two or several subsequent cross-relaxation steps can cause spin diffusion, which greatly influences the observed NOE intensities and is the most common reason for a mistaken interpretation of NOESY data. This is why NOE build up experiments are done to examine the most suitable τ_m value and to prevent the loss or falsification of part of the information content of NOESY spectra. Such experiments involve plotting intensities of cross peaks for different τ_m values.

The 2D NOESY spectra of native and MFI species of AAI protein were done at 600.1507820 MHz spectrometer frequency, with spectral width in both dimensions, ω_1 and ω_2, of 8000 Hz. The number of points was 1024 and 512 in ω_1 and ω_2, respectively. The spectra were zero-filled in the second dimension to 1024 points. The spectra in ω_1 were averaged over 32 acquisitions and 336 in ω_2. The delay time between the acquisitions was 2.0 s. Mixing time, τ_m, was set to 200 ms. The experiments were done with a Shigemi tube, 1.0 mM protein, 99.9% D_2O, at pHs: 3.0 and 7.0 .

Chapter 3

AAI peptide preparation

3.1 Introduction

Protein samples used in biophysical and biochemical studies of protein folding need to satisfy two major requirements. First, they should be available in substantial amounts of high purity. Second, they need to be ready for use in as short as possible time, because they are the starting point of time consuming experiments, which provide novel data about the folding phenomena. Solid Phase Peptide Synthesis (SPPS) is a technique able to provide several tens or hundreds of milligrams of peptide in a matter of several weeks and is therefore extensively used for preparation of peptides and short proteins.

It has been confirmed previously that synthetic AAI peptide from SPPS is able to refold *in vitro* and achieve its biological activity [19]. In the present chapter, the results from an improved synthesis approach are presented, which increased the overall yield of peptide synthesis and recovered a greater amount of the peptide after the refolding step. The details of the experimental strategies are given in Chapter 2, section 2.2.3.

3.2 Solid phase peptide synthesis

3.2.1 Results of AAI peptide synthesis

It can be anticipated from AAI's primary sequence and predicted by Peptide Companion software [120] that certain coupling steps in the synthesis would be more difficult, for example residues Cys1, Ile2, Lys11, Gly14, Val15, Tyr21 and Tyr27. This was in part confirmed, because the introduction of residues Cys1, Ile2 and Lys11 required double coupling steps. Apart from these residues, seven out of eight N-terminal residues required multiple couplings, probably because of the increasing aggregation potential of the growing peptide chain towards the end of the synthesis. Especially difficult was residue Lys4, which required a triple coupling. The amino acid derivatives used in SPPS of AAI peptide, their molar excesses and the times of coupling steps are summarised in Table 3.1.

The synthetic reaction was monitored during the individual coupling steps. Before attempting Fmoc-deprotection, Kaiser [92] or chloranil [93] tests were proven negative for all residues. This confirmed the absence of unreacted amines and the completion of the acylation reaction. After the introduction of residues Ser25, Thr22, Cys17 and Cys8, small amounts of the incomplete peptide were cleaved off from the resin, sidechains were deprotected and the crude mixtures were analysed with RP-HPLC and ESI-MS. The results from this analysis are seen in Figure 3.1A. The four peptides have 8, 11, 16 and 23 sequential C-terminal amino acids from the AAI sequence, respectively. They are all >90% pure and confirm the correct molecular masses. In Figure 3.1B, the RP-HPLC profile of the crude, full-sequence AAI peptide is shown. The molecular mass of the main peak corresponds to the completely reduced AAI peptide with six cysteine

step	derivative	couplings	molar excess	time [min]
1	Fmoc-Cys(Trt)-OPfp	1	4	60
2	Fmoc-Asn(Trt)-OH	1	4	85
3	Fmoc-Gly-OH	1	4	110
4	Fmoc-Tyr(tBu)-OH	1	4	60
5	Fmoc-Tyr(tBu)-OH	1	4	125
6	Fmoc-Asp(OtBu)-OH	1	4	120
7	Fmoc-Ser(tBu)-OH	1	4	60
8	Fmoc-Thr(tBu)-OH	1	4	60
9	Fmoc-Cys(Trt)-OPfp	2	4 + 2	60 + 60
10	Fmoc-Thr(tBu)-OH	1	4	90
11	Fmoc-Tyr(tBu)-OH	1	4	45
12	Fmoc-Pro-OH	1	4	90
13	Fmoc-Glu(OtBu)-OH	1	4	90
14	Fmoc-Cys(Trt)-OPfp	1	3	120
15	Fmoc-Cys(Trt)-OPfp	1	3	80
16	Fmoc-Pro-OH	1	4	110
17	Fmoc-Val-OH	1	4	70
18	Fmoc-Gly-OH	1	4	60
19	Fmoc-Asp(OtBu)-OH	1	4	80
20	Fmoc-Met-OH	1	4	90
21	Fmoc-Lys(Boc)-OH	2	4 + 4	90 + 30
22	Fmoc-Pro-OH	1	4	90
23	Fmoc-Gly-OH	1	4	90
24	Fmoc-Cys(Trt)-OPfp	2	3 + 3	120 + 60
25	Fmoc-Arg(Pbf)-OH	2	4 + 4	120 + 90
26	Fmoc-Asn(Trt)-OH	2	4 + 4	120 + 60
27	Fmoc-Trp(Boc)-OH	2	4 + 4	120 + 60
28	Fmoc-Lys(Boc)-OH	3	3 + 3 + 3	120 + 60 + 120
29	Fmoc-Pro-OH	2	4	120
30	Fmoc-Ile-OH	2	4 + 4	120 + 120
31	Fmoc-Cys(Trt)-OPfp	2	4 + 4	120 + 120

Table 3.1: **Amino acid derivatives used in solid phase synthesis of AAI**: All first couplings for cysteine residues (underlined), were done with -OPfp derivatives in the absence of base. The second round of couplings for Cys23, Cys8 and Cys1, were done with pre-activation with 2,6-dimethylpyridine (DMP) as base. Instead of using 100% N,N-dimethylformamide (DMF), the cysteine couplings were done in CH_2Cl_2 : DMF = 1:1.

thiols and its purity was 92.1%. Both the average yield per synthetic step and the overall experimental yield for this SPPS are very high, 99.9% and 92.1%, respectively. The exact quantification and yield calculation are given in Equations 3.1 to 3.10.

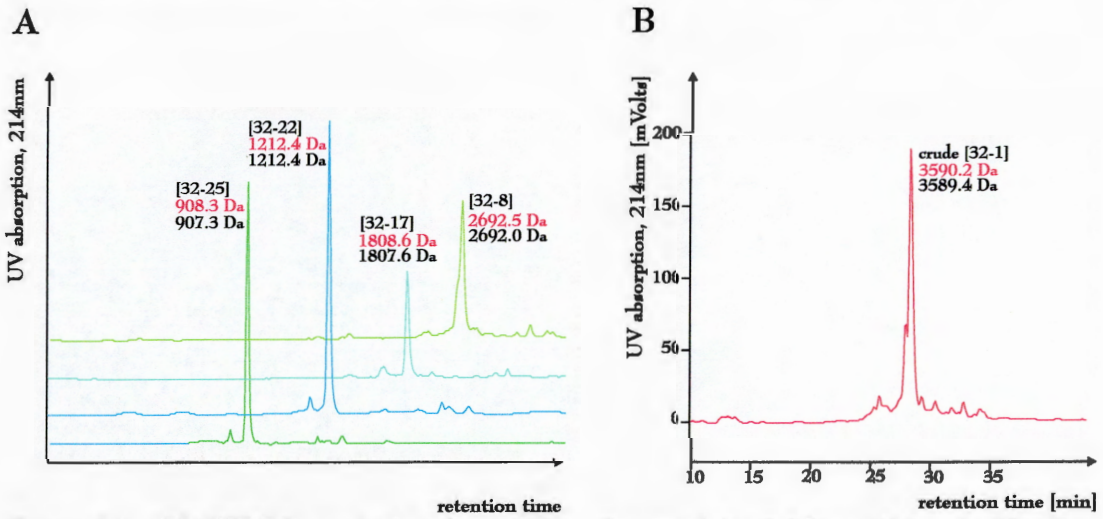

Figure 3.1: **RP-HPLC analysis of partial and complete AAI peptide:** A. The four partial cleavages of the incomplete AAI synthesis. B. Profile of the crude, complete sequence AAI. The amino acid sequences, experimental MW (red) and theoretical MW (black) are written next to the main peaks. The profiles were obtained with a RP C18 analytic column (5 μm) and a gradient of 0-20% B (5 min) followed by 20-30% B (30 min) at a flow rate of 0.8 ml min^{-1}.

$$\text{initial resin substitution} = 0.283 \text{ mmol / g} \tag{3.1}$$

$$\text{initial resin mass} = 2644.7 \text{ mg} \tag{3.2}$$

$$\text{synthesis scale} = 0.750 \text{ mmol} \tag{3.3}$$

$$\text{MR (protected AAI)} = 6658.5 \text{ mg / mmol} \tag{3.4}$$

$$-\text{ initial Fmoc-Ser(tBu)-OH on resin} = 383.5 \times 0.750 = 287.6 \text{ mg} \tag{3.5}$$

$$+\text{ final peptide weight} = 0.750 \times 6658.5 = 4994.3 \text{ mg} \tag{3.6}$$

$$\text{theoretical yield} = 7351.7 \text{ mg} \tag{3.7}$$

$$\text{experimental yield} = 6767.0 \text{ mg} = 92.1\ \% \tag{3.8}$$

$$\text{synthetic steps}^1 = 94 \tag{3.9}$$

$$\text{average yield per synthetic step} = \sqrt[94]{0.9205} = 99.9\ \% \tag{3.10}$$

[1]Three chemical reactions are repeated for each amino acid that is added to the peptide chain: deprotection, activation and coupling. An additional step is the Fmoc-deprotection of the N-terminal residue. Therefore for AAI, we have $3 \times 31 + 1 = 94$ steps. This number is used for calculation of average yield per synthetic step.

The impurities, which constitute ~8% of crude AAI peptide and include mainly side products of SPPS and racemised AAI products, were studied by LC-MS and ESI-MS. Almost 6% of the crude product is constituted by partial deletions of two to five sequential amino acid residues, see Figure 3.2B. The deletions are in places of the primary sequence that were observed as difficult synthetic points and required multiple couplings (Table 3.1). On the basis of a LC-MS experiment (data not shown), it was concluded that ~2% of the crude product underwent racemisation. This is a common side reaction in SPPS and is for AAI identified by the existence of a peak in the chromatogram immediately to the right of the main peak, with the same molecular mass. If we assume that racemisation happens mostly at the introduction of cysteine residues, this averages at ~0.3% per cysteine residue. This is the expected rate of racemisation for cysteines found in the literature [121]. The remaining impurities are below 0.5% and include for example the oxidation of methionine residue into methionine sulphoxide.

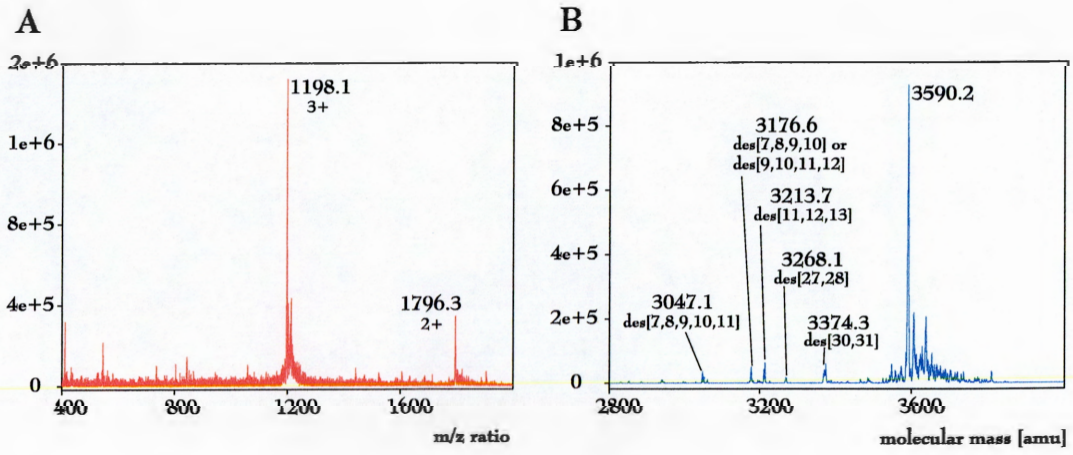

Figure 3.2: **ESI-MS of crude AAI peptide**: A. ESI mass spectrum of the crude AAI peptide B. The deconvoluted mass spectrum to show the molecular mass of the crude product, of which the RP-HPLC profile was shown in Figure 3.1. The side products of SPPS are identified with their molecular masses and the deletions in their sequence.

3.2.2 Improvement from previous strategies

In this section the improvements of the synthetic strategy are summarised with respect to the previously published material [19]. The greatest advantage of the present protocol is the substantial yield and purity increase of the crude peptide.

A major factor influencing the increase in yield is the minimisation of racemisation at cysteine residues. Racemisation in peptide synthesis is usually a base-induced side reaction, which can be prevented at the activation step of the coupling reaction. The mechanism of racemisation that is of importance to AAI synthesis is direct enolisation that occurs at the activation of the cysteine residues for coupling. Deprotonation by strong unhindered bases at the α-carbon results in racemisation because the carbanion can re-protonate on either side as is schematically shown in Figure 3.3. This is why the strong base N,N-diisopropylethylamine (DIEA), used for activation of the coupling reactions, was avoided for cysteine residues. It has been observed that racemisation is minimised when cysteine residues are introduced as pentafluorophenyl ester (OPfp) derivatives in the absence of any base [121]. The choice of OPfp ester was made because these derivatives are efficient acylating agents despite being slow, yielding products of high purity in efficient coupling reactions. As an alternative, when cysteines are introduced as free carboxylic acids, a weaker base, 2,6-dimethylpyridine, which has a significantly lower pK_a value than DIEA (6.7 instead of 11.4), was used. As a result, the racemisation rate was successfully minimised to below 2% (averaging at $\sim$0.3% per cysteine coupling step) from the previous 17% [19].

Another improvement in the strategy was the change in the side-chain protection of the cysteine residues from Acm to Trt, which enabled the deprotection and cleavage from the resin in one step. This is important because any additional steps in synthetic

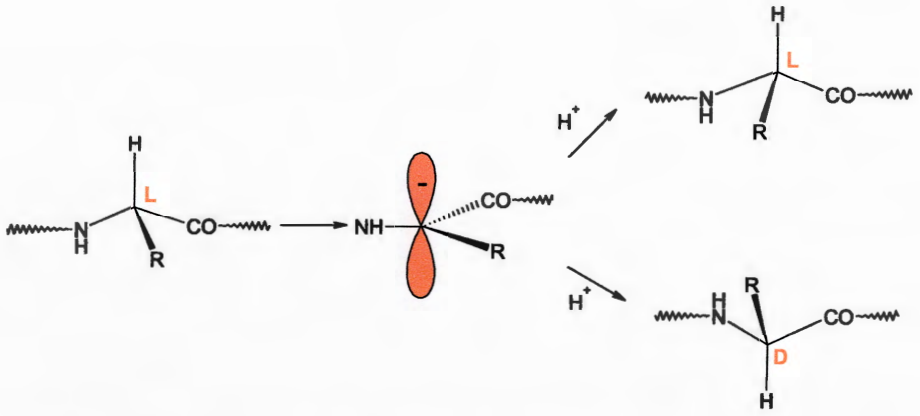

Figure 3.3: **Racemisation through direct enolisation pathway**: This mechanistic scheme shows the two possible pathways for the attack of carbanion after deprotonation of the α-carbon.

chemistry can cause product loss and lower overall yield. The Acm protecting group, which was used previously [19], requires deprotection with $Hg(Ac)_2$, which results in 5-10% aggregated (possibly dimerised) peptide. Additionally, the use of heavy metals can contaminate laboratory equipment and are noxious in case of environmental diffusion.

Further, in our synthetic scheme no significant oxidation of the Met5 residue to sulphoxide was detected (less than 0.5%), which is usually observed as a peak in the mass spectrum with molecular mass increase of 16 Da. This oxidation was prevented by the use of an inert atmosphere of N_2 throughout the synthesis reaction. Previously, ~8% of the crude peptide was observed with an oxidised methionine residue [19]. Overall, this optimised protocol increased the yield of the product from ~61% to 92.1%.

3.3 Folding and purification

3.3.1 Optimisation of oxidative refolding

Disulfide rich proteins, which are either chemically synthesised or obtained from recombinant expression, require an additional step in order to oxidise the cysteine residues into their native disulfide bonds. Oxidative refolding *in vitro* requires conditions, which

give a balance between the control of aggregation and the formation of the correct disulfide bonds. The methods to find the optimum conditions that give a better yield in a shorter amount of time are still relatively empirical. A systematic approach, which evaluates a matrix of conditions, is preferred.

The folding process of AAI peptide was studied in order to optimise the yield of the native protein recovery after about 20 to 24 hours. The folding conditions that were varied included: pH (7.25 to 8.50), AAI protein concentration (20 to 200 mg/l), temperature (4°C to 25°C), concentration of denaturant (urea or guanidine hydrochloride: concentration 0 to 2M) and redox reagents and their ratios (cysteine/cystine vs. glutathione reduced/glutathione oxidised, ratios 1:20 to 1:5). Apart from the pH value and the protein concentration, the conditions that gave the best renaturation yield of native AAI were confirmed to be the ones used previously by Lozanov et al [19]. In the following section the optimisation of pH value and concentration of AAI peptide are described.

The folding yield was first increased by determining the optimal pH of the oxidative folding buffer (Figure 3.4A). The yield of native AAI protein was compared after 20h of the folding reaction at various pH values, keeping all other conditions fixed. The pH values examined were from 7.25 to 8.50, which is the range of pH values found in the literature. The yield increased markedly at higher pH values; from ~60% of native AAI being recovered at pH 7.25 rising to over 93% at pH 8.50. Higher pH values were not taken into consideration because different phenomena, which are unfavorable for the recovery of pure native protein, are present at pH values over 8.50. At pH 8.70 for example, which is known to be close to the pK_a of cysteine thiols, a heterogeneous population of intermediates with different ionisation states for different cysteine residues

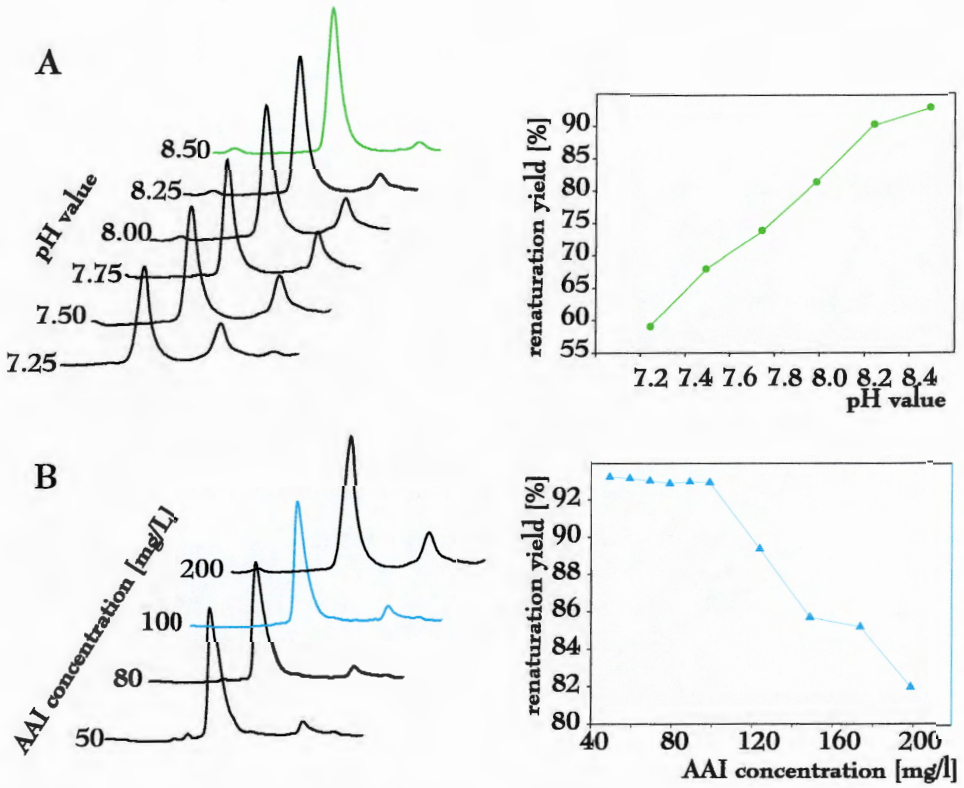

Figure 3.4: **Dependence of native AAI recovery on pH and concentration**: The renaturation yield of the native AAI protein was examined after 20 hours of reaction at different pH values (A) and different protein concentrations (B). Panel A represents oxidative folding at various pH values at a fixed protein concentration of 100 mg/l, whereas panel B. represents experiments at various peptide concentrations at a fixed pH value of 8.50. In both panels a graph of renaturation yield versus the changing condition is added for clarity. The remaining conditions were as follows: cystine:cysteine=0.05 mM :1 mM, 1.0 M GdnHCl, 0.1 M NH_4OAc, 2 mM EDTA, bubbled with N_2

can be detected [41]. The free thiols cause fast reshuffling of the native disulfide bonds, hence making the folding reaction more heterogeneous, erroneous and indirectly also longer. A particular disadvantage for cysteine containing peptides at high pH values is also the possibility of a β-elimination reaction, which can take place even at room temperature [122]. Another process to consider is racemisation, which can be accelerated at high pH (and high temperature) via the direct enolisation pathway (Figure 3.3) even in aqueous solution of deprotected, chirally pure peptides [122]. As a result, the pH value of 8.50 was chosen, giving >93% recovery of the native peptide after 20 hours

of oxidative folding.

Further, the maximum amount of protein per litre of refolding buffer was studied in order to increase the amount of production of the native protein. In Figure 3.4B, the renaturation yield of native protein over 20h of folding is examined as a function of AAI concentration. The yield decreases slowly as the concentration is raised from 50 to 100 mg per litre of refolding buffer, staying >90%. However, at concentrations higher than 100 mg/l, the renaturation yield is rapidly decreasing as the protein concentration is increased, showing a drop to a renaturation yield of just over 80% at 200 mg/l. In order to maximise the amount of AAI protein refolded efficiently per litre of refolding buffer and to minimise the effects of aggregation and polymerisation, the protein concentration was kept at 100 mg/l.

It is believed that protein concentration in the folding buffer is the main factor governing aggregation, which is often the cause of low recovery yields. Folding intermediates tend to aggregate more readily than native conformations through normally unexposed hydrophobic regions on their surface. In the kinetic competition between folding and aggregation, the latter is favoured at high protein concentration, because of its higher reaction rate order, estimated to be $\geq$2.5, whereas folding is usually a first order reaction [123]. Also, in a cysteine rich peptide such as AAI, high concentration may increase aggregation through dimerisation and polymerisation via formation of intermolecular disulfides. The inhibition of intermolecular interactions, which are responsible for aggregation, with the use of low molecular weight additives is an efficient strategy to increase renaturation yields [124]. The exact mechanism of the action of these additives is still unknown. They may influence the solubility and stability of the native, denatured and intermediate states, they may act by solubilising aggregates

or by changing the ratio of the rates for folding vs. aggregation. AAI peptide does not have major aggregation problems, when a small amount of additives is used. It was found that 1.0 M GdnHCl and 2mM EDTA were the optimum combination for preventing the aggregation in AAI oxidative folding, although other combinations and concentrations were examined (data not shown).

A summary of the final conditions used for the optimised refolding buffer, which give a final folding yield >93%, is given below:

1. AAI protein concentration: 100 mg/l
2. temperature: 25°C
3. pH value: 8.50
4. denaturant: 1.0 M GdnHCl
5. salts: 0.1 M NH_4OAc
6. redox pair, cysteine / cystine: 1 mM : 0.05 mM
7. metal chelator: 2 mM EDTA
8. bubbled with N_2: 20 min

3.3.2 Preparative purification of native AAI

The scale up process of the refolding protocol at optimised conditions to several hundreds of milligrams gave the same results for the yield of renaturation (>93%). Two sets of refolding experiments were done. The first set repeated the same procedure of refolding as described by Lozanov et al [19], whereas the second set of experiments was done under the optimum refolding conditions, described previously in section 3.3.1. Each sets of experiments repeated the folding starting from the crude AAI product and the purified reduced AAI, see Figure 3.5.

In Figure 3.5 the four RP-HPLC profiles of folding mixtures after 20 hours of reaction are shown. Interestingly, under the optimum refolding conditions, native AAI

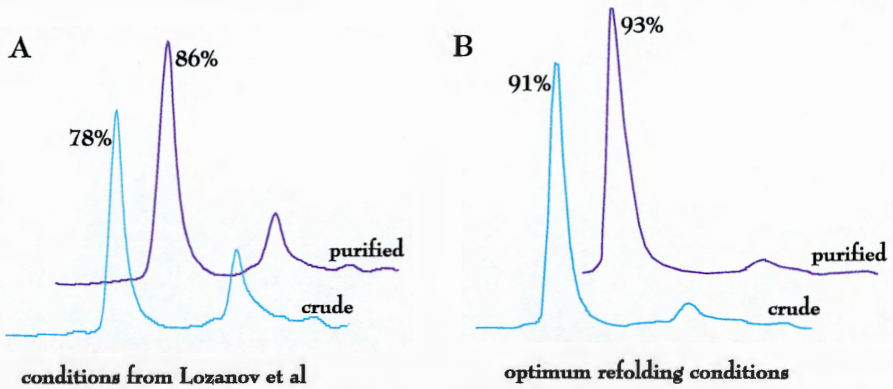

Figure 3.5: **Preparative AAI refolding**: Comparison of RP-HPLC profiles of the folding mixtures after 20h: A. folding conditions taken from Lozanov et al [19], B. optimum folding conditions, described in section 3.3.1. The blue chromatograms result from oxidative folding starting from crude reduced AAI, whereas the purple ones result from purified reduced AAI.

protein does not regenerate in a considerably greater amount if the starting material (reduced AAI) has been previously purified from the crude mixture. The difference in renaturation yields is only 2% (Figure 3.5B). Therefore, after the extraction of the crude product from ether, no further purification step is necessary before oxidative folding. Every additional step in product preparation lowers the overall yield, so elimination of an additional purification step which may cause a loss of 10% is favourable over the 2% lower oxidative folding yield.

On the other hand, the difference in renaturation yields between folding from crude or purified reduced AAI is 8% at the conditions described previously in the literature [19]. The main difference between the two folding protocols is a higher pH value of the optimum refolding protocol (8.50 versus 7.80). The reason for the greater difference in yield in non-optimised conditions could be a lower solubility of impurities that are present in crude AAI at the low pH. These species could cause more aggregation or promote the formation of non-native disulfides. It should be also noted that a major advantage for obtaining higher yields of renaturation described in this chapter was the better quality of the starting material, the crude reduced AAI. This is confirmed with

a ~16% increase in refolding yield when the experimental procedure was reproduced from [19] with our crude material (bottom chromatogram in Figure 3.5A).

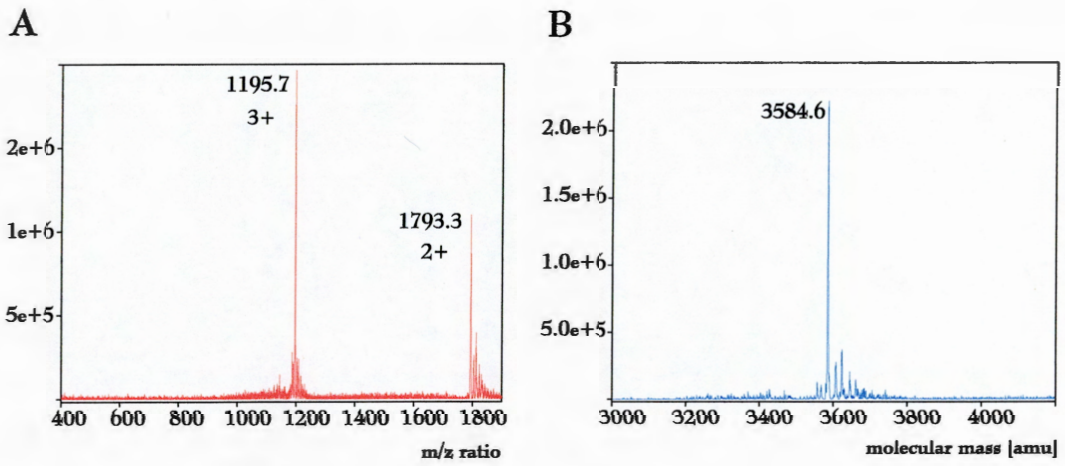

Figure 3.6: **ESI-MS of native AAI peptide**: A. ESI mass spectrum of native AAI peptide B. the deconvoluted spectrum on the right. The molecular mass of the product is shown and its RP-HPLC profile was shown as the bottom chromatogram in Figure 3.5B.

After preparative purification and isolation of the main peak on the refolding profile, ESI-MS confirmed the mass of the native product, which is shown in Figure 3.6. After lyophilisation, a small amount of native AAI was analysed by LC-MS, which demonstrated >99% purity. Finally, the disulfide bridge connectivity was proven by proteolytic digestion and ESI-MS (see section 4.3 and Table 4.2). The refolded AAI was lyophilised and kept at room temperature under a N_2 atmosphere and proved to be extremely stable, even over longer periods of months. Similarly, the purified reduced AAI peptide was also stable when purified and stored for longer periods of time. The reason for this inherent stability is also the absence of heavy metal ions, like Hg^{2+}, which even in trace amounts accelerate greatly the formation of disulfides, including the non-native ones. This can produce insoluble and heterogeneous products, which are prone to racemisation and other non-desirable modifications.

3.4 Conclusions

New synthetic and refolding strategies with high yields of native AAI protein were developed. The improvement of the SPPS approach increased the yield of crude, reduced AAI from the previous value of ~61% to 92.1%. This was done by minimising the racemisation at cysteine residues, decreasing the number of steps in the synthetic protocol and eliminating almost entirely the oxidation of the Met5 residue. Further, the matrix of oxidative folding conditions was optimised to give a maximum refolding yield >93%. It was also observed that the high quality of crude reduced AAI allows the oxidative refolding step to be started directly from it without further purification. The concentration of protein in the refolding buffer was increased, so that about 90 mg of purified native AAI protein is recovered per litre of refolding buffer.

Chapter 4

Oxidative folding intermediates

4.1 Introduction

A nascent linear protein chain has to adopt a series of conformational intermediate states before it can reach its native structure. This is why protein folding studies, including the ones presented in this thesis, are in part focussed on the detection and characterisation of the folding intermediates. One approach to study the intermediate states is oxidative folding, where a denatured, reduced, disulfide rich protein is allowed to refold in the presence of a redox buffer. In this composite process the protein regains both its native 3D structure and its native disulfide bonds [37]. An important attribute of an oxidative folding study is the ability to trap, isolate and hence characterise all disulfide intermediates present on the pathway, since they are covalently stable species. The intermediate conformations define the folding of a protein in terms of their heterogeneity, disulfide content and three-dimensional structures. The technique to track the formation and stabilisation of native disulfide bonds was pioneered by Creighton [40, 45], who trapped the folding intermediates of bovine pancreatic trypsin

inhibitor (BPTI). A further study of the BPTI folding pathway [41] improved this technique, which now involves acid trapping of disulfide intermediate species that are later separated with RP-HPLC.

A protein that contains three disulfide bridges has 75 different possible disulfide isomers, among which is also the native disulfide connectivity. These 75 isomers include 15 one-disulfide, 45 two-disulfide and 15 three-disulfide species. The starting point of oxidative folding is the completely reduced state, with 6 cysteine thiol groups, which needs to oxidise into a three-disulfide isomer with the native disulfide connectivity. In the present chapter, disulfide intermediate species that appear on the folding pathway of *Amaranthus α-amylase inhibitor*, are identified and characterised in terms of their disulfide content and heterogeneity. For five intermediate species, which have three disulfide bonds, the disulfide bond connectivities were determined by means of proteolytic digestion and mass spectrometry. Further, three "milestone" species of AAI oxidative folding were studied by means of mass spectrometry and circular dichroism to compare their molecular masses and molecular structure. As a final section in this chapter, a study on vicinal disulfide bonds is included, composed of a database search and molecular modelling.

4.2 Disulfide content of intermediate species

The folding pathway of AAI protein is very heterogeneous in terms of the content of disulfide structures which are present. Altogether, ten species could be isolated and detected in sufficient amount to be studied in terms of their disulfide content. Out of these, five species were observed to have one or two reduced cysteine residues and two disulfide bonds and the remaining five had three disulfide bonds. The distribution

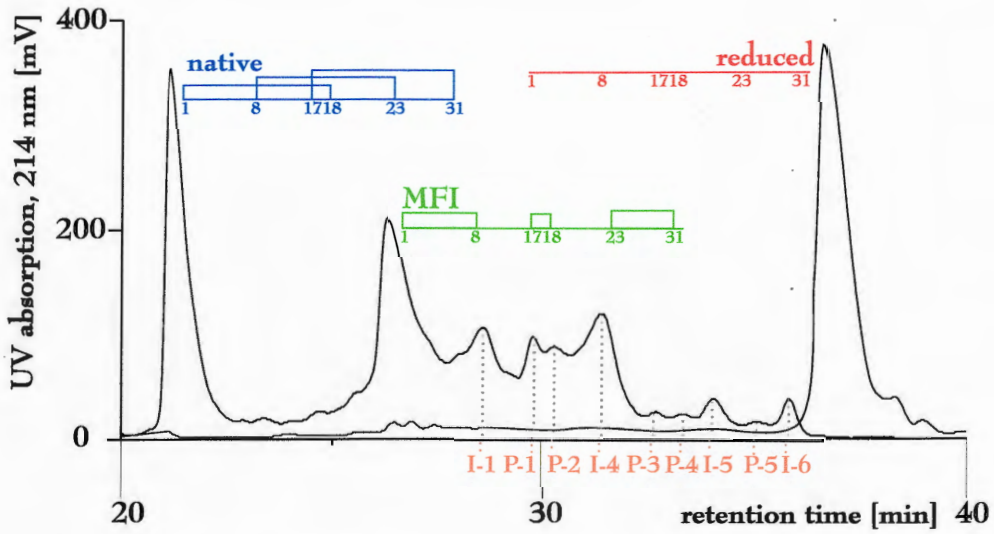

Figure 4.1: **Reversed-phase HPLC separation of intermediate species:** Two RP-HPLC chromatograms are shown: one is the separation of the folding mixture after 45 min (showing a heterogeneous mixture of intermediate species) and the second one after 1 min (showing mainly the reduced AAI) of the initiation of folding at the optimum conditions. The emphasis is given on the reduced, MFI and native AAI species. Their disulfide structures are drawn next to the peaks belonging to the corresponding species. Additionally, the elution positions of the remaining fully oxidised (I-1, I-4, I-5, I-6) and partially oxidised (P-1, P-2, P-3, P-4, P-5) intermediates are shown.

of fully oxidised (three-disulfide) intermediates (denoted with **I**) and partially reduced species (denoted with **P**) when separated with RP-HPLC is shown in Figure 4.1. At least four species on the folding pathway were detected to bind an exogenous cysteine molecule. In partially reduced species cysteine sometimes remains attached, because it is the reducing agent in the refolding buffer, which drives the isomerisation of non-native disulfide bonds (via thiol-disulfide exchange mechanism).

All ten intermediate species were first isolated and purified and then two separate experiments were done on each of the intermediate species in order to determine the number of disulfide bonds and free cysteine residues. First a simple molecular weight determination by LC-MS system was carried out, followed by an experiment involving iodoacetamide modification with subsequent LC-MS analysis. The results for molecular weights from both experiments are shown in Table 4.1.

species	mol. weigth (Da)		disulfides	free thiols	exogenous cysteine
	LC-MS	after C.M.			
native	3586.2	3586.1	3	0	no
MFI	3585.9	3586.1	3	0	no
P-1	3587.9 + 120.5	3702.5	2	1 or 2	yes
I-1	3586.3	3586.2	3	0	no
P-2	3588.4 + 120.5	3701.5	2	1 or 2	yes
I-4	3586.2	3586.2	3	0	no
P-3	3587.8 + 120.5	3702.7	2	1 or 2	yes
P-4	3588.4 + 120.4	3701.5	2	1 or 2	yes
I-5	3586.4	3586.1	3	0	no
P-5	3588.4	3703.5	2	2	no
I-6	3586.5	3586.8	3	0	no

Table 4.1: **Disulfide bond content of intermediate species**: all significantly abundant intermediate species were first subjected to an LC-MS analysis of their molecular weight (see "LC-MS" column). Later, they were carboxymethylated with iodoacetamide and analysed again with LC-MS (see "after C.M." column). In some species, exogenous cysteine was observed. Both analyses were done with a micro-bore C-18 column (3.5 μm) and a gradient of 0-20% B in 5min, followed by 20-27% B in 30min. The MS data were obtained at orifice +30 V in the positive ion mode.

Oxidative folding pathways should be distinguished and defined both in terms of heterogeneity, stability, structure and kinetics of disulfide folding intermediates on one hand and in terms of the developments in the three-dimensional conformations on the other hand. A complex diversity among oxidative folding pathways has been observed. For the model protein BPTI, for example, only one- and two-disulfide intermediates with native disulfides and native-like tertiary structures were observed [40, 41]. On the other hand, Chang and co-workers studied extensively the oxidative folding mechanisms of three-disulfide proteins and defined various distinctive mechanisms, which should reflect the distinctive ways the native disulfide bonds in proteins are stabilised [56]. One group of proteins sharing a common mechanism includes hirudin [47], tick anti-coagulant protein (TAP) [52] and potato carboxypeptidase inhibitor (PCI) [49, 50]. Dissimilarly to BPTI, these proteins have an oxidative folding mechanism with a high

heterogeneity of one- and two- and three-disulfide intermediates. The presence of three-disulfide species with non-native disulfide bonds is also a characteristic of this mechanism. In these proteins, there is also no evidence for predominant intermediates, which is on the other hand the case for epidermial growth factor (EGF) [47, 53], whose oxidative folding pathway rests between that of hirudin and BPTI. EGF's pathway comprises both of three-disulfide isomers like the one of hirudin as well as of a limited number of one- and two-disulfide intermediates, one of which is the predominant intermediate with two native disulfides (EGF-II). There have been other proteins like EGF, where the predominant intermediate is only partially oxidised and contains only native disulfide bonds (usually one less than the native protein), for example hen egg white lysozyme (HEWL) [63], EETI-II [125], the cyclotide kalata B1 [73] and insulin-like growth factor-I [69].

Putting the oxidative folding mechanism of AAI into perspective with other proteins, it is clear that it bears a significant resemblance to other small, three-disulfide proteins, such as for example the cystine-knot PCI. AAI presents a pathway with a high heterogeneity of one-, two- and three-disulfide intermediates, including fully oxidised intermediates involving only non-native disulfides. At the same time it clearly has an intermediate species, which prevails in abundance and has therefore been termed the main folding intermediate, MFI. This intermediate, dissimilarly to EGF, HEWL and others, is fully oxidised and contains three non-native disulfides.

4.3 Disulfide bonds in three-disulfide species

For a three disulfide bridge protein such as AAI, there are 15 possible disulfide arrangements with three disulfides, but only one of these combinations represents the native

protein pairing. Disulfide intermediate species, which have all cysteine residues oxidised are in some way off-pathway intermediate species. They all possess at least two disulfide bonds, which do not appear in the native species, hence termed *non-native disulfides.* The non-native disulfides have to undergo reduction before the correct cysteine residues pair into the native bonds. Some of these species are extremely abundant and can be thought of as an energy well on the protein folding energy surface, into which a certain activation energy needs to be put in to convert it to a partially reduced species able to further oxidise into the native disulfide bonds.

On the folding pathway of AAI, five species apart from the native protein have three disulfide bonds: MFI, I-1, I-4, I-5, I-6 (see Figure 4.2). These species were digested with trypsin and α-chymotrypsin simultaneously and the resulting fragments were separated and analysed by LC-MS. The masses and disulfide connectivities of the resulting fragments are given in Table 4.2. The disulfide connectivity of the native AAI protein was confirmed to have an *abc abc* topology as noted previously [20]. The disulfide bonds of the remaining five species were determined and the comparison of these is shown in Figure 4.2.

The two adjacent cysteines (Cys17 and Cys18) in the primary sequence of AAI and the absence of a specific reagent, which cuts between two sequential cysteine residues is the reason that it is sometimes impossible to distinguish between two combinations of disulfide bond topologies. This is the case for I-1 and I-4 species for which two topologies are noted in last column of Table 4.2 and in Figure 4.2. Interestingly, three species contain the vicinal disulfide bond Cys17-Cys18 and it seems this is the first time that an oxidative folding intermediate has been discovered with such a feature. This is why section 4.5 is dedicated to a discussion of the structural features

SPECIES	FRAGMENT	MOLECULAR WEIGHT		DISULFIDE
		experimental (charge)	theoretical	
N	[1-5] + [12-21] + [29-32]	2134.5 (1+)	2136.2	1-17 and 18-31 or 1-18 and 17-31
		1068.8 (2+)	1068.3	
	[1-4] + [12-21] + [29-32]	1948.8 (1+)	1948.3	
		974.5 (2+)	975.2	
	[8-11] + [22-27]	1090.7 (1+)	1091.2	8-23
		545.8 (2+)	546.1	
MFI	[1-4] + [8-11]	861.8 (1+)	862.1	1-8
	[22-27] + [29-32]	1066.5 (1+)	1067.1	23-31
		534.0 (2+)	534.1	
	[22-32]	1211.5 (1+)	1211.3	
	[12-21]	1111.8 (1+)	1112.3	17-18
I-1	[1-4] + [8-11] + [12-21]	1972.3 (1+)	1973.4	1-17 and 8-18 or 8-17 and 1-18
		987.0 (2+)	987.2	
		658.3 (3+)	658.5	
	[1-5] + [8-11] + [12-21]	1080.0 (2+)	1080.8	
		720.3 (3+)	720.9	
	[22-32]	1212.5 (1+)	1213.3	23-31
I-4	[8-11]+ [12-21]+ [22-27]	1100.8 (2+)	1101.8	8-17 and 18-23 or 8-18 and 17-23
	[8-11] + [12-27]	2184.3 (1+)	2184.5	
		1092.3 (2+)	1092.8	
		728.8 (3+)	728.8	
	[8-27]	1083.3 (2+)	1083.8	
	[1-4] + [29-32]	837.5 (1+)	838.0	1-31
	[1-5] + [29-32]	1023.8 (1+)	1025.2	
	[1-4] + [28-32]	1000.8 (1+)	1002.2	
I-5	[12-21]	1111.8 (1+)	1112.3	17-18
		556.3 (2+)	556.7	
	[1-4] + [22-27]	1146.8 (1+)	1147.3	1-23
		574.0 (2+)	574.2	
	[1-7] + [22-27]	802.8 (2+)	803.4	
	[8-11] + [29-32]	781.5 (1+)	781.9	8-31
I-6	[1-4] + [29-32]	837.5 (1+)	838.0	1-31
	[1-5] + [29-32]	1023.8 (1+)	1025.2	
	[8-11] + [22-27]	1090.5 (1+)	1091.2	8-23
		546.0 (2+)	546.1	
	[12-21]	1111.8 (1+)	1112.3	17-18
		556.3 (2+)	556.7	

Table 4.2: **Proteolytic digestion of three-disulfide species**: native AAI protein and 5 intermediate species were isolated by quenching the oxidative folding reaction after 30, 35, 40, 45 or 1200 min by acidification to pH ~2 with 4% TFA and separated by RP-HPLC. Over 98% pure species were incubated at 37°C in 0.1 M NH_4OAc solution at pH 5.2 for 44 h. The ratio of trypsin : α-chymotrypsin : protein was 1:1:20. The reaction was quenched with 4% TFA and the proteolytic fragments were subjected to analysis on a LC-MS system with a micro-bore C-18 column (3.5 μm), applying a gradient of 0-60%B (60 min) at a flow rate of 40μl min^{-1}. For some species, the two possible disulfide connectivities are given in the last column.

Figure 4.2: **Disulfide bonds in three-disulfide species**: a summary and a comparison of all disulfide bonds, which appear in three-disulfide species of AAI protein. Apart from the native protein, which was confirmed to have the *abc abc* topology, the prevailing disulfide bond appearing in three different species, is the vicinal disulfide bond Cys17-Cys18. For I-1 and I-4 the two possible arrangements are shown.

that a vicinal disulfide bond can induce and to the comparison of the existing protein structures in the Protein Data Bank with this feature.

On the other hand, it is interesting to examine which intermediates have one disulfide in common, because this means they are only one redox step away[2] from each other on the folding pathway. I-6 is the only species that has a native bond (Cys8-Cys23), which suggests that this species is one of the closest (in folding time and energy) to the native protein. Three other intermediates share one disulfide exactly with the I-6 species: MFI and I-5 have the vicinal disulfide Cys17-Cys18 and I-4 shares Cys1-Cys31. I-6 could therefore be a "funnelling" intermediate, since it is the species, which shares disulfides with the biggest number of species. I-1 is the only fully oxidised species, which does not share a disulfide with I-6. It is possible that I-1 has a native

[2]To get for example from MFI to I-6, the bridge 17-18 is retained, while the connectivities 1-8, 23-31 are swapped for 1-31, 8-23. Where only one such swap occurs and one disulfide bond is retained, the species are considered to be one redox step away.

disulfide, Cys1-Cys18, but this was not possible to confirm because of the ambiguity of the digestion strategy. I-1 also shares one disulfide with MFI (Cys23-Cys31) and could be connected to I-6 through this intermediate. More discussion on the kinetics of AAI folding is included in Chapter 6.

4.4 Three "milestone" species on the folding pathway

Three disulfide arrangements of AAI protein appear consecutively on the oxidative folding reaction and dominate each at a different point. At the start, the completely reduced peptide (R) with six thiol groups folds rapidly into a heterogeneous mixture of intermediate species. Soon after, a second species accumulates to over 30 % (at optimum folding conditions) and prevails over all other intermediates in abundance, which is why it was termed the main folding intermediate (MFI). The third species is the native (N) species with its native three-dimensional structure and disulfide pairing, which reaches its maximum abundance at the end of the folding reaction. These three species can be seen as the principal three milestones of AAI oxidative folding.

These three characteristic species are strikingly different in their cysteine residue pairings (see sections 4.2 and 4.3). The reduced species has no disulfide bonds, since it is the starting point of oxidative folding. The native protein was confirmed to involve a cystine knot motif, whereas MFI has disulfide bonds between cysteines, which are closest to each other in sequence, also called a "bead-like" structure.

These prominent species were subjected to a RP-HPLC study in order to determine their retention times, which would allow easy identification in further analytical HPLC studies (Figure 4.1). The three species were well separated (N, MFI and R eluted at 19, 24 and 35 min, respectively) with a linear HPLC gradient (20-27% B in 30min).

The retention time is a good indicator of the hydrophobicity. The most hydrophilic species is usually and in the case of AAI the native protein. This is to be expected, because it possesses both 6 oxidised cysteine residues and its three dimensional scaffold has optimised the shielding of hydrophobic groups from the solvent as is usual for native structures. The second most hydrophilic is MFI, because the three disulfide bridges with respect to the six thiol groups in R make a considerable difference in solvation energy.

Mass spectrometry of N, MFI and R

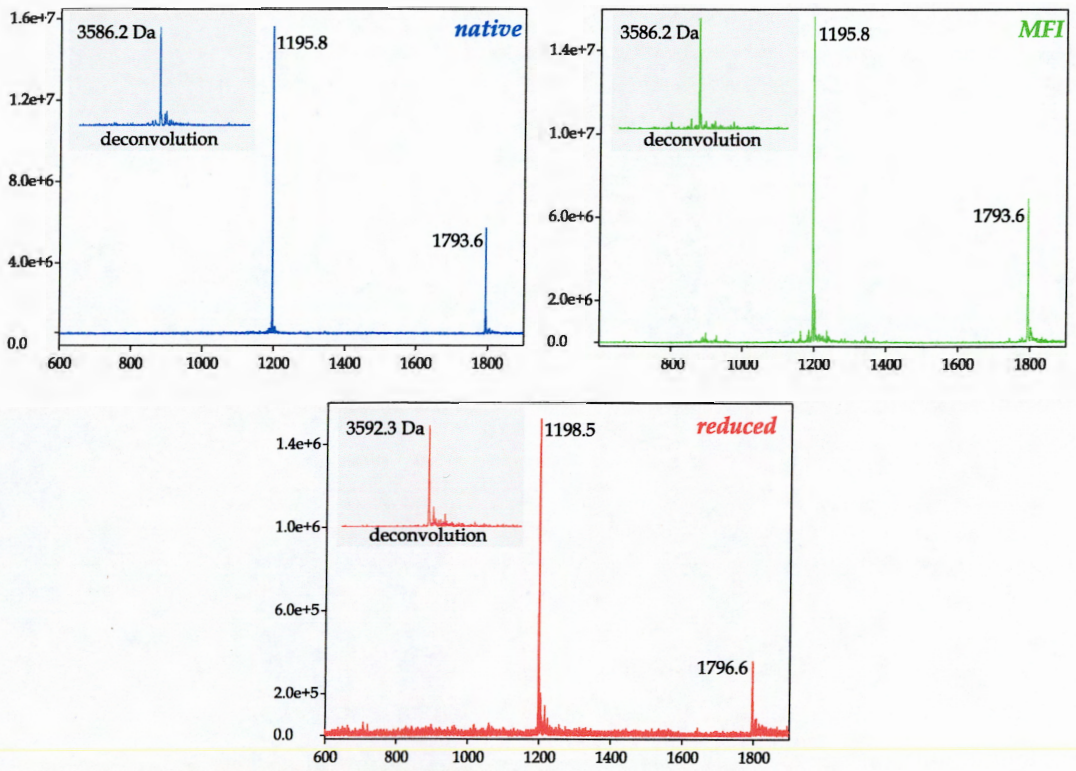

Figure 4.3: **ESI-MS spectra of native, MFI and reduced AAI species**: All three experiments were done in positive ion mode at orifice +40 V. The m/z ratio of each peak in the mass spectra is shown next to it and the molecular mass for each species is shown in the top left corner insert. BioMultiView software was used for deconvolution of the spectra.

The reduced, MFI and native AAI states were studied with ESI-MS to determine their molecular masses. The experiments were done in positive ion mode at pH 2, where

the species are invariably stable in terms of their disulfide content. At this low pH value the API ESI-MS instrument has a higher sensitivity and the basic groups in the protein are more readily protonated. In Figure 4.3 the ESI-MS spectra are shown, together with their deconvoluted spectra, which determine the molecular masses of the three species. The masses obtained are in excellent agreement with their theoretical values (for N and MFI 3586.2 Da vs. theoretical 3586.1 Da and for R 3592.3 Da vs. theoretical 3592.2 Da). Experiments were done at a variety of ionising conditions, but the MS spectra of the three species always appear very similar in terms of the distributions of the charge states. This is probably because in a such a small protein even the most compact conformation, the native protein, has most or all of the groups, which can be protonated, solvent exposed and therefore prone to ionisation. The four charge states in AAI protein observed at low pH (data not shown) are due to protonation on three side chain NH_2 groups (Lys4, Lys11 and Arg7) and N_α of the N-terminus. The +4 ionisation state was observed only at lower orifice values (15 to 40 V).

Circular dichroism of N, MFI and R

To be able to say more about the secondary and tertiary structure content, CD spectra (Figure 4.4) of the native, fully reduced and MFI state were compared both in the far-UV and the near-UV region. In the far-UV region (190-260 nm), the CD spectra of the reduced and the MFI species are very similar to the characteristic random coil state without significant secondary structure. The native state, on the other hand, has a non-random coil structure, although it does not show many properties of a β-sheet protein. This is probably because such a small protein has very short β strands (4 or 5 residues in each of the three anti-parallel strands). On the other hand, in the near-UV

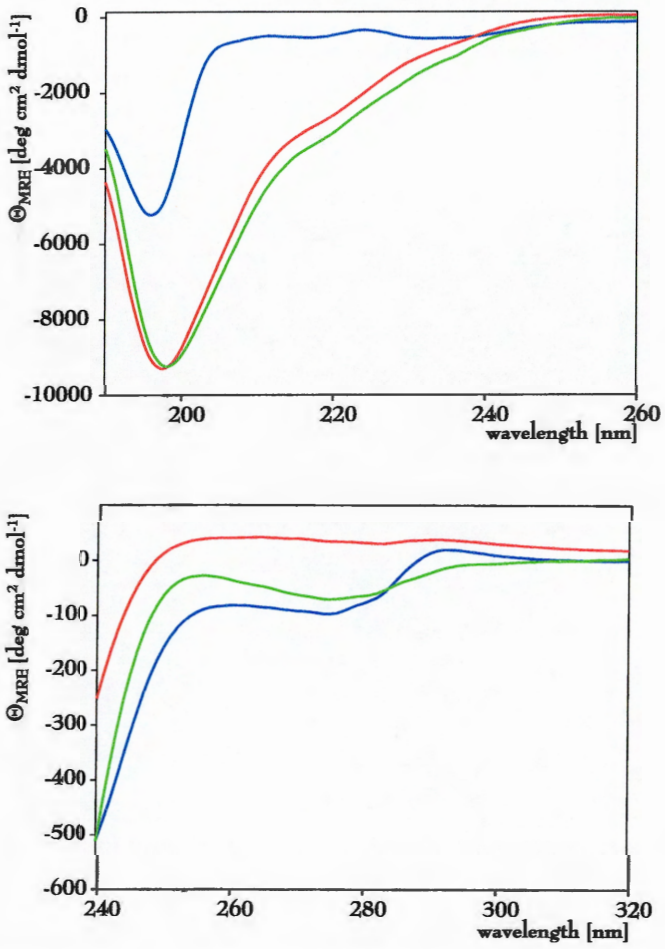

Figure 4.4: **Far-UV and near-UV circular dichroism spectra** of native (blue), MFI (green) and reduced (red) AAI species. The spectra were acquired with 1 nm step resolution, 50 nm min^{-1} speed, 1 second response time, 1 nm band width, 20 milli-degree sensitivity and 16 acquisitions. Cuvettes with 0.2 mm and 1.0 mm path-length were used in the far-UV and near-UV region, respectively. AAI protein concentration was 0.1 mM and 2 mM in the far-UV and the near-UV regions, respectively. The spectra were acquired at pH 7.0 and 25°C.

region (260-320 nm), the native and the MFI species had quite similar spectra, while the reduced showed again very little intensity, due to its random-coil like structure. These results indicate that in reduced AAI state there is an absence of regular structure. MFI possesses a disordered backbone, without a significant secondary structure content. At the same time, some of its aromatic side chains are constrained and together with the three disulfide bonds, this gives a near-UV CD spectrum, similar to the native state.

4.5 Vicinal disulfide bridges

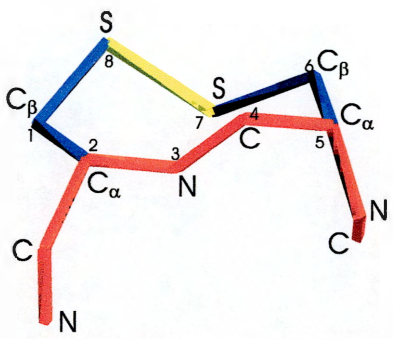

Figure 4.5: **Vicinal disulfide bridge**: a schematic representation of this structural feature made with Molmol software [14]. The backbone bonds are drawn in red and the side chains in blue. The disulfide bond is emphasised in yellow.

As mentioned earlier, three species (MFI, I-5 and I-6) appear to be the first observed oxidative folding intermediates possessing a vicinal disulfide bond. This is why this section is dedicated to the study of vicinal disulfide bonds, their structural features and the existence of proteins with this feature in the Protein Data Bank.

A vicinal disulfide bridge is a rare structural motif, represented by an eight-membered ring, composed of the main and side chain atoms of two adjacent cysteine residues (see Figure 4.5). Disulfide bridges usually link sequentially distant parts of polypeptide chains. It is therefore surprising to find covalent bonds between side chains of neighboring cysteine residues in some native proteins that have been structurally characterized (Figure 4.6). It can be anticipated that these rare structural motifs have a structural and/or functional role.

The Protein Data Bank was searched for experimentally determined protein structures containing two sequential cysteine residues. There are at present 18 published structures, where the two cysteines are oxidised into a vicinal disulfide bridge. It is a relatively rare structural element, so far observed in seven different types of proteins. They include bacterial redox enzymes, acid-base enzymes from bacteria, yeast, and mammals, plant insecticidal toxins and fish and molluscan receptors involved in signal transduction. Seven representative structures, each for one type of protein, are shown in Figure 4.6. They are J-Atracotoxin from *Hadronyche versuta* [126], methanol

dehydrogenase from *Methylobacterium extorquens* [127, 128, 129, 130, 131], the ethanol dehydrogenase from *Pseudomonas aeruginosa* [132, 133], acetylcholine-binding protein (AchBP) from *Lymnaea stagnalis* [134, 135], thioesterase I from *Bos taurus* [136, 137], carboxypeptidase T from *Thermoacinomyces vulgaris* [138] and anti-microbial peptide hepcidin-25 from *Homo sapiens* [139] (Figure 4.6,A-G).

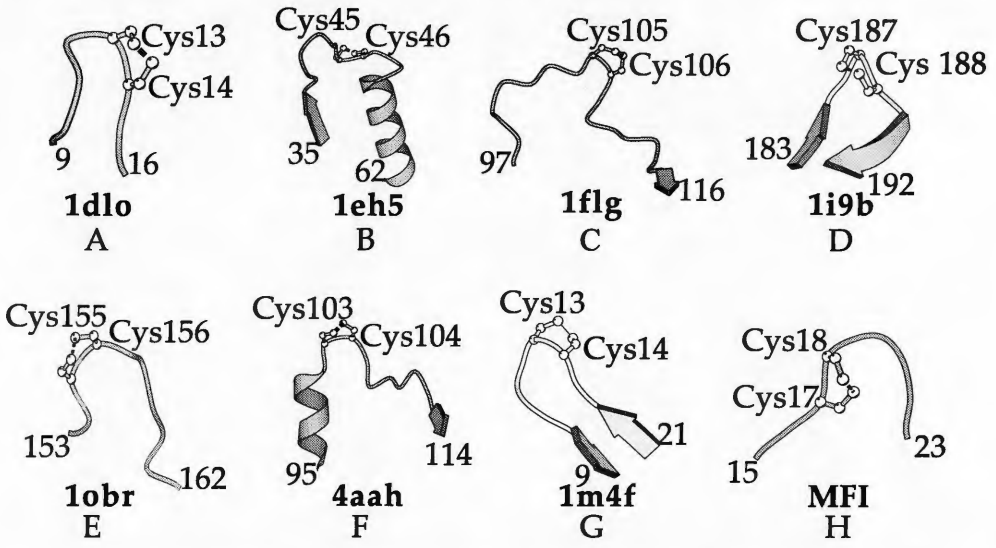

Figure 4.6: **Vicinal disulfide bonds in the Protein Data Bank**: Schematic view of the vicinal disulfide bonds, which have been structurally characterised. Under each structure the code of the Protein Data Bank file, which was used to make the image is written. A) 1dlo: J-Atracotoxin (*Hadronyche versuta*): C13-C14; B) 1eh5: palmitoyl protein thioesterase I (*Bos taurus*): C45-C46; C) 1flg: ethanol dehydrogenase (*Pseudomonas aeruginosa*): C105-C106; D) 1ib9 acetylcholine-binding protein, AchBP (*Lymnaea stagnalis*): C187-C188; E) 1obr: carboxypeptidase T (*Thermoactinomyces vulgaris*): C155-C156; F) 4aah: methanol dehydrogenase (*Methylobacterium extorquens*): C103-C104; G) 1m4f: hepcidin-25 (*Homo sapiens*): C13-C14; H) computer model of MFI: C17-C18 (see section 5.6).

On the other hand, there are at least 62 resolved structures, which include two sequential cysteines in a reduced form. If the structures of all octapeptides that contain two sequential cysteine residues are superimposed, a clear difference between the ones containing the reduced and the oxidised form is seen [140]. The ones containing the oxidised Cys-Cys pair are clearly bent and the two cysteine side-chains protrude on the same side of the backbone. On the contrary, protein segments that contain adjacent cysteines in the reduced state, possess an extended backbone and the sulfhydril side-

chains tend to be on opposite sides. The oxidized segments also seem to be more constrained than the reduced structures.

In the past, the conformational analysis of the oxidized Cys-Cys dipeptides suggested that the ring might lead to a turn when found in the middle of a peptide chain [141, 142, 143, 144, 145, 146]. The analysis of model peptides confirmed that the formation of the vicinal disulfide bond is accompanied by the appearance of a tight, β-type turn, and this phenomenon is also used in rational peptide design [126, 130, 131, 147]. The overall stereochemical features of the eight-membered heterocycle containing the disulfide bond of the structures in Figure 4.6 appear to be close to those observed in Ac-ox-[Cys-Cys]-NH_2 [148].

A systematic comparison was made between experimental and computational data on vicinal disulfide bridges in proteins and peptides [149]. It was found that the *ox*-[Cys-Cys] unit is able to accept types I, II, VIa, VIb and VIII β-turn structures, while without the vicinal disulfide bridge the Cys-Cys sequence unit is not constrained to form a turn. This is clearly seen in Figure 4.6, where all the protein backbones of the proteins containing a vicinal disulfide form a more or less tight turn. At the same time, the computer model of the main folding intermediate of AAI (see section 5.6) is shown for comparison in Figure 4.6H and it shows the possibility for a similar turn in the region of vicinal cysteine residues 17 and 18. The structural features are well conserved amongst the proteins reported in Figure 4.6, although the topology of the turn is much less conserved than in the case of model peptides.

The results may be useful in protein engineering, modelling, and design. It is also worth noting the possibility of using the redox-regulated transition from turn to extended conformation in nano-technology. The turn induced by the oxidation of two

adjacent cysteine side-chains, brings the residues preceding and following the Cys-Cys pair closer to each other. This fact could be used to construct spectroscopic probes, which would give redox-dependent signals. It is worth noting the recent study by Park et al [150], where the activity of ribonuclease A has been modulated through the insertion of two vicinal cysteines, which can be reversibly oxidized and reduced.

The biochemical relevance of vicinal disulfides is characterized in detail only for a few proteins, but little insight has been obtained about their role at the molecular level. Vicinal disulfide bonds can have both a structural role, by stabilizing the local backbone conformation, or a redox role, by allowing or assisting the electron transfer. In quinoprotein alcohol dehydrogenases [129, 128, 127, 132] and atracotoxin J-ACTX-Hv1c [126] the vicinal disulfide is proved to be essential for biological activity by site-directed mutagenesis studies. The atracotoxin J-ACTX-Hv1c, for example, keeps its three-dimensional structure unchanged after reduction of its vicinal disulfide bond, although its neurotoxic activity is completely lost. A vicinal disulfide bond is supposed to be a significant feature in the function of hepcidin, a polypeptide involved in iron uptake in the human intestine and iron release in macrophages [139]. The disulfide ring is also believed to be necessary in binding and reducing the cation in mercuric ion reductase [151].

4.6 Conclusions

The complex oxidative folding pathway of AAI peptide is very heterogeneous in terms of the disulfide content of its intermediate species. one-, two- and three-disulfide species are created as the reduced protein is allowed to refold to its native conformation. The three milestones of the folding pathway are identified to be the reduced, the MFI and

the native species. The most abundant intermediate species, MFI, has no significant secondary structural features, but seems to have some similarity in the orientation of the aromatic side-chains to the native protein. Non-native disulfides are formed in five fully oxidised intermediate species. One type of non-native disulfide, which has previously not been found in oxidative folding intermediates, is present in all three combinations that are theoretically possible. This is the vicinal disulfide between cysteines 17 and 18, which is found in MFI, I-5 and I-6, and is a surprisingly abundant feature of AAI's pathway. Vicinal disulfide bonds are found in at least seven wild type proteins, for which structures have been deposited in the Protein Data Bank. Structurally, such a disulfide is a turn inducing feature and is also thought to have an important functional role in several existing proteins. Some results from this chapter have been recently accepted for publication [140, 149].

Chapter 5

Structural studies of MFI

5.1 The main folding intermediate

The main folding intermediate (MFI) was named this way, because it is the most abundant species on the oxidative folding pathway of the AAI protein. This does not necessarily mean that MFI is the kinetically the most important or most productive species, a question which will be addressed in Chapter 6. The interesting question to ask is what kind of 3D conformation does a species with three non-native disulfides adopt. Does MFI's disulfide pattern, the so-called "bead-like" arrangement with a vicinal bridge, have a specific 3D fold, similar or dissimilar to the native protein?

Some data about MFI's structure have been obtained with circular dichroism in Chapter 4, which indicated that MFI does not have any secondary structure, but has some similarity to the native protein in the tertiary orientation of aromatic residues. The present chapter will deal with a more detailed structural characterisation of MFI's structure with NMR, photo-CIDNP and computer modelling techniques. Most results from this chapter have been recently published [152].

5.2 ^{1}H NMR and photo-CIDNP spectroscopy

The first step to elucidate MFI's three-dimensional structure was to compare one-dimensional NMR and photo-CIDNP spectra of the MFI and N species. The similarity of the circular dichroism spectra in the near-UV region suggest that some of the tertiary structure of the two species is similar. NMR spectroscopy was used to obtain information about the general conformational similarity and photo-CIDNP spectra to compare the accessibility of the aromatic residues and to try to assign the MFI NMR spectrum.

A detailed assignment of the MFI spectrum was done mainly with a 2D COSY experiment (section 5.3). However, a few guesses could be made on the basis of the multiplet patterns and phases of the amino acids peaks in the photo-CIDNP spectrum. For example, it was seen in Figure 2.8 that tyrosine H(3,5) doublet is the only emissive peak, which appears in the aromatic region of ^{1}H photo-CIDNP spectra. This led to the assignment of the region between 6.7 and 6.9 ppm in the MFI spectrum.

The 1D ^{1}H NMR spectrum of the purified MFI species (Figure 5.1A) resembles that of the N state: the chemical shifts are well dispersed and some of the lines are sharp, but the MFI-spectrum is clearly more complex. In addition to the major resonances, many of which have obvious counterparts in the spectrum of the N state, there are several weaker and broader signals close to the principal tyrosine and tryptophan lines. This is clearer in the 1D ^{1}H 750 MHz spectrum, where these additional signals, probably arising from minor conformations, are better resolved and denoted by asterisks (bottom panel in Figure 5.2).

AAI contains four aromatic residues (Trp5, Tyr21, Tyr27, Tyr28) whose photo-CIDNP spectra at pH 7 are compared in Figure 5.1B for N and MFI species. The

N spectrum is well-resolved, with chemical shifts characteristic of a single well-defined tertiary structure. The pattern of photo-CIDNP effects indicates that Trp5, Tyr27 and Tyr28 are exposed, but that Tyr21, which is not polarized and showed no photo-CIDNP signal, exists in a more crowded environment.

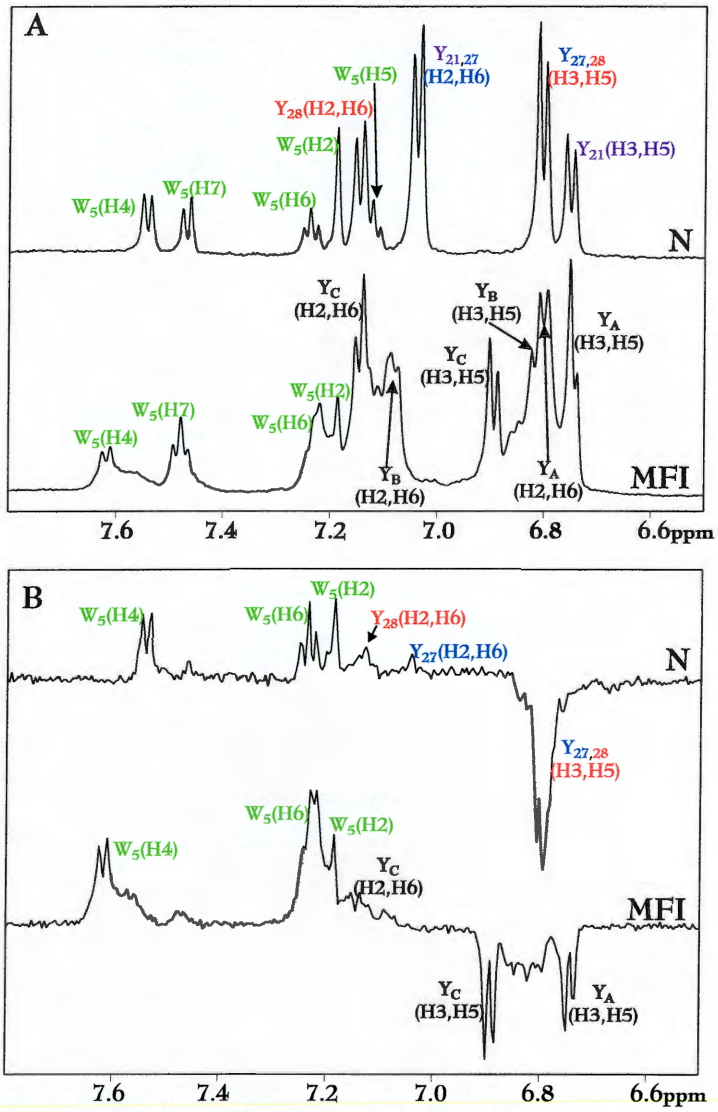

Figure 5.1: **600 MHz 1D ^{1}H NMR and photo-CIDNP spectra of native and MFI**: A. NMR and B. photo-CIDNP spectra of the purified native state (N) and of the main folding intermediate (MFI) (0.3 mM protein, 0.2 mM FMN, pH* 7.0, in D_2O at 25°C). NMR and photo-CIDNP spectra were recorded with 512 and 16 scans, respectively. The assignments of the native state tyrosine resonances are taken from [20]. The tyrosine resonances of the MFI species, labelled Y_A, Y_B, and Y_C, have not been assigned to specific residues. pH* is the measured pH of a D_2O solution, uncorrected for the deuterium isotope effect.

This result is consistent with the static accessibilities calculated from the three-

dimensional structure of AAI [153]. The NAccess programme gives relative accessibility of each residue calculated as the % accessibility compared to the accessibility of that residue type in an extended Ala-X-Ala tripeptide. The relative values for side chain atoms of the aromatic groups are expressed for all experimentally calculated structures given in the file of the Protein Data Bank. The average accessibility values of the ten structures in the 1qfd file (NMR structure of the free AAI peptide, [20]) for Trp5, Tyr21, Tyr27 and Tyr28 were 14.4, 2.4, 10.9 and 41.38, respectively. This means that the most exposed residue is Tyr28 and the least exposed Tyr21. This result indicates that signals in a photo-CIDNP spectrum will probably not be seen from Tyr21, which agrees with the experimental data.

The corresponding photo-CIDNP spectrum of MFI has enhanced signals from Trp5 and all three tyrosines; presumably the constraints on the backbone fold imposed by the non-native disulfides, allow the exposure of the side chain of Tyr21, itself in close proximity to the vicinal disulfide. Additionally, some of the subsidiary lines are also polarized, suggesting that the minor conformation(s) also possess similar exposure of aromatic residues.

5.3 Assignment of NMR spectrum for MFI species

In order to better resolve and to assign the peaks in the aromatic region of the spectrum of MFI, a two dimensional homonuclear ^{1}H COSY (COrrelated SpectroscopY) experiment was done. The experiment was performed at 750 MHz to obtain better resolution in the spectrum. In fact, if we compare the one dimensional spectra at 600 and 750 MHz, the difference in resolution is clear (see Figure 5.2). As noted before there are some smaller, broader peaks, probably belonging to minor conformations.

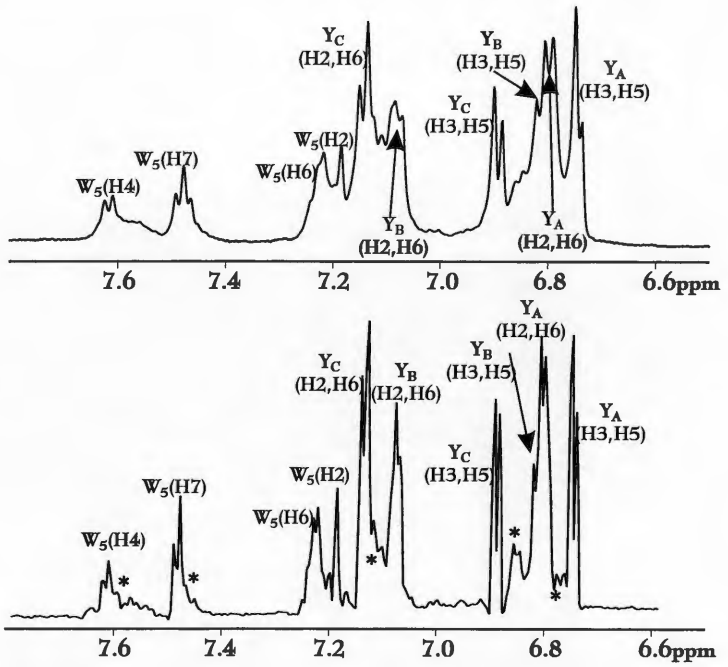

Figure 5.2: **1D ^{1}H NMR spectra of MFI species** at 600 MHz (upper spectrum) and 750 MHz (lower spectrum): the comparison of the two 1D spectra in the aromatic region, which consists of peaks from Tyr21, Tyr27, Tyr28 and Trp5 residues. The conditions of the two experiments were identical. MFI was dissolved at 1.0 mM concentration in D_2O, pH* 7.0, 25°C. Both spectra are averaged over 64 acquisitions, have 4096 points and were acquired with a 12500 Hz sweep width. The delay time between each acquisition was 1.8 s. It is clear that the 750 MHz spectrum is better resolved and the asterisk denotes peaks, which probably belong to minor conformations and were not resolved at 600 MHz.

In the 2D homonuclear 750 MHz COSY experiment on MFI species (Figure 5.3) the cross peaks in the aromatic region were partially assigned. The MFI aromatic spectrum was assigned with the help of photo-CIDNP experiments, which can be seen in Figure 5.1. Because a full sequential assignment was not possible from the NOESY spectra (see section 5.4), the three tyrosine residues were not distinguished from each other. The assignment procedure was done by first recognising the specific patterns of peaks (belonging to either a Tyr or a Trp residue), followed by linking the diagonal peaks to the off-diagonal peaks.

It was revealed that the signal from the H(2,6) protons of one of the tyrosine residues ("Y_A") has moved ~0.3 ppm from its position in the spectrum of the native

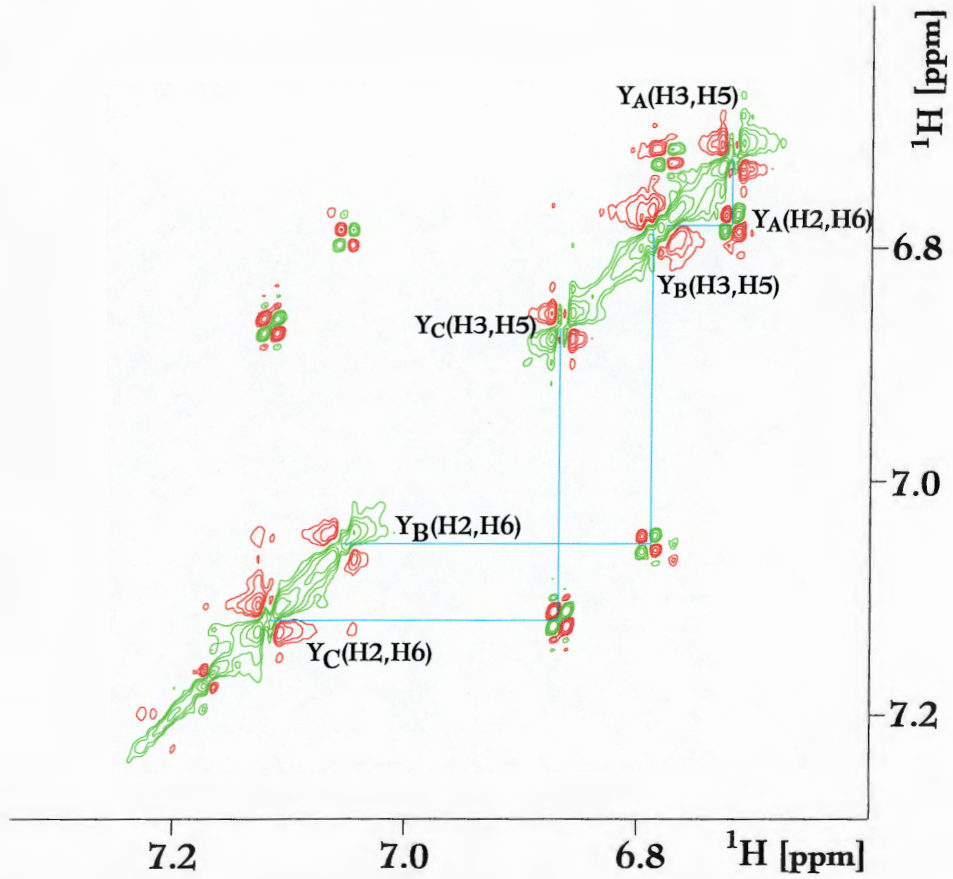

Figure 5.3: **750 MHz 2D COSY of MFI species**: This section of the aromatic region shows mainly H(2,6) and H(3,5) protons of the three tyrosine residues Y_A,Y_B, Y_C and some protons from Trp5. The spectral width was 6666 Hz, with 1024 complex points in both direct and indirect dimensions. Purified MFI was dissolved at 1.0 mM concentration in D_2O, pH* 7.0, 25°C.

state (7.0 - 7.2 ppm), to overlap with the H(3,5) resonances. This chemical shift, which gradually increased when the temperature was lowered to 4°C (data not shown), may arise from stacking of aromatic side-chains permitted by a changing backbone fold in the MFI-state. Closer inspection of this portion of the spectrum (6.7 - 6.9 ppm) revealed at least two weak doublets (denoted by an asterisk in Figure 5.2) in addition to four strong resonances: the three labelled H(3,5) signals and the shifted H(2,6) peak. The cross peaks in the two-dimensional COSY spectrum are surrounded by small, "satellite" cross peaks, which belong to the minor conformation(s) (Figure 5.3).

5.4 Tertiary structure comparison of N and MFI

The spectrum of MFI was partially assigned in its aromatic region by means of photo-CIDNP and 2D COSY experiments, and from 1D NMR spectroscopy it proved to have reasonably similar characteristics to the native protein. Attempts were made to determine the atomic resolution structure of MFI using 2D ^{1}H NMR spectroscopy via ^{1}H-^{1}H nuclear Overhauser effects (NOE) which reveal pairs of protons separated by less than ~5 Å. The principal experiment in such a process is two-dimensional Nuclear Overhauser Effect Spectroscopy (NOESY), which first requires the determination of an optimum mixing time value (τ_m) for each protein sample under investigation. The intensities of cross peaks at different τ_m values were plotted for both the native and the MFI species in Figure 5.4.

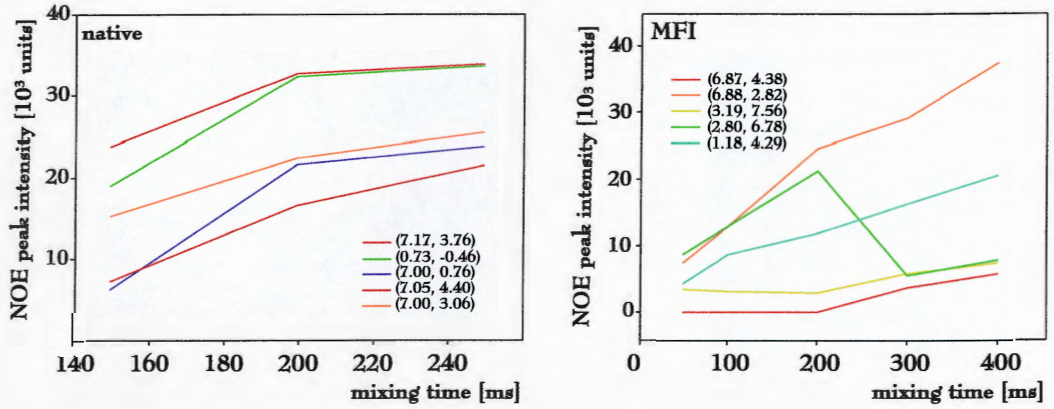

Figure 5.4: **Optimisation of mixing time for 2D NOESY experiment**: The intensities of five off-diagonal peaks in the native (left panel) and the MFI (right panel) species are plotted for different values of the mixing times, τ_m. Each cross-peak is identified with its chemical shift values in the first and second dimension, respectively: (δ_1,δ_2). Apart from τ_m, all other conditions were kept constant in all experiments. For experimental details, see captions of Figures 5.5 and 5.6.

The NOESY cross peaks for the native species get to their maximum intensity at ~220-250 ms of mixing time, so a safe value for τ_m is 200 ms, to avoid contributions from spin diffusion. For the MFI species it was more tricky to define an optimum τ_m

value, because the NOE cross peak intensities seem to vary irregularly with the τ_m values. This may be because of the dynamic properties of this extremely non-rigid species. The 2D NOESY experiment for MFI was performed at various τ_m values. However, most NOE's were observed at 200 ms, which is also the optimum value for the native species and hence the 2D NOESY spectra for the two species are compared at this τ_m value.

600 MHz NOESY spectra of the native and MFI states (Figures 5.5 and 5.6, respectively) show similar chemical shift dispersion, as noted for the 1D spectra in Figure 5.1, and many NOE cross peaks. Compared to the native state, however, the MFI spectrum was much less well resolved, showing a large number of overlapping and broader resonances. This indicates the presence of slowly inter-converting conformations with different side chain environments. Rapid inter-conversion between major and minor conformations would tend to average chemical shift differences, and lead to weak NOEs. Given the poor spectral resolution resulting from multiple conformations, an atomic resolution structure determination was not feasible.

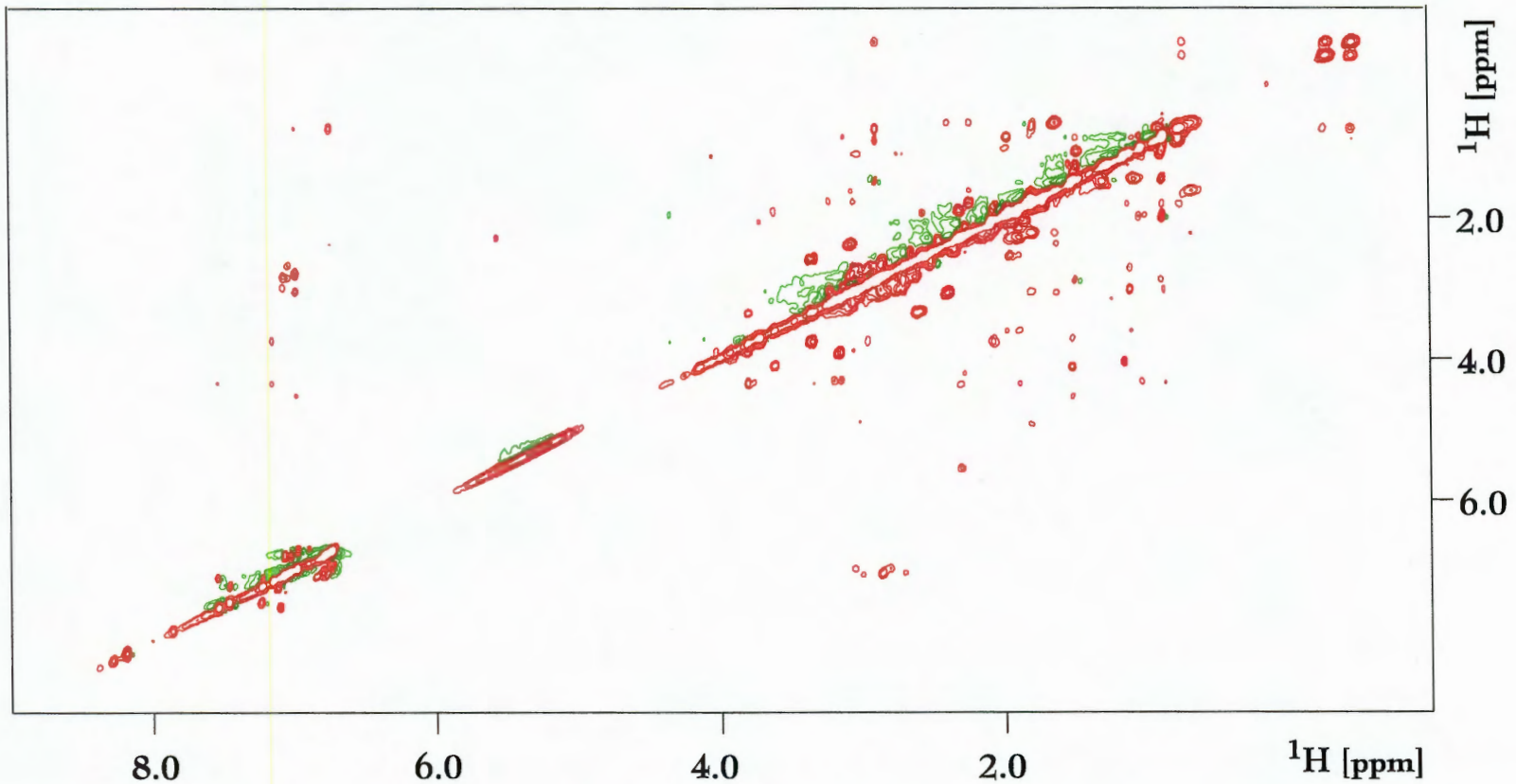

Figure 5.5: **600 MHz 2D ^{1}H NOESY of native AAI**: the spectrum was acquired with 200 ms mixing time, spectral width 8000 Hz in both dimensions, 1024 complex points in direct and 512 complex points in indirect dimension. The data were zero-filled in the second dimension to 1024 points.

Figure 5.6: **600 MHz 2D ^{1}H NOESY of MFI**: the spectrum was acquired with 200 ms mixing time, spectral width 8000 Hz in both dimensions, 1024 complex points in direct and 512 complex points in indirect dimension. The data were zero-filled in the second dimension to 1024 points.

5.5 Molecular dimensions of N and MFI

In characterising non-native states of proteins in solution, a measurement of molecular dimensions is of considerable importance. A measure of average dimensions assumed by the conformational ensemble can provide information about the nature of structures adopted by the polypeptide chain [32]. It also has the importance in understanding the role of conformationally disordered states in protein folding, stability and aggregation. Mostly, molecular dimensions have been given in terms of measurements of the effective radius of gyration by small-angle scattering studies, using neutrons (SANS) or X-rays (SAXS). The alternative is to determine the hydrodynamic or Stokes radii with measurements of translational diffusion coefficients, using dynamic light scattering or pulsed field gradient NMR techniques [91, 154, 155], the last of which was used in the study of MFI.

The molecular dimensions of N and MFI species were compared by performing NMR pulsed field gradient diffusion measurements. The approach focused on the estimation of the effective hydrodynamic radii of both species with a Pulsed Gradient Stimulated Echo Longitudinal Encode-Decode (PG-SLED) sequence [91, 156, 157]. In this experiment, the protein NMR signal decay, s, is measured as a function of gradient field strength, g, as indicated in Equation 5.1. The signal decay rate, d, is proportional to the diffusion coefficient, D, which is itself inversely proportional to the effective hydrodynamic radius, R_H. By adding a small molecule, in our case dioxane, as an internal radius standard [91, 154], absolute values for protein radii can be estimated as indicated in Equation 5.2.

$$s = A \cdot e^{-dg^2} \tag{5.1}$$

$$R_H^{protein} = \frac{D_{dioxane}}{D_{protein}} \cdot R_H^{dioxane} \tag{5.2}$$

Native	d(dioxane)	d(aliphatic)	d(aromatic)	d(tyrosine H3,5)
	7.75	1.35	1.26	1.40
	7.75	1.38	1.27	1.39
	7.77	1.45	1.29	1.44
	7.77	1.42	1.29	1.45
	7.78	1.29	1.23	1.36
	7.83	1.37	1.27	1.41
d(dioxane): d(native)		5.66 ± 0.23	6.13 ± 0.11	5.52 ± 0.13
total average	5.77 ± 0.32			
MFI	d(dioxane)	d(aliphatic)	d(aromatic)	d(tyrosine H3,5)
	7.74	1.13	1.14	1.28
	7.89	1.14	1.16	1.35
	7.84	1.30	1.19	1.38
	7.91	1.21	1.19	1.31
	7.86	1.33	1.24	1.36
	7.92	1.18	1.17	1.35
d(dioxane): d(MFI)		6.47 ± 0.41	6.66 ± 0.18	5.88 ± 0.14
total average	6.33 ± 0.41			

Table 5.1: **NMR diffusion measurements**: the table shows values for the signal decay rate, d, derived from the fitting of curves from NMR diffusion measurements. Each measurement was repeated six times and the result of each experiment is given in the table. Additionally, the first column for each species is filled with the values derived for the reference molecule: 2,4-dioxane. At the bottom of each column the average ratio d(dioxane): d(MFI) is given and below the total average ratio for each species.

The measurements with PG-SLED pulse sequence were completed by acquiring 20 spectra and varying the strength of diffusion gradient between 100% and 5% of its maximum value. Each measurement was repeated six times to improve the signal-to-noise ratio and allow the estimation of experimental error. The length of the diffusion gradient was optimised to give a total decay of protein signal between 80 and 90% at 100% gradient strength. Results are summarised in Table 5.1. The data for the diffusion coefficient, d, were obtained by integrating three regions of the NMR spectra from each species and fitting the resulting intensities to Equation 5.2. The regions

chosen for integration were a part of the aromatic region (6.70-7.70 ppm), a part of the aliphatic region (0.80-1.40 ppm) and a region of a tyrosine H(3,5) signal, which is well separated from other signals (7.00-7.08 ppm for the native species and 6.88-6.93 ppm for MFI species). The decay rates have been extracted in each experiment also for the reference molecule, dioxane. The signal decay rate, d, is, as noted above, proportional to D, the diffusion coefficient, which is in turn inversely proportional to the effective hydrodynamic radius, R_H. This is extracted from the Stokes-Einstein equation (Equation 5.3), where η is viscosity of the solvent. The ratio of R_H values for the native and MFI species is given in Equation 5.4.

$$D = \frac{kT}{6\,\pi\,R_H\,\eta} \tag{5.3}$$

$$R_H^N : R_H^{MFI} = d_{MFI} : d_N = 0.91 \pm 0.09 \tag{5.4}$$

Although absolute sizes of protein hydrodynamic radii must be interpreted with care [154], measurements of the relative sizes of similar species measured under similar conditions are much more robust. The ratio of effective hydrodynamic radii for the N and MFI species was 0.91 ± 0.09, and so the radii of the two species are indistinguishable within the experimental error.

5.6 Molecular simulations of MFI structure

A closer examination of the native AAI structure reveals that the six cysteine residues are near enough to each other that, in principle, all the 15 theoretically possible disulfide topologies can form via minor displacements of the main chain.

A three-dimensional structure of MFI was modelled with the Biopolymer and

Discover modules of InsightII. 10000 steps of the steepest descents method, followed by 1000 steps of the conjugate gradient technique were performed for energy minimization. The final root-mean-square (rms) deviation was smaller than 0.001 Å. Five random shakes of the starting atom coordinates resulted in identical final models. The root-mean-square-distances (rmsd) between the C_α atoms of the cysteine residues were calculated by optimal superposition of MFI's modelled structure with the native AAI.

The average rms distance calculated for cysteine C_α atoms with respect to the native structure is 1.72 Å. The MFI model superimposed onto the experimental native structure is seen in Figure 5.7. It suggests a globular, compact structure resembling the native state. It should be emphasized that modelling of this sort shows that stereochemically allowed structures with non-native disulfide bridges can be generated from the native tertiary scaffold of AAI protein, but not that these structures are necessarily observed. Energy minimization methods, which were used to generate the model are known to be unreliable at predicting loop structures. However, the region of the vicinal disulfide bridge in the model of MFI features a turn reminiscent of those found in the protein structures containing this feature (Chapter 4).

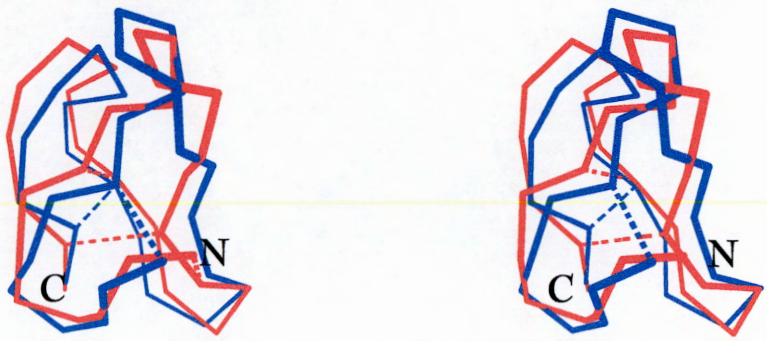

Figure 5.7: **Molecular model of MFI**: Stereo-view of the superposition of MFI species (blue) native AAI conformation (red). The N- and C-termini are denoted. The disulfide bonds are drawn in a dashed line.

5.7 Conclusions

Although it proved impossible to determine an atomic resolution structure of MFI, a consistent picture nevertheless emerges from the experiments. The structure is as compact as the native state, as indicated by the NMR diffusion measurements. There is little evidence from far-UV CD for secondary structure, but at least some of the side chains are ordered, as evidenced by the near-UV CD, the number of NOEs, the chemical shift dispersion and the large change in the chemical shift of the H(3,5) resonance of one of the tyrosine residues from its position in the native state. The additional lines visible in the NMR spectra, responsible for the poor resolution of the NOESY cross peaks, point strongly towards slowly inter-converting multiple conformations of the polypeptide backbone. Modelling suggests that the non-native disulfides in MFI do not constrain the backbone to the extent that native-like conformations of the backbone and native-like side chain interactions cannot exist. Nevertheless, the side chain exposures revealed by CIDNP and the differences in chemical shifts show that there are clear differences between the tertiary structures of the native and MFI states.

Chapter 6

Kinetic analysis of oxidative folding

6.1 Introduction

The study of kinetics of a folding process is essential for the understanding of the formation of the native structure. The kinetics of oxidative folding of AAI protein was studied to examine the stabilisation of the native disulfide bonds *in vitro* in different folding conditions. In addition, oxidative folding *in vivo* was mimicked with the addition of a protein disulfide isomerase.

This chapter deals both with the kinetic process at the level of a single disulfide species and at the level of a single amino acid residue. Two different time-resolved techniques were used to follow the oxidative folding and their results were compared to examine the credibility of application of such techniques to oxidative folding. The kinetic data were analysed, fitted, plotted and the rate constants were calculated using Sigma Plot software.

6.2 Oxidative folding at the level of a single species

The oxidative folding process of AAI was first studied on the level of a single disulfide species. The intermediate species, which take part in the folding process, were all previously characterised in terms of their disulfide content (see Chapter 4). They are covalently stable species, which are efficiently separated with an analytic RP-HPLC technique at acidic pH [41]. A time-resolved analysis of the transformation on the level of disulfide species was done in different folding conditions.

6.2.1 Optimum oxidative folding conditions

The first experiment was at the optimum oxidative folding conditions, where over >93% of the native protein is regenerated in 20 hours. The presence of disulfide species at different times after the start of folding was examined by removing aliquots from the folding mixture at selected time points and analysing them with RP-HPLC. Figure 6.1 shows time-resolved chromatograms (A) and the abundances of disulfide species with respect to the folding time (B). It is clear from the chromatograms that the folding pathway is dominated by a few disulfide species, among which are especially prominent the five fully oxidised species with three disulfide bonds.

The optimised oxidative folding of AAI is characterised by a high heterogeneity of one-, two- and three-disulfide species as discussed in detail before in Chapter 4. The rapid disappearance of the reduced species is fitted well to a 1^{st} order exponential decay (R^2 >0.99, where R is the multiple correlation coefficient) with a rate constant of 0.14 min^{-1}. The native species accumulates with a lower rate constant and is fitted well to a 1^{st} order exponential rise curve with a rate constant of 0.0098 min^{-1}. Non-native disulfide species (i.e. MFI, I-1, I-4, I-5, I-6) are present on the pathway of oxidative

folding. It may be that they play a specific role in directing the folding process through structural information that they contain.

It is clear that the first non-native disulfide intermediate that accumulates is I-6, which is most abundant at about 5 minutes after the start of the folding. It is also the only fully oxidised intermediate on the folding pathway, which contains a native disulfide bond (Cys8-Cys23). It could therefore be a close precursor of one- and two-disulfide intermediate(s), which are in turn direct precursors of the native protein. Soon afterwards, at about 15 minutes after the start of folding, three other fully oxidised species (I-1, I-4 and I-5) reach their maximum abundance at almost the same time. The last three-disulfide intermediate to accumulate is MFI, which reaches its maximum abundance of almost 40% at 30 minutes. This means that reshuffling and isomerisation of non-native disulfides of this intermediate is a slow, rate-determining step in the process.

6.2.2 Oxidative folding in denaturing conditions

The fully reduced AAI protein was refolded in the presence of different concentrations of guanidine hydrochloride (GdnHCl), which was used to disrupt the native structural preferences and interactions. The refolding reaction was monitored by time-resolved RP-HPLC. The chromatograms obtained from the folding mixtures 24 hours after the initiation of the folding reaction, containing from 0 to 6.0 M GdnHCl are compared in Figure 6.2. The reactions in the absence of GdnHCl and in its presence at 1.0 M concentration yield very similar results, reaching over 90% of native AAI product. As the concentration of GdnHCl is gradually increased to 6.0 M, the amount of native species, which is obtained after 24 hours of refolding, decreases rapidly and native

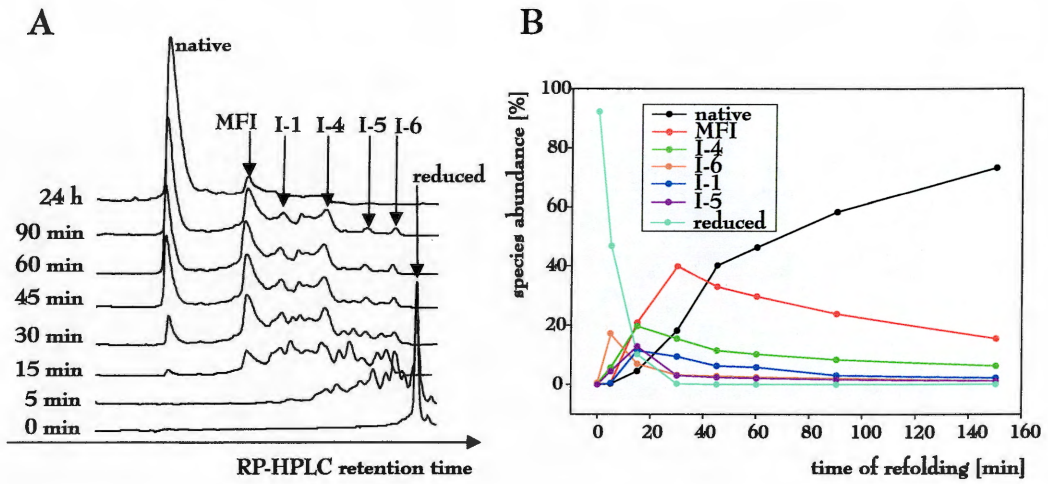

Figure 6.1: **Oxidative folding at optimum conditions**: 100mg/l AAI, 0.1 M NH_4OAc, 2 mM EDTA, 1.0 M GdnHCl, pH 8.50, 1mM cysteine, 0.05 mM cystine, 25°C. A. Time-resolved chromatograms of the folding reaction, isolating all disulfide intermediate species, which appear on the pathway. Chromatographic profiles were obtained with a RP C18 analytic column (5 μm) and a gradient of 0-20% B (5 min), followed by 20-27% B (30 min) at a flow rate of 0.8 ml min^{-1}. B. The abundances of all fully oxidised species and the fully reduced protein are plotted versus the refolding time. Their abundances are taken from the intensities of the RP-HPLC peaks in A. and are calculated from their absorbance at 214 nm.

AAI represents only 38% of all protein conformations at 6.0 M GdnHCl. At the same time, the other three-disulfide species (MFI, I-1, I-4, I-5 and I-6) are becoming more abundant, especially MFI, which reaches 32% at 6.0 M GdnHCl. This species is clearly favoured over the other three-disulfide species, whose abundances never reach more than 10% per single species.

Denaturants are an important part of oxidative folding buffers, because they prevent aggregation and hence significantly increase the yield of the native protein. However, the concentration of denaturant is usually kept low (from 0 to 2 M urea or from 0 to 1 M GdnHCl) in order to balance aggregation and the formation of the native structural elements, such as the native disulfide bonds. In fact, the optimum refolding conditions for AAI oxidative folding involve 1.0 M GdnHCl in order to prevent aggregation and at the same time provide a high yield of the native conformation.

When the concentration of denaturant is increased, other disulfide connectivities might prevail or become significantly abundant. This is because in absence of a denaturant, secondary and tertiary preferences encoded in the amino acid sequence enable efficient conversion to the close packed native structure, which promotes the native disulfide pairing [25]. It is highly unlikely that a native 3D array of the backbone would allow two different disulfide pairings. At the same time, in small proteins such as AAI the native-like 3D conformations are not unlikely to be able to accommodate non-native disulfide bonds and non-native disulfide topologies.

It has been confirmed in recent years that even proteins under highly denaturing conditions with reduced disulfide bonds keep a non-random structural preference for formation of secondary and tertiary structures [158]. Also, hydrophobic interactions persist under strongly denaturing conditions [159] and compact molten globule structures are present in the absence of disulfides [160]. A similar phenomenon is observed as the concentration of denaturant is increased in the oxidative folding of AAI. It does not lose the primary sequence's preference for the formation of the native structure and native disulfide bridges, which remains the most abundant conformation.

At the same time, the probability of formation of disulfides in an unstructured chain decreases with the number of residues between the two cysteines [161]. Therefore under conditions, where some secondary and tertiary interactions are disrupted, such as in the presence of a high concentration of a denaturant, the vicinal disulfides should exhibit the highest probability of formation. In the same way, the "bead-like" structures - in which disulphide bridges are formed from the cysteines nearest to each other within the sequence - have the highest probability of formation [48, 162]. Out of the three intermediates with a vicinal bridge, which would be favoured *per se*, the most

abundant, MFI, is indeed the one with the "bead-like" structure. It is also well known that single domain proteins featuring primarily sequence-local contacts (high "contact order") tend to fold with lower energy barriers than those characterized by more non-local interactions [163]. This is consistent with the accumulation of a significant amount of MFI.

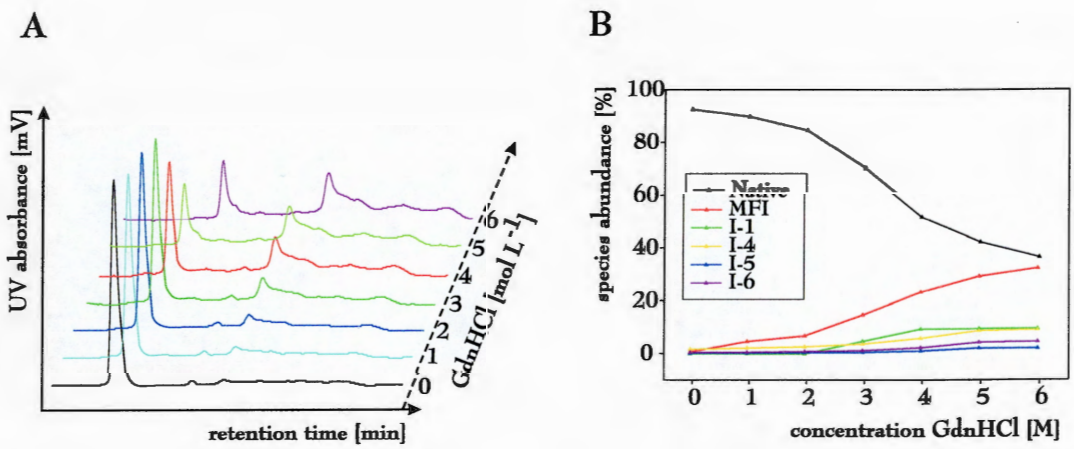

Figure 6.2: **Oxidative folding in the presence of guanidine hydrochloride**: A. Time-resolved RP-HPLC chromatograms are shown for the folding mixtures after 24h of oxidative folding. The oxidative folding conditions are kept constant (100mg/l AAI, 0.1 M NH_4OAc, 2 mM EDTA, 1.0 M GdnHCl, pH 8.25, 1mM cysteine, 0.05 mM cystine, 25°C), while GdnHCl concentration is varied from 0 to 6.0 M. The profiles were obtained with a RP C18 analytic column (5 μm) and a gradient of 0-20% B (5 min) followed by 20-30% B (30 min) at a flow rate of 0.8 ml min^{-1}. B. The abundances of the three-disulfide species after 24h of refolding are plotted versus the concentration of GdnHCl.

6.2.3 Reductive unfolding from native and MFI species

The mechanism of reductive unfolding of AAI protein was analysed. Purified native AAI protein was isolated and placed into a reducing buffer containing various concentrations of dithiothreitol (DTT) at pH 8.50. The unfolding experiments were quenched in a time course manner and analysed by RP-HPLC. The results show that the reduction of the native species leads to the direct conversion to the fully reduced species, without

accumulation of intermediate one-, two- or three-disulfide species. This phenomenon of a concurrent reduction is observed for different concentrations of the reducing agent, ranging from 5 to 100 mM. In Figure 6.3A, the time-resolved RP-HPLC chromatograms are shown for the reduction of native AAI with 5 mM DTT. The rate constant for the conversion of native to the reduced species ($k_{N \Rightarrow R}$), which was extracted for four different DTT concentrations, shows a linear dependence upon the concentration of DTT, as seen in Figure 6.3C. The unfolding pathway was also examined with TCEP as a reducing agent (pH 3.0) and was found mechanistically unchanged.

Because AAI either exists in the native three-disulfide topology or is reduced directly to the reduced species, it follows the *all-or-none* reductive unfolding mechanism. This mechanism is shared also by other small cystine-rich proteins, such as hirudin, PCI and TAP [56]. The native disulfides of such proteins are stabilised in a concerted and interdependent manner, because they do not allow for intermediate species when the native disulfide bonds are reduced. Another important observation is that these proteins, which share a common reductive unfolding mechanism, also have the same characteristics in the oxidative folding pathway. AAI and the above mentioned proteins all have a high heterogeneity of one-, two- and three-disulfide species, including non-native disulfide isomers.

In contrast to the reductive unfolding of the native species, the mechanism of MFI for the same process was studied by LC-MS (Figure 6.4B) and it turned out to be very different. MFI forms two two-disulfide isomers and three one-disulfide isomers as intermediates during reductive unfolding (Figure 6.4B). The retention times of these isomers do not coincide with those of any other species on the oxidative folding pathway from reduced AAI, so they probably contain only MFI-disulfide bonds and are formed

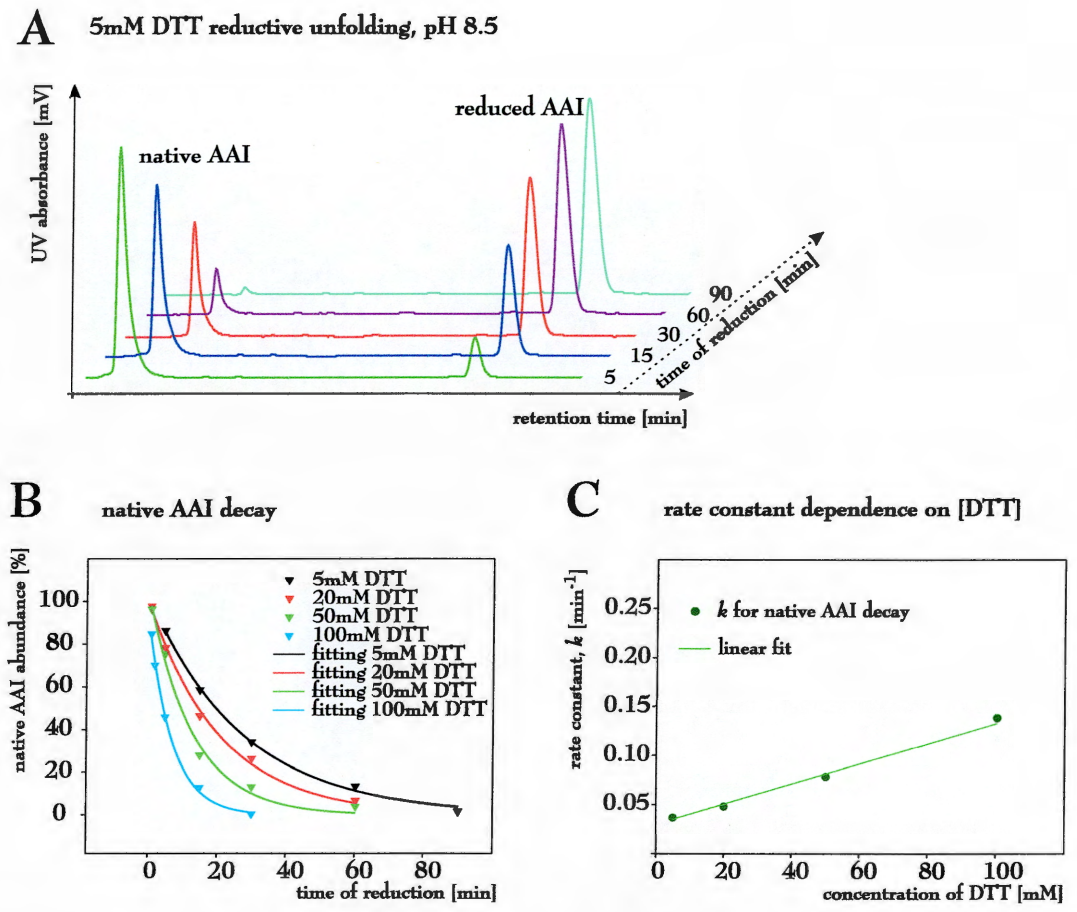

Figure 6.3: **Reductive unfolding pathway of native AAI**: A. Time-resolved RP-HPLC profiles for reduction with 5 mM DTT at different times after the initiation. The reverse process of oxidative folding starts from purified native AAI species in the presence of a reducing agent, unfolds and gets reduced to form the fully reduced state (conditions: 0.5 mg/ml AAI, 0.1 M NH_4OAc, pH 8.50, 25°C). B. Native AAI decay at different concentrations of DTT (5, 20, 50 or 100 mM DTT) against reduction time. The decay is fitted to a 1^{st} order exponential decay curve for each DTT concentration. C. The rate constants for native AAI decay (extracted from B.) plotted versus DTT concentration.

when each MFI-disulfide is reduced in turn. This points to the fact that the disulfide bonds of MFI are not interdependent and are stabilised independently, probably because of the preference for forming disulfides between cysteines, which are closest to each other in sequence.

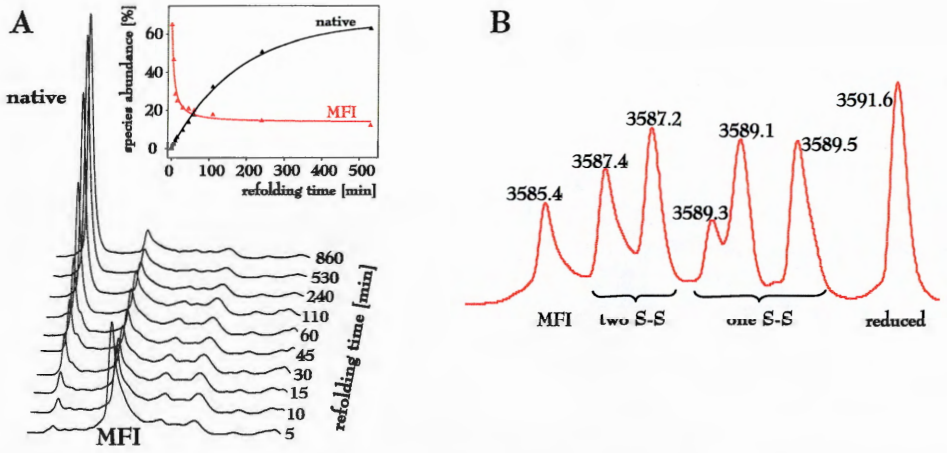

Figure 6.4: **Oxidative folding and reductive unfolding from MFI**: A. Oxidative folding starting from purified MFI species (100 mg/l MFI, 0.1 M NH_4OAc, 2 mM EDTA, 1.0 M GdnHCl, pH 8.50, 1 mM cysteine, 0.05 mM cystine, 25°C). Time-resolved RP-HPLC chromatograms of folding mixtures at different times after the start of oxidative folding. B. Reductive unfolding of MFI (0.5 mg/ml MFI, 0.1 M NH_4OAc, pH 3.0, 100 mM TCEP, 25°C) examined by LC-MS. The profile shows the mixture of products after 2 min reduction. The molecular masses of each species are shown, which determine their disulfide content. The profiles in both A. and B. were obtained with a RP C18 analytic column (5 μm) and a gradient of 0-20% B (5 min) followed by 20-30% B (30 min) at a flow rate of 0.8 ml min^{-1}.

6.2.4 Oxidative folding from a three-disulfide intermediate

The folding pathway from purified MFI species was studied to examine the kinetic role of MFI on the oxidative folding pathway of AAI protein. MFI was purified from the refolding mixture after 30 min of the start of folding from reduced protein, when it is present in ~40% abundance. It was then re-inserted again into the oxidative folding buffer. The chromatograms obtained at different times of folding are shown in the Figure 6.4A. The rate constant for formation of native AAI is 0.0057 min^{-1}, which is about half that found for folding from the reduced AAI protein at the same conditions (0.0095 min^{-1}).

When MFI is isolated and inserted into the refolding buffer, its non-native disulfides are likely to be reduced independently. This is suggested by MFI's reductive unfolding mechanism, described in section 6.2.3 and seen in Figure 6.4B. A very sim-

ilar folding pathway to the one proceeding from reduced AAI is observed in terms of the presence of disulfide species that are formed. All three-disulfide intermediates (I-1, I-4, I-5 and I-6) accumulate to significant amounts. If MFI's disulfides are reduced independently, I-5 and I-6 can form from the one-disulfide species, containing only the vicinal disulfide bridge. I-1 can similarly form from the one-disulfide isomer containing the Cys23-Cys31 bond, whereas I-4 can be made by partial reduction and isomerisation of disulfides in I-6.

The reduced state was not detected, which could signify that MFI does not need to be completely reduced in order to start the folding process towards the native state. This affirms the independence of MFI's disulfide bonds. It has been noted previously that fully oxidised intermediates involving non-native disulfides do not need to unfold and be reduced entirely for conversion to the native structure [47]. It has to be noted however, that MFI is far from an immediate precursor of the native structure. The one- and two-disulfide species, which form from MFI, are either short lived, exist in a minute concentration or their conversion to other three-disulfide species is spontaneous and irreversible [47]. Even though the direct precursors of the native species are one- and two-disulfide species with native disulfides, the rate-determining step of folding will be their formation from the three-disulfide isomers, such as MFI, as suggested previously in section 6.2.1.

Species containing non-native disulfide bonds in oxidative folding that are stabilised by non-native tertiary interactions are kinetic traps. Such species are off-pathway intermediates, which have to overcome an energy barrier to unravel the structure they have already formed [53]. MFI is in this sense an off-pathway intermediate in the oxidative folding from reduced AAI to the native state. It has to undergo a

complex process of disulfide reshuffling from non-native to native bonds, during which it forms other fully oxidised species (I-1, I-4, I-5 and I-6).

6.2.5 Mimicking oxidative folding *in vivo*

In vivo, the reduction and isomerisation of non-native disulfides made early in the oxidative folding process is essential to avoid misfolding. It is mostly achieved with the help of a folding catalyst, for example, protein disulfide isomerase (PDI), which catalyses the rearrangements that lead to native disulfides and native tertiary structure [164]. In an *in vitro* oxidative folding experiment the isomerisation can be induced with a reducing agent, which is frequently a core part of the refolding buffer. In the case of AAI this role is taken up by cysteine, which keeps a high reducing potential in order to convert non-native disulfides into reactive thiols. Disulfide isomerase proteins have been previously found in numerous other plant organisms, such as soybean [165], wheat [166, 167], maize [168] and rice [169]. It is therefore highly probable that Amaranthus crop plants have a gene for a disulfide isomerase. In experiments with AAI, the bacterial homologue DsbC was used, whose 3D structure is known to contain two active sites for disulfide reduction and isomerisation [83].

In order to approximate the conditions of oxidative folding to those in the endoplasmic reticulum, AAI was refolded in the presence of DsbC enzyme. A time-resolved RP-HPLC study of oxidative folding in a buffer containing disulfide isomerase DsbC is seen in Figure 6.5. It can be seen that all three-disulfide species, previously observed at optimum refolding conditions (MFI, I-1, I-4, I-5, I-6), are present on this pathway. However, MFI is not the most abundant intermediate species, but this is now I-1. The kinetics of native AAI regeneration was examined and compared to a study at the same

conditions in the absence of DsbC, seen in Figure 6.5B. Both kinetics fit well to a 1^{st} order exponential growth curve (R^2 >0.99) and have the following rate constants for the formation of native AAI protein: 0.014 min^{-1} with Dsbc, 0.0083 min^{-1} without DsbC. The final yield of the native AAI at this conditions is larger for the folding assisted by DsbC (94% versus 75% for the folding without DsbC). It has to be noted, however, that the high yield of formation of native AAI peptide in the presence of DsbC (94%) is comparable to the one obtained in the optimum refolding conditions (93%), described in section 6.2.1. The rate constant of formation of native AAI, is increased from 0.0095 min^{-1} in optimum refolding conditions (section 6.2.1) to 0.014 min^{-1} in the presence of DsbC (Figure 6.5B).

AAI is a small protein with respect to the DsbC molecule, which is a homodimer containing over 400 amino acids residues. The efficiency of reduction and isomerisation of AAI's non-native disulfides depends on the accessibility of such disulfides to the DsbC active sites. Although these active sites are well exposed on the surface of DsbC protein (Figure 1.8), it might not be easy to access the "hidden" disulfides of a cystine knot. Structures with more solvent exposed disulfides are probably isomerised more efficiently than others. This last observation could be an explanation for the fact that MFI is not as abundant in the folding assisted by DsbC, because a "bead-like" disulfide pattern probably has a high disulfide accessibility. The energy barrier for conversion of MFI to one- and two-disulfide species could be lowered in this way.

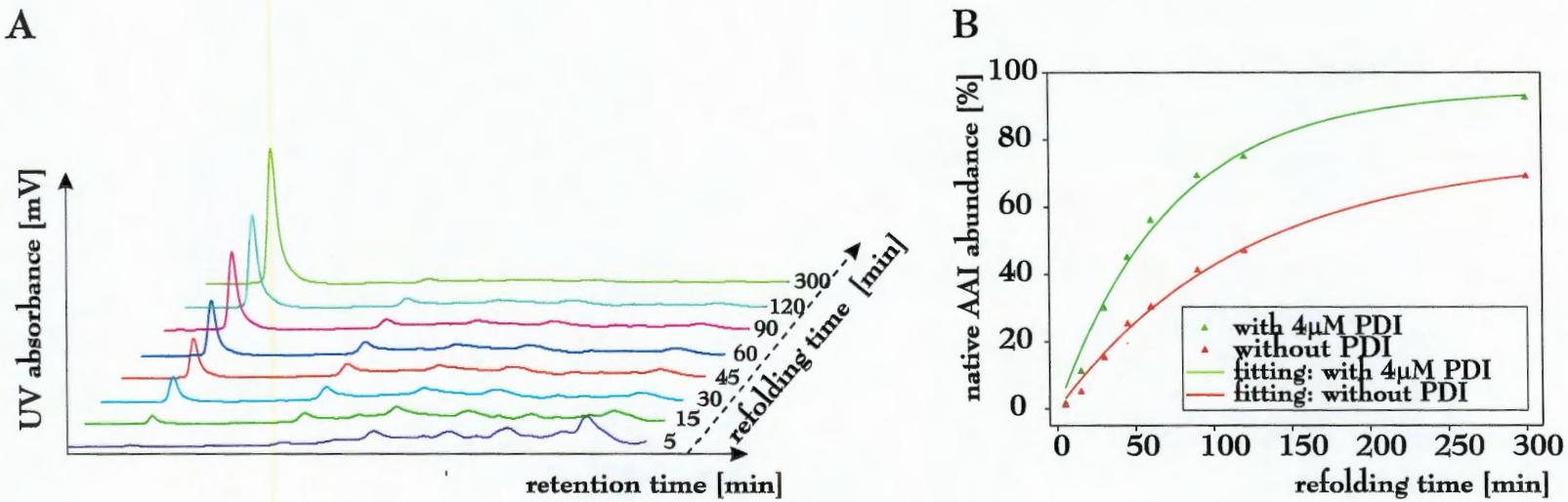

Figure 6.5: **DsbC assisted oxidative folding**: A. Time-resolved RP-HPLC profiles of DsbC assisted oxidative folding of AAI protein (100mg/l AAI, 0.1 M NH_4OAc, 2 mM EDTA, 1.0 M GdnHCl, pH 7.40, 1 mM cysteine, 0.05 mM cystine, 25°C, 4 μM DsbC). The profiles were obtained with a RP C18 analytic column (5 μm) and a gradient of 0-20% B (5 min) followed by 20-30% B (30 min) at a flow rate of 0.8 ml min^{-1}. B. Comparison of formation of native AAI with and without DsbC at the conditions described under A. The abundance data were calculated from absorbance detected at 214 nm. Both sets of data are fitted to a 1^{st} order exponential growth curve.

6.3 Oxidative folding at the level of a single residue

In order to follow the conformational changes that accompany the oxidative folding process on the level of a single amino acid residue, a series of experiments using 1D ^{1}H NMR and photo-CIDNP spectroscopy was carried out. As before for the study at the level of a single disulfide species, the properties of the known equilibrium species, which appear on the folding pathway were elucidated. In the case of NMR and photo-CIDNP study, the 1D spectra of the three milestone species of the oxidative folding reaction were compared to the time-resolved spectra obtained from the folding mixture.

6.3.1 Time-resolved NMR study

The tertiary folds of the three milestones of the folding pathway of AAI were first elucidated with equilibrium NMR. Purified reduced (R), MFI and native (N) samples were used in order to obtain orientation spectra for the conformational arrangements along the folding pathway (right panels in Figures 6.6 and 6.7. Chemical shift assignments of the native state were taken from Lu et al [20], assignments for the aromatic region as well as some assignments in the aliphatic region of MFI were obtained previously from 1D photo-CIDNP and 2D COSY experiments (see Chapter 5). Some reduced species assignments were obtained from 1D photo-CIDNP. The spectra of R show a series of fairly well resolved lines corresponding to the resonances of the component amino-acid residues in similar environments. However, this species does not display a completely random coil conformation, a fact that has been noted previously also for highly denatured proteins [170]. The corresponding spectra of the N state include significant dispersion in chemical shifts, a typical characteristic of native tertiary folds with persistent tertiary features. The NMR spectrum of the purified MFI species resembles that

of the N state as the chemical shifts are well dispersed and the lines are narrow, but the spectrum of MFI shows additional weaker signals, owing to minor conformation(s) present aside from the main conformation.

The equilibrium spectra of these three species were compared with the time-resolved spectra in order to look for the characteristics of the conformational folding assembly at different times of the refolding reaction. For the time-resolved experiments the protein conformations were separated from the refolding buffer with RP-HPLC, also to render the aliquots more stable while acquiring the NMR data. The NMR experiments were done on aliquots of protein assemblies from the folding reaction, which was stopped at different times after the initiation.

The aromatic region of the NMR spectra is shown in Figure 6.6. Here the NMR signals of four aromatic residues (Tyr21, Tyr27, Tyr28, Trp5) are shown for the equilibrium species and the time-resolved samples. There are several examples of how a peak with a native chemical shift develops through the folding reaction, for example of Tyr27(H2,H6) and Trp(H7), which are distinguishable already after 45 and 90 min of the folding reaction, respectively. The peaks belonging to MFI conformation appear almost immediately after the start of the reaction (15min) and slowly, but not completely, disappear from the folding assembly. Even after 1200 min, there is some MFI present in the mixture and the best-resolved example of an MFI peak is TyrC(H3,H5). The aliphatic region of ^{1}H NMR (Figure 6.7) shows a similar picture, although there are fewer lines as well resolved and as representative of the three equilibrium species. Three native resonances are partially restored as early as 45 min after the initiation of the folding reaction: Thr22(Hγ2), Val15(Hγ1) and Val15(Hγ2). The chemical shifts of the signals for MFI and the reduced species in this region are almost entirely over-

lapping, so one cannot say with certainty much about the disappearance of reduced state and the appearance of MFI. The only peak which is distinguishably non-native is Thr22(Hγ2), but even here MFI and reduced states overlap.

It has to be noted however, that some intensity in the NMR peaks, which are thought to belong to a specific species, for example MFI, could also be partially contributed from other disulfide intermediates with similar structural features. Therefore, peaks could be looked at as representatives of a specific structural environment for the residue in question.

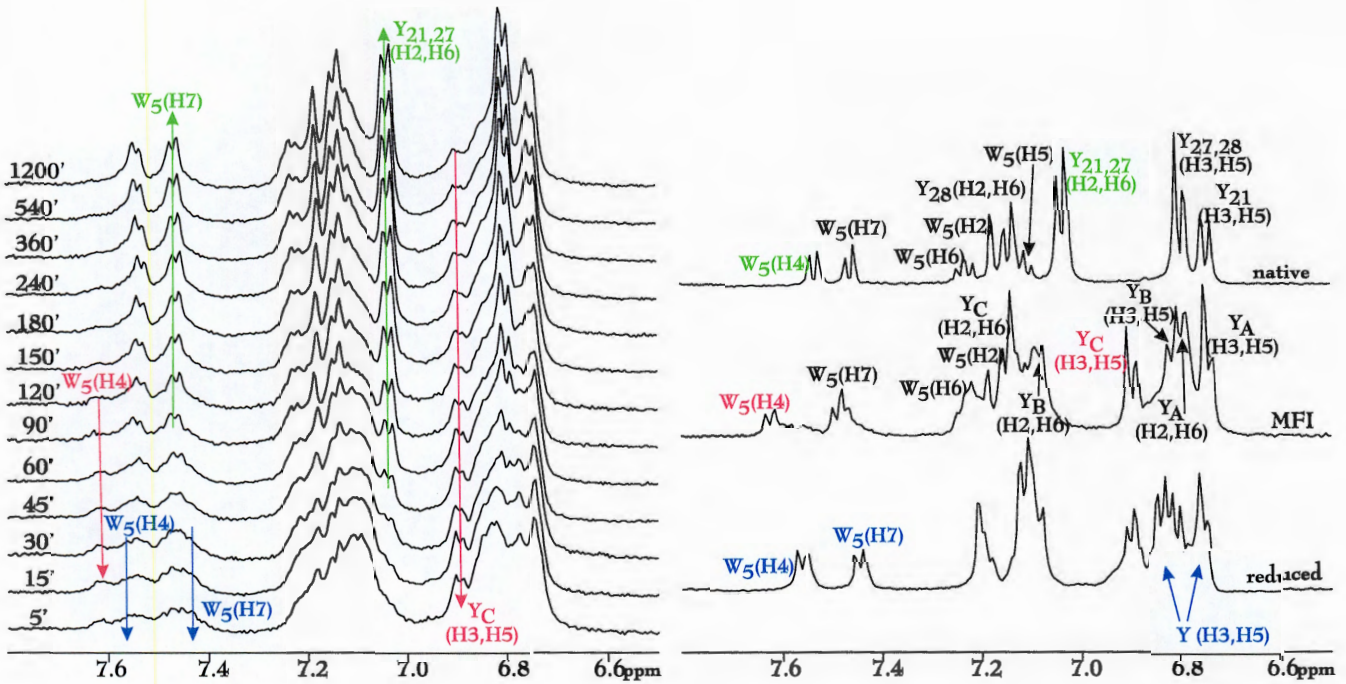

Figure 6.6: **600 MHz 1D ^{1}H spectra of oxidative folding of AAI, aromatic region**: Samples for the equilibrium spectra were purified on analytic RP-HPLC, lyophilised and re-dissolved at 1.0 mM in D_20 at pH 7.0. Samples for the time-resolved spectra were prepared in the following way. Purified reduced AAI was dissolved in the refolding buffer (100mg/l AAI, 0.1 M NH_4OAc, 2 mM EDTA, 1.0 M GdnHCl, pH 8.50, 1 mM cysteine, 0.05 mM cystine, 25°C). 10.0 ml aliquots were quenched at with 4 % TFA at various refolding times. The protein content of these aliquots was separated from the refolding buffer with RP-HPLC (gradient: 0-60% B in 10 min). Samples were then lyophilised and kept under N_2 at 4°C before they were re-dissolved in D_2O at pH 7.0. Sample stability was confirmed by comparing analytical RP-HPLC profiles before lyophilisation and after the NMR experiment. Spectra were recorded at 25°C, pH 7.0, with 10000 Hz spectral width, 4096 complex points and 512 acquisitions.

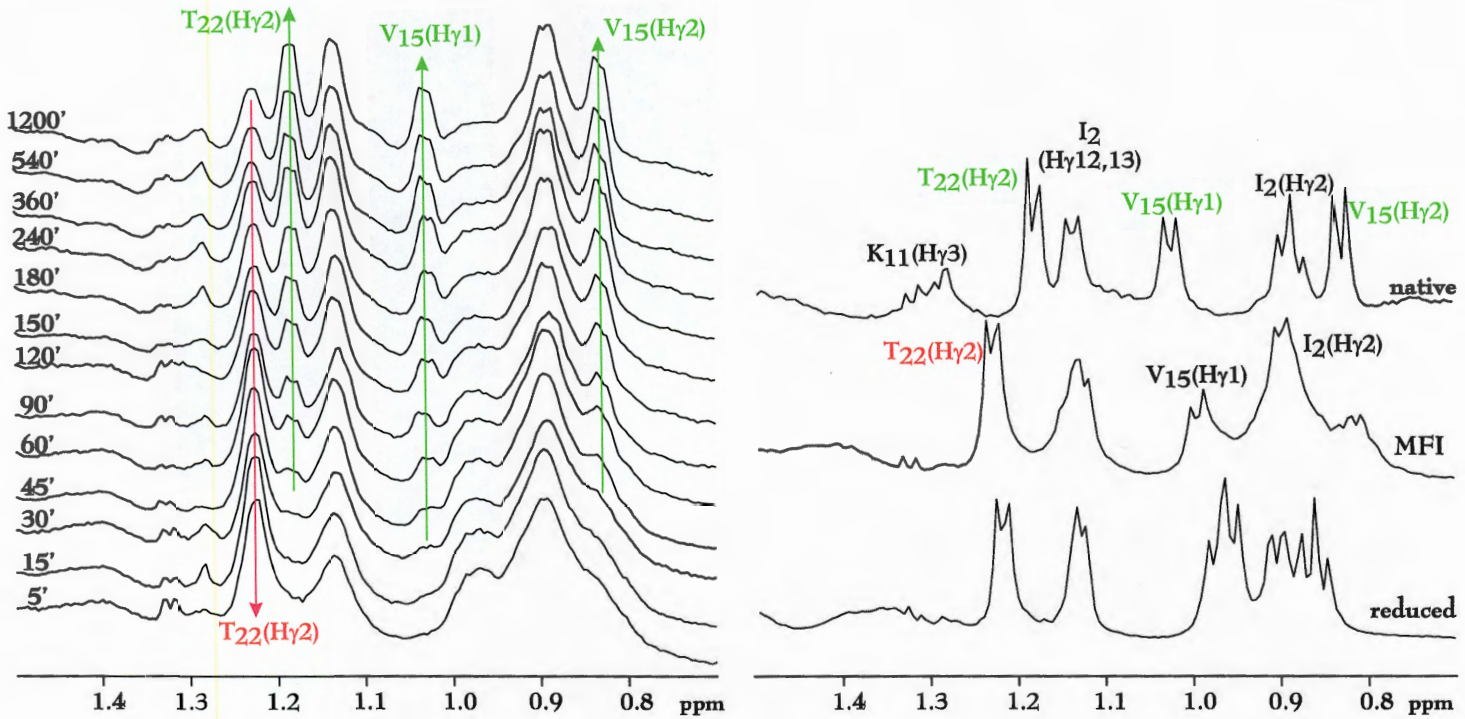

Figure 6.7: **600 MHz 1D ^{1}H spectra of oxidative folding of AAI, aliphatic region**: Samples for the equilibrium and the time-resolved spectra were prepared as described in caption to Figure 6.6. Spectra were recorded at 25°C, pH 7.0, with 10000 Hz spectral width, 4096 complex points and 512 acquisitions.

6.3.2 Time-resolved photo-CIDNP study

The photo-CIDNP experiments monitor not only the restoration of the native state and the abundance of various intermediates at different times, but also the accessibility of the four aromatic residues during the oxidative folding process. In this experiment the NMR signal intensity can be diminished by several factors, but primarily by the presence of the cysteine thiol groups.

First, preliminary experiments were done on all principal components of the folding buffer to test them for quenching effects of the FMN dye polarisation in the photo-CIDNP reaction. Cysteine, which is a present at 1.0 mM concentration, was found to have a strong quenching effect on the triplet state of FMN, which is responsible for polarization transfer and signal enhancement of the aromatic side chains of amino acid residues. This quenching phenomenon was assigned to its thiol group, well known to be an FMN triplet quencher, without production of CIDNP [113]. The oxidative folding buffer contains several other components at reasonably high concentrations (i.e. 0.1 M NH_4OAc, 1.0 M GdnHCl) with respect to the total protein concentration and the photo-sensitiser dye. This is why non-protein components of the refolding buffer were removed as described in the caption to Figure 6.6.

In the equilibrium photo-CIDNP spectra (Figure 6.8) of the three milestone species the four aromatic residues (Tyr21, Tyr27, Tyr28 and Trp5) participate in giving the signals. In the photo-CIDNP spectrum of the N species Tyr27 and Tyr28 are very well accessible, but Tyr21 seems less accessible. MFI shows accessibility for all three tyrosine residues, both in its major and minor conformation(s). This was discussed in detail in Chapter 5. The spectrum of the reduced species is similar to that of MFI and although this spectrum has not been assigned entirely, we can expect that the signals

appearing are from all four aromatic side chains. The signal intensity of this spectrum is somewhat hampered by the presence of the six thiol groups.

The time-resolved photo-CIDNP spectra can again be analysed in terms of the appearances of native, MFI and reduced resonances. The reduced peaks disappear within 30 min of the initiation of the folding reaction and this is especially clear for the Trp5(H4) peak, which is not overlapping with its native and MFI equivalents. The MFI peaks overlap both with the peaks of N and R. However, the emissive peaks of the (H3,H5) protons of TyrC and TyrA residues are well separated. They appear soon after the start of the folding and have significant abundance up to 150 min of refolding. The restoration of the native state is confirmed with Trp5 and Tyr27,28(H3,H5) resonances. It can be seen that from 45 min onwards there is some accumulation of the native state, which prevails from about 120 min of folding onwards.

One interesting observation is the disappearance of the number of peaks and their intensities in the region between 6.7 and 6.9 ppm of the time-resolved photo-CIDNP spectra with time. In this region of AAI spectrum, the resonances belonging to Tyr(H3,H5) protons appear as emissive peaks. At the beginning of oxidative folding, there is a large number of emissive peaks, which represents the large number of conformations where Tyr(H3,H5) protons are accessible. The intensities of some of this peaks gradually decrease and eventually some disappear. As the folding proceeds, one main peak is appearing (belonging to native Tyr27,28) and its intensity is growing. This process of disappearance of the emissive peaks qualitatively represents the disappearance of the number of conformations and Tyr residues, where the side chains are accessible to the solvent. Indirectly, this phenomenon could also be seen as the funnelling of a big number of conformations towards a more compact, native tertiary scaffold with fewer

and less accessible side chains.

In the above time-resolved ^{1}H NMR experiments, the signal intensities were directly proportional to the percentage population of an equilibrium species, but here there are many more factors, which can influence the signal intensity. In the equilibrium spectrum of the reduced species the signal-to-noise ratio is substantially lowered because this species possesses 6 thiol groups, which may all quench the triplet state of FMN. The accessibility of different aromatic residues changes as disulfide bonds are being oxidized and reshuffled to the three native disulfide bridges. It seems as though the tryptophan signals are higher than tyrosine in the spectra of MFI and reduced species. This could mean that Trp5 residue is relatively more accessible in these two species than in the native species, where its signals are comparable to those of the tyrosine residues. However, care must be take with such interpretation, because both types of amino acids compete for the triplet flavin and the reaction rate constant is higher for Trp than for Tyr [171].

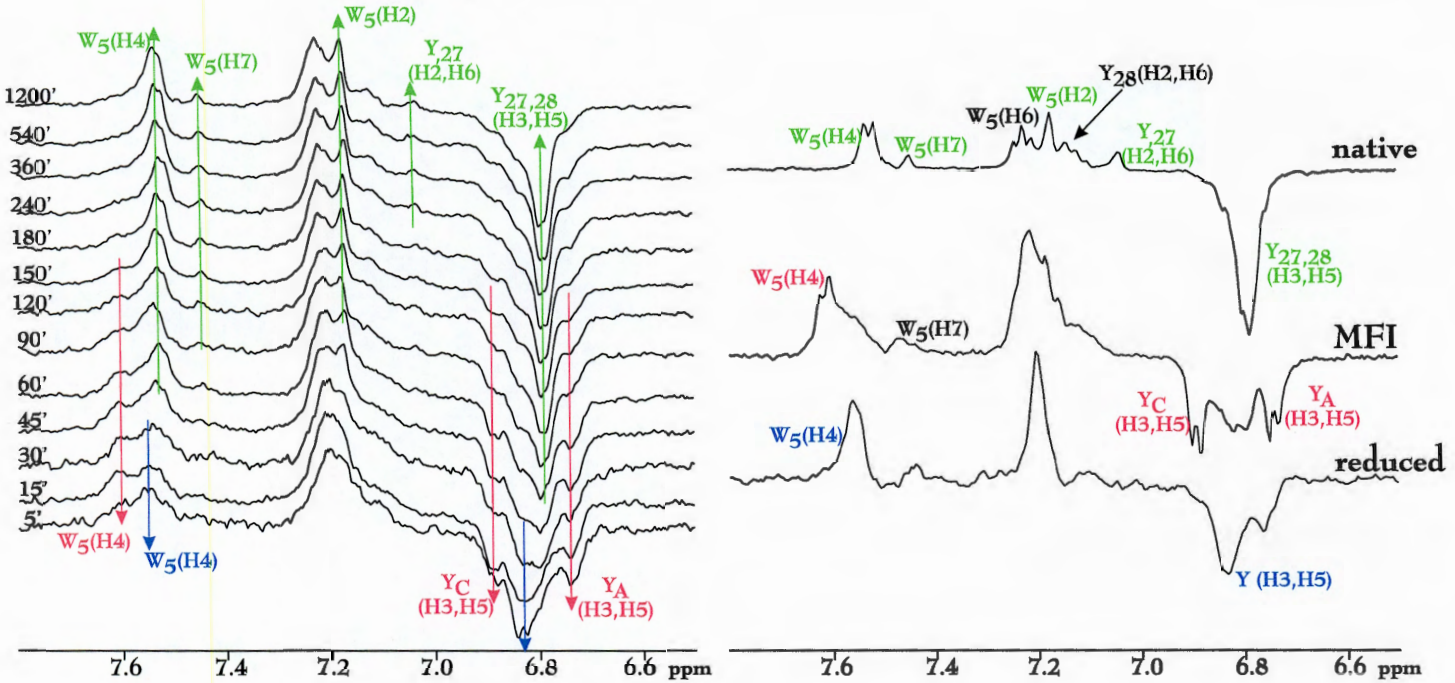

Figure 6.8: **600 MHz 1D ^{1}H photo-CIDNP spectra of oxidative folding, aromatic region**: Samples for the equilibrium and the time-resolved spectra were prepared as described in caption to Figure 6.6. Flavin mononucleotide (FMN) was added at 0.2 mM concentration. Spectra were recorded at 1.0 mM AAI, 25°C, pH 7, with 10,000 Hz spectral width, 4096 complex points and 16 acquisitions.

6.3.3 Comparison between single species and single residue studies

This is the first time that time-resolved NMR and photo-CIDNP studies are used to analyse a complete oxidative folding pathway. This is why it is important that the interpretation of results from the application of these techniques are examined for their credibility. These results from single-residue techniques are compared to the results from the single-species HPLC techniques (Figure 6.9).

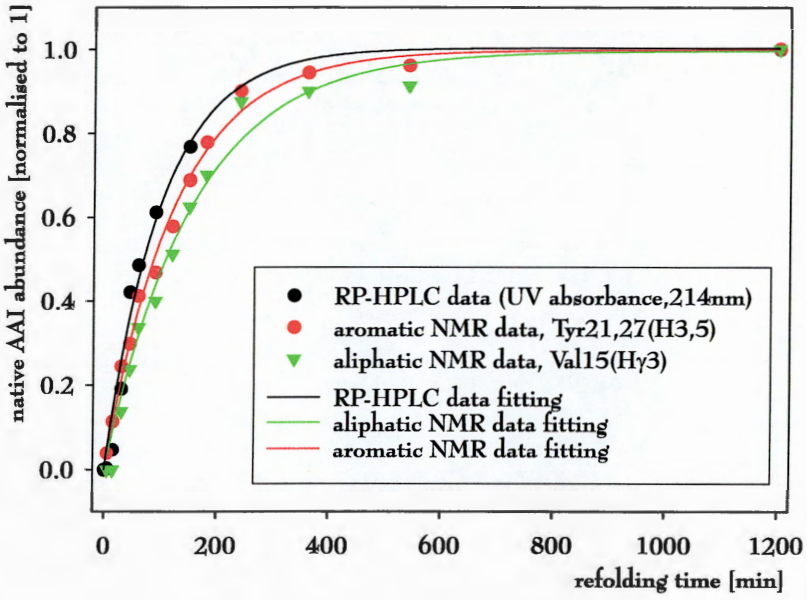

Figure 6.9: **Comparison of NMR and HPLC study**: The data for the formation of native AAI species are compared for the NMR and RP-HPLC study. The rate constants are extracted from the 1^{st} order exponential curves, which are fitted to the data.

The rate constants calculated for the three curves seen in Figure 6.9 are of the same order. The RP-HPLC study curve has a rate constant of 0.010 min^{-1}, the aromatic NMR data 0.0080 min^{-1} and the aliphatic NMR data 0.0064 min^{-1}. This indicates that time-resolved NMR experiments of this kind are a valuable technique for the study of conformational developments on an oxidative folding pathway. Photo-CIDNP intensities can not be used in quantitative comparisons of data, because they

do not only reflect the abundances of different species. An example for qualitative interpretation of the data in give in section 6.3.2.

6.4 Conclusions

The kinetics of oxidative folding was studied at the levels of a single species and a single residue. At the level of a single disulfide species, the kinetics in different conditions and the rate constants of formation of native AAI were compared. The configuration of the energy landscape was found to be independent of the rate of formation of the native AAI protein, which depends on the conditions of the folding buffer. This fact has been previously observed for other oxidative folding pathways [55].

MFI is statistically the most probable disulfide conformation when reduced AAI protein is allowed to refold. In the presence of high concentrations of denaturant that weakens the secondary and tertiary interactions, which usually drive the folding process, MFI was the most abundant conformation apart from the native state. The examination of the reductive unfolding mechanisms for the native and MFI species revealed the interdependence of disulfide bonds in a cystine knot arrangement of the former and their independence in a "bead-like" arrangement of the latter species. This defines MFI as an off-pathway species of the oxidative folding pathway, because it needs to reduce its disulfide bonds in order to proceed with the search for the native pairing. Oxidative folding was also examined in the presence of a protein disulfide isomerase, DsbC. All three-disulfide species observed before in optimum refolding conditions were present on the pathway assisted by this enzyme and MFI was not the most abundant one.

The study of AAI oxidative folding at the level of a single amino acid residue

was done with time-resolved NMR and photo-CIDNP spectroscopies and this is the first time that the latter technique is applied to this kind of process. Both techniques proved to be valuable for the study of conformational developments and aromatic accessibility changes along oxidative folding pathways. The data obtained for rate constants of native AAI formation were successfully compared with the data from single-species RP-HPLC studies.

The high statistical probability of formation of a "bead-like" disulfide pattern may well be responsible for the accumulation of MFI in the early stages of oxidative folding. It is difficult to say whether it is the structural constraints imposed by the non-native disulfides or the preferences encoded in the primary sequence that can be considered more important in directing the oxidative folding process by creating a compact fold. Non-native interactions in protein folding have been suggested to play a potentially significant role even for small proteins with simple topologies [172]. In the same way, the non-native disulfides on AAI's oxidative folding pathway could have a role in directing the folding process. More importantly, the native-like tertiary interactions, which have to be on average more stabilising than the non-native ones, will bias the system towards the native state, which will favour also the native-like disulfide pairing.

Bibliography

[1] D. J. Craik, N. L. Daly, and C. Waine. The cystine knot motif in toxins and implications for drug design. *Toxicon*, 39(1):43–60, 2001.

[2] O. Carugo, S. Lu, J. Luo, X. Gu, S. Liang, S. Strobl, and S. Pongor. Structural analysis of free and enzyme-bound amaranth α-amylase inhibitor: classification within the knottin fold superfamily and analysis of its functional flexibility. *Protein Eng.*, 14(9):639–646, 2001.

[3] N. Q. McDonald and W. A. Hendrickson. A structural superfamily of growth factors containing a cystine knot motif. *Cell*, 73(3):421–424, 1993.

[4] J. Murray-Rust, N. Q. McDonald, T. L. Blundell, M. Hosang, C. Oefner, F. Winkler, and R. A. Bradshaw. Topological similarities in TGF-β 2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure*, 1(2):153–159, 1993.

[5] P. K. Pallaghy, K. J. Nielsen, D. J. Craik, and R. S. Norton. A common structural motif incorporating a cystine knot and a triple-stranded β-sheet in toxic and inhibitory polypeptides. *Protein Sci.*, 3(10):1833–1839, 1994.

[6] N. W. Isaacs. Cystine knots. *Curr. Opin. Struct. Biol.*, 5(3):391–395, 1995.

[7] D. J. Craik, N. L. Daly, T. Bond, and C. Waine. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.*, 294:1327–1336, 1999.

[8] L. Narasimhan, J. Singh, C. Humblet, K. Guruprasad, and T. Blundell. Snail and spider toxins share a similar tertiary structure and 'cystine motif'. *Nat. Struct. Biol.*, 1(12):850–852, 1994.

[9] P. M. Harrison and Sternberg J. E. The disulfide β-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. *J. Mol. Biol.*, 264:603–623, 1996.

[10] H. Tamaoki, R. Miura, M. Kusunoki, Y. Kyogoku, Y. Kobayashi, and L. Moroder. Folding motifs induced and stabilized by distinct cystine frameworks. *Protein Eng.*, 11(8):649–659, 1998.

[11] R. S. Norton and P. K. Pallaghy. The cystine knot structure of ion channel toxins and related polypeptides. *Toxicon*, 36(11):1573–1583, 1998.

[12] M. Price-Carter, G. Bulaj, and D. P. Goldenberg. Initial disulfide formation steps in the folding of an ω-conotoxin. *Biochemistry*, 41(10):3507–3519, 2002.

[13] O. L. Franco, D. J. Rigden, F. R. Melo, and M. F. Grossi-De-Sa. Plant α-amylase inhibitors and their interaction with insect α-amylases. *Eur. J. Biochem.*, 269(2):397–412, 2002.

[14] R. Koradi, M. Billeter, and K. Wuthrich. MOLMOL: a programme for display and analysis of macromolecular structures. *J. Mol. Graph.*, 14(1):51–55, 1996.

[15] E. A. MacGregor, S. Janecek, and B. Svensson. Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim. Biophys. Acta*, 1546(1):1–20, 2001.

[16] S. Strobl, K. Maskos, M. Betz, G. Wiegand, R. Huber, F. X. Gomis-Ruth, and R. Glockshuber. Crystal structure of yellow meal worm α-amylase at 1.64 Å resolution. *J. Mol. Biol.*, 278(3):617–628, 1998.

[17] M. Richardson. *Seed Storage Proteins: The Enzyme Inhibitors*, volume 5, pages 261–307. Academic Press, London, 1990.

[18] A. Chagolla-Lopez, A. Blanco-Labra, A. Patthy, R. Sanchez, and S. Pongor. A novel α-amylase inhibitor from amaranth (Amaranthus hypocondriacus) seeds. *J. Biol. Chem.*, 269(38):23675–23680, 1994.

[19] V. Lozanov, C. Guarnaccia, A. Patthy, S. Foti, and S. Pongor. Synthesis and cystine/cysteine-catalyzed oxidative folding of the amaranth α-amylase inhibitor. *J. Pept. Res.*, 50(1):65–72, 1997.

[20] S. Lu, P. Deng, X. Liu, J. Luo, R. Han, X. Gu, S. Liang, X. Wang, F. Li, V. Lozanov, A. Patthy, and S. Pongor. Solution structure of the major α-amylase inhibitor of the crop plant amaranth. *J. Biol. Chem.*, 274(29):20473–20478, 1999.

[21] P. J. Pereira, V. Lozanov, A. Patthy, R. Huber, W. Bode, S. Pongor, and S. Strobl. Specific inhibition of insect α-amylases: yellow meal worm α-amylase in complex with the amaranth α-amylase inhibitor at 2.0 Å resolution. *Structure Fold. Des.*, 7(9):1079–1088, 1999.

[22] J. C. Martins, M. Enassar, R. Willem, J. M. Wieruzeski, G. Lippens, and S. J. Wodak. Solution structure of the main α-amylase inhibitor from amaranth seeds. *Eur. J. Biochem.*, 268(8):2379–2389, 2001.

[23] C. Venclovas, A. Zemla, K. Fidelis, and J. Moult. Comparison of performance in successive CASP experiments. *Proteins*, (Suppl. 5):163–170, 2001.

[24] C. Levinthal. Are there pathways for protein folding. *J. Chim. Phys.*, 65:44–45, 1968.

[25] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(96):223–230, 1973.

[26] P. S. Kim and R. L. Baldwin. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Ann. Rev. Biochem.*, 51:459–489, 1982.

[27] M. Karplus and D. L. Weaver. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.*, 3(4):650–668, 1994.

[28] T. Y. Tsong and R. L. Baldwin. A sequential model of nucleation-dependent protein folding: kinetic studies of ribonuclease A. *J. Mol. Biol.*, 63(3):453–469, 1972.

[29] K. A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985.

[30] S. C. Harrison and R. Durbin. Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. USA*, 82(12):4028–4030, 1985.

[31] M. Karplus. The Levinthal paradox: yesterday and today. *Fold. Des.*, 2:S69–75, 1997.

[32] C. M. Dobson, A. Sali, and M. Karplus. Protein folding: A perspective from theory and experiment. *Angew. Chem. Int. Ed.*, 37:868–893, 1998.

[33] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA*, 92:3626–3630, 1995.

[34] C. P. Schultz. Illuminating folding intermediates. *Nat. Struct. Biol.*, 7(1):7–10, 2000.

[35] C. M. Dobson. *Mechanisms of Protein Folding*, chapter The nature and significance of protein folding, pages 1–33. Oxford University Press, second edition, 2000.

[36] C. M. Dobson. Protein folding and its links with human disease. *Biochem. Soc. Symp.*, 68:1–26, 2001.

[37] W. J. Wedemeyer, E. Welker, M. Narayan, and H. A. Scheraga. Disulfide bonds and protein folding. *Biochemistry*, 39(23):7032, 2000.

[38] G. Bulaj, T. Kortemme, and D. P. Goldenberg. Ionization-reactivity relationships for cysteine thiols in polypeptides. *Biochemistry*, 37(25):8965–8972, 1998.

[39] T. E. Creighton. The disulfide folding pathway of BPTI. *Science*, 256(5053):111–114, 1992.

[40] T. E. Creighton. Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. *J. Mol. Biol.*, 113(2):275–293, 1977.

[41] J. S. Weissman and P. S. Kim. Reexamination of the folding of BPTI: predominance of native intermediates. *Science*, 253(5026):1386–1393, 1991.

[42] J. S. Weissman and P. S. Kim. Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, 89(20):9900–9904, 1992.

[43] J. S. Weissman and P. S. Kim. Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. *Nature*, 365(6442):185–188, 1993.

[44] J. S. Weissman and P. S. Kim. A kinetic explanation for the rearrangement pathway of BPTI folding. *Nat. Struct. Biol.*, 2(12):1123–1130, 1995.

[45] T. E. Creighton. *Mechanisms of Protein Folding*, chapter Protein folding coupled to disulphide-bond formation, pages 250–278. Oxford University Press, second edition, 2000.

[46] T. E. Creighton, C. J. Bagley, L. Cooper, N. J. Darby, R. B. Freedman, J. Kemmink, and A. Sheikh. On the biosynthesis of bovine pancreatic trypsin inhibitor (BPTI): structure, processing, folding and disulphide bond formation of the precursor in vitro and in microsomes. *J. Mol. Biol.*, 232(4):1176–1196, 1993.

[47] J. Y. Chang. The properties of scrambled hirudins. *J. Biol. Chem.*, 270(43):25661–25666, 1995.

[48] J. Y. Chang, P. Schindler, and B. Chatrenet. The disulfide structures of scrambled hirudins. *J. Biol. Chem.*, 270(20):11992–11997, 1995.

[49] J. Y. Chang, F. Canals, P. Schindler, E. Querol, and F. X. Aviles. The disulfide folding pathway of potato carboxypeptidase inhibitor. *J. Biol. Chem.*, 269(35):22087–22094, 1994.

[50] J. Y. Chang, L. Li, F. Canals, and F. X. Aviles. The unfolding pathway and conformational stability of potato carboxypeptidase inhibitor. *J. Biol. Chem.*, 275(19):14205–14211, 2000.

[51] J. Y. Chang. The disulfide folding pathway of tick anticoagulant peptide (TAP), a kunitz-type inhibitor structurally homologous to bpti. *Biochemistry*, 35(36):11702–11709, 1996.

[52] J. Y. Chang. Denatured states of tick anticoagulant peptide: Compositional analysis of unfolded scrambled isomers. *J. Biol. Chem.*, 274(1):123–128, 1999.

[53] J. Y. Chang and L. Li. The disulfide structure of denatured epidermal growth factor: preparation of scrambled disulfide isomers. *J. Protein Chem.*, 21(3):203–213, 2002.

[54] J. Y. Chang, L. Li, and P. H. Lai. A major kinetic trap for the oxidative folding of human epidermal growth factor. *J. Biol. Chem.*, 276(7):4845–4852, 2001.

[55] J. Y. Chang. The folding pathway of α-lactalbumin elucidated by the technique of disulfide scrambling: isolation of on-pathway and off-pathway intermediates. *J. Biol. Chem.*, 277(1):120–126, 2002.

[56] J. Y. Chang, L. Li, and A. Bulychev. The underlying mechanism for the diversity of disulfide folding pathways. *J. Biol. Chem.*, 275(12):8287–8289, 2000.

[57] J. Y. Chang. A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J. Biol. Chem.*, 272(1):69–75, 1997.

[58] D. M. Rothwarf and H. A. Scheraga. Regeneration of bovine pancreatic ribonuclease A. 1. steady-state distribution. *Biochemistry*, 32(10):2671–2679, 1993.

[59] D. M. Rothwarf, Y. J. Li, and H. A. Scheraga. Regeneration of bovine pancreatic ribonuclease A: detailed kinetic analysis of two independent folding pathways. *Biochemistry*, 37(11):3767–3776, 1998.

[60] E. Welker, M. Narayan, W. J. Wedemeyer, and H. A. Scheraga. Structural determinants of oxidative folding in proteins. *Proc. Natl. Acad. Sci. USA*, 98(5):2312–2316, 2001.

[61] H. A. Scheraga, W. J. Wedemeyer, and E. Welker. Bovine pancreatic ribonuclease A: oxidative and conformational folding studies. *Methods Enzymol.*, 341:189–221, 2001.

[62] Q. X. Hua, S. N. Gozani, R. E. Chance, J. A. Hoffmann, B. H. Frank, and M. A. Weiss. Structure of a protein in a kinetic trap. *Nat. Struct. Biol.*, 2(2):129–138, 1995.

[63] B. van den Berg, E. W. Chung, C. V. Robinson, and C. M. Dobson. Characterisation of the dominant oxidative folding intermediate of hen lysozyme. *J. Mol. Biol.*, 290(3):781–796, 1999.

[64] B. van den Berg, E. W. Chung, C. V. Robinson, P. L. Mateo, and C. M. Dobson. The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. *EMBO J.*, 18(17):4794–4803, 1999.

[65] Y. Kurokawa, N. Koganesawa, Y. Kobashigawa, T. Koshiba, M. Demura, and K. Niita. Oxidative folding of human lysozyme: effects of the loss of two disulfide bonds and the introduction of a calcium-binding site. *J. Protein Chem.*, 20(4):293–303, 2001.

[66] V. Guez, P. Roux, A. Navon, and M. E. Goldberg. Role of individual disulfide bonds in hen lysozyme early folding steps. *Protein Sci.*, 11(5):1136–1151, 2002.

[67] S. Hober, M. Uhlen, and B. Nilsson. Disulfide exchange folding of disulfide mutants of insulin-like growth factor I in vitro. *Biochemistry*, 36(15):4616–4622, 1997.

[68] Y. Yang, J. Wu, and J. T. Watson. Probing the folding pathways of long R(3) insulin-like growth factor-I (LR(3) IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J. Biol. Chem.*, 274(53):37598–37604, 1999.

[69] S. J. Milner, J. A. Carver, F. J. Ballard, and G. L. Francis. Probing the disulfide folding pathway of insulin-like growth factor-I. *Biotechnol. Bioeng.*, 62(6):693–703, 1999.

[70] S. Liang, Q. Shu, X. Wang, and X. Zong. Oxidative folding of reduced and denatured huwentoxin-I. *J. Protein Chem.*, 18(6):619–625, 1999.

[71] M. Ruoppolo, F. Talamo, P. Pucci, M. Moutiez, E. Quemeneur, A. Menez, and G. Marino. Slow folding of three-fingered toxins is associated with the accumulation of native disulfide-bonded intermediates. *Biochemistry*, 40(50):15257–15266, 2001.

[72] A. Rattenholl, M. Ruoppolo, A. Flagiello, M. Monti, F. Vinci, G. Marino, H. Lilie, E. Schwarz, and R. Rudolph. Pro-sequence assisted folding and disulfide bond formation of human nerve growth factor. *J. Mol. Biol.*, 305(3):523–33, 2001.

[73] N. L. Daly, R. J. Clark, and D. J. Craik. Disulfide folding pathways of cystine knot proteins. Tying the knot within the circular backbone of the cyclotides. *J. Biol. Chem.*, 278(8):6314–6322, 2003.

[74] J. W. Cuozzo and C. A. Kaiser. Competition between glutathione and protein thiols for disulphide-bond formation. *Nat. Cell. Biol.*, 1(3):130–135, 1999.

[75] A. Zapun, C. A. Jakob, D. Y. Thomas, and J. J. Bergeron. Protein folding in a specialized compartment: the endoplasmic reticulum. *Structure Fold. Des.*, 7(8):R173–182, 1999.

[76] A. R. Frand and C. A. Kaiser. Two pairs of conserved cysteines are required for the oxidative activity of Ero1p in protein disulfide bond formation in the endoplasmic reticulum. *Mol. Biol. Cell.*, 11(9):2833–2843, 1999.

[77] M. G. Pollard, K. J. Travers, and J. S. Weissman. Ero1p: a novel and ubiquitous protein with an essential role in oxidative protein folding in the endoplasmic reticulum. *Mol. Cell.*, 1(2):171–182, 1998.

[78] B. P. Tu, S. C. Ho-Schleyer, K. J. Travers, and J. S. Weissman. Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science*, 290(5496):1571–1574, 2000.

[79] D. M. Ferrari and H. D. Soling. The protein disulphide-isomerase family: unravelling a string of folds. *Biochem. J.*, 339:1–10, 1999.

[80] J. Kemmink, N. J. Darby, K. Dijkstra, M. Nilges, and T. E. Creighton. Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear ^{13}C-^{15}N NMR spectroscopy. *Biochemistry*, 35(24):7684–7691, 1996.

[81] J. Kemmink, K. Dijkstra, M. Mariani, R. M. Scheek, E. Penka, M. Nilges, and N. J. Darby. The structure in solution of the b-domain of protein disulfide isomerase. *J. Biomol. NMR*, 13(4):357–368, 1999.

[82] F. De Lorenzo, R. F. Goldberger, Jr. Steers, E., D. Givol, and B. Anfinsen. Purification and properties of an enzyme from beef liver which catalyzes sulfhydryl-disulfide interchange in proteins. *J. Biol. Chem.*, 241(7):1562–1567, 1966.

[83] A. A. McCarthy, P. W. Haebel, A. Torronen, V. Rybin, E. N. Baker, and P. Metcalf. Crystal structure of the protein disulfide bond isomerase, DsbC, from Escherichia coli. *Nat. Struct. Biol.*, 7(3):196–199, 2000. 1072-8368.

[84] C. S. Sevier, J. W. Cuozzo, A. Vala, F. Aslund, and C. A. Kaiser. A flavoprotein oxidase defines a new endoplasmic reticulum pathway for biosynthetic disulphide bond formation. *Nat. Cell. Biol.*, 3(10):874–882, 2001.

[85] E. Gross, C. S. Sevier, A. Vala, C. A. Kaiser, and D. Fass. A new FAD-binding fold and intersubunit disulfide shuttle in the thiol oxidase Erv2p. *Nat. Struct. Biol.*, 9(1):61–67, 2002.

[86] D.J. Brockwell, D.A. Smith, and S.E. Radford. Protein folding mechanisms: new methods and emerging ideas. *Curr. Opp. Struc. Biol.*, 10:16–25, 2000.

[87] R. B. Merrifield. *J. Am. Chem. Soc.*, 85:2149, 1963.

[88] M. Bergmann and L. Zervas. *Chem. Berichte*, 65:1192, 1932.

[89] E. Atherton, M. J. Gait, R. C. Sheppard, and B. J. Williams. *Bioorg. Chem.*, (8):351, 1979.

[90] E. Atherton and R.C. Sheppard. *Solid phase peptide synthesis: A practical approach.* The practical approach series. OU Press, 1989.

[91] J. H. Jones. *Amino acid and peptide synthesis.* Oxford Chemistry Primers. OU press, 1997.

[92] E. Kaiser, R. L. Colescott, C. D. Bossinger, and P. I. Cook. *Anal. Biochem.*, 34:565, 1970.

[93] T. Vojkovski. *Pept. Res.*, 8:236, 1995.

[94] M. Tswett. *Ber. Deutsch. Botan. Ges.*, 24:384, 1906.

[95] A. J. P. Martin and R. L. M. Synge. *Biochem. J.*, 35:1358–1368, 1941.

[96] L. Konermann and D. J. Douglas. Unfolding of proteins monitored by electrospray ionization mass spectrometry: a comparison of positive and negative ion modes. *J. Am. Soc. Mass. Spectrom.*, 9(12):1248–1254, 1998.

[97] D. L. Smith, Y. Deng, and Z. Zhang. Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J. Mass. Spectrom.*, 32(2):135–146, 1997.

[98] T. D. Veenstra. Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes. *Biochem. Biophys. Res. Commun.*, 257(1):1–5, 1999.

[99] H. Hernandez and C. V. Robinson. Dynamic protein complexes: insights from mass spectrometry. *J. Biol. Chem.*, 276(50):46685–46688, 2001.

[100] S. M. Kelly and N. C. Price. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.*, 1(4):349–384, 2000.

[101] R. W. Woody. Circular dichroism. *Methods Enzymol*, 246:34–71, 1995.

[102] F. Bloch, W. W. Hansen, and M. Packard. The nuclear induction experiment. *Phys. Rev.*, 70(7):474–485, 1946.

[103] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69:37–38, 1946.

[104] B. Jacobson, W. A. Anderson, and J. T. Arnold. *Nature*, 17:772–773, 1954.

[105] M. Saunders, A. Wishnia, and J.G. Kirkwood. *J. Am. Chem. Soc.*, 79:3289, 1957.

[106] J. Bargon, H. Fischer, and U. Johnsen. *Zeitschrift fur Naturforschung*, 22a:1551, 1967.

[107] H. R. Ward and R. G. Lawler. Nuclear magnetic resonance emission and enhanced absorption in rapid organometallic reactions. *J. Am. Chem. Soc.*, 89:5518–5519, 1967.

[108] G. L. Closs. A mechanism explaining nuclear spin polarizations in radical combination reactions. *J. Am. Chem. Soc.*, 91(16):4552–4554, 1969.

[109] R. Kaptein and L. J. Oosterhoff. Chemically induced dynamic nuclear polarization III: anomalous multiplets of radical coupling and disproportionation products. *Chem. Phys. Lett.*, 4(4):214–216, 1969.

[110] R. Kaptein, K. Dijkstra, and K. Nicolay. Laser photo-CIDNP as a surface probe for proteins in solution. *Nature*, 274(5668):293–294, 1978.

[111] R. Kaptein. *Structural Information from photo-CIDNP in Proteins.* Nuclear Magnetic Resonance Spectroscopy in Molecular Biology. D. Reidel Publishing, Dordrecht, Holland, 1978.

[112] R. Kaptein, K. Dijkstra, F. Mueller, C.G. van Schagen, and A.J. Visser. 360-mhz laser-induced photo-CIDNP in photoreactions of flavins. *J. Magn. Reson.*, 31:171–176, 1978.

[113] K. Maeda, C. E. Lyon, J. J. Lopez, M. Cemazar, C. M. Dobson, and P. J. Hore. Improved photo-CIDNP methods for studying protein structure and folding. *J. Biomol. NMR*, 16(3):235–244, 2000.

[114] C. E. Lyon. *Photo-CIDNP and Protein Folding.* PhD thesis, Oxford Univeristy, 1999.

[115] J. J. Lopez. *Photo-CIDNP studies of Amino Acids and Proteins.* PhD thesis, Oxford Univeristy, 2001.

[116] R. W. Broadhurst. *Photo-CIDNP studies of Proteins.* PhD thesis, Oxford Univeristy, 1991.

[117] C. E. Lyon, J.A. Jones, C. Redfield, C. M. Dobson, and P. J. Hore. Two-dimensional ^{15}N-^{1}H photo-CIDNP as a surface probe of native and partially structured proteins. *J. Am. Chem. Soc.*, 121(27):6505–6506, 1999.

[118] C. E. Lyon, E. S. Suh, C. M. Dobson, and P. J. Hore. Probing the exposure of tyrosine and tryptophan residues in partially folded proteins and folding intermediates by CIDNP pulse-labeling. *J. Am. Chem. Soc.*, 124(44):13018–13024, 2002.

[119] M. Cemazar. Photo-CIDNP studies of amino acids and proteins. Master's thesis, Oxford Univeristy, 1999.

[120] M. Lebl, G. Lebl, and V. Krchnak. Peptide Companion. computer programme, 1994.

[121] Y. Han, F. Albercio, and G. Barany. Occurence and minimisation of cysteine racemisation during stepwise solid-phase peptide sythesis. *J. Org. Chem.*, 62(13):4307–4312, 1997.

[122] J. W. van Nispen. *Topics in Pharmaceutical Sciences*, chapter Chemical and physical stability of peptides, pages 293–307. Elsevier Science Publishers, 1987.

[123] A. Trans-Moseman, N. Schauer, and E. De Bernardez Clark. Renaturation of Escherichia coli-derived recombinant human macrophage colony-stimulating factor. *Prot. Expr. Purif.*, 16:181–89, 1999.

[124] E. D. Clark. Protein refolding for industrial processes. *Curr. Opin. Biotech.*, (12):202–207, 2001.

[125] A. Heitz, L. Chiche, D. Le-Nguyen, and B. Castro. Folding of the squash trypsin inhibitor EETI II. evidence of native and non-native local structural preferences in a linear analogue. *Eur. J. Biochem.*, 233(3):837–846, 1995.

[126] X. Wang, M. Connor, R. Smith, M. W. Maciejewski, M. E. Howden, G. M. Nicholson, M. J. Christie, and G. F. King. Discovery and characterization of a family of insecticidal neurotoxins with a rare vicinal disulfide bridge. *Nat. Struct. Biol.*, 7(6):505–513, 2000.

[127] M. Ghosh, C. Anthony, K. Harlos, M. G. Goodwin, and C. Blake. The refined structure of the quinoprotein methanol dehydrogenase from methylobacterium extorquens at 1.94 Å. *Structure*, 3(2):177–187, 1995.

[128] Y. J. Zheng, Zx Xia, Zw Chen, F. S. Mathews, and T. C. Bruice. Catalytic mechanism of quinoprotein methanol dehydrogenase: A theoretical and X-ray crystallographic investigation. *Proc. Natl. Acad. Sci. USA*, 98(2):432–434, 2001.

[129] P. R. Afolabi, F. Mohammed, K. Amaratunga, O. Majekodunmi, S. L. Dales, R. Gill, D. Thompson, J. B. Cooper, S. P. Wood, P. M. Goodwin, and C. Anthony. Site-directed mutagenesis and X-ray crystallography of the PQQ-containing quinoprotein methanol dehydrogenase and its electron acceptor, cytochrome c(L). *Biochemistry*, 40(33):9799–9809, 2001.

[130] Z. Xia, W. Dai, Y. Zhang, S. A. White, G. D. Boyd, and F. S. Mathews. Determination of the gene sequence and the three-dimensional structure at 2.4 Å resolution of methanol dehydrogenase from methylophilus W3A1. *J. Mol. Biol.*, 259(3):480–501, 1996.

[131] Z. X. Xia, Y. N. He, W. W. Dai, S. A. White, G. D. Boyd, and F. S. Mathews. Detailed active site configuration of a new crystal form of methanol dehydrogenase from methylophilus W3A1 at 1.9 Å resolution. *Biochemistry*, 38(4):1214–1220, 1999.

[132] T. Keitel, A. Diehl, T. Knaute, J. J. Stezowski, W. Hohne, and H. Gorisch. X-ray structure of the quinoprotein ethanol dehydrogenase from Pseudomonas aeruginosa: basis of substrate specificity. *J. Mol. Biol.*, 297(4):961–974, 2000.

[133] A. Oubrie, H. J. Rozeboom, K. H. Kalk, E. G. Huizinga, and B. W. Dijkstra. Crystal structure of quinohemoprotein alcohol dehydrogenase from Comamonas testosteroni: structural basis for substrate oxidation and electron transfer. *J. Biol. Chem.*, 277(5):3727–3732, 2002.

[134] K. Brejc, W. J. van Dijk, R. V. Klaassen, M. Schuurmans, J. van Der Oost, A. B. Smit, and T. K. Sixma. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*, 411(6835):269–276, 2001.

[135] H. Zeng, L. Moise, M. A. Grant, and E. Hawrot. The solution structure of the complex formed between α-bungarotoxin and an 18-mer cognate peptide derived from the α1 subunit of the nicotinic acetylcholine receptor from Torpedo californica. *J. Biol. Chem.*, 276(25):22930–22940, 2001.

[136] J. J. Bellizzi, J. Widom, C. Kemp, J. Y. Lu, A. K. Das, S. L. Hofmann, and J. Clardy. The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. *Proc. Natl. Acad. Sci. USA*, 97(9):4573–4578, 2000.

[137] M. Das, B. N. Mallick, S. C. Dasgupta, and A. Gomes. A sleep inducing factor from common indian toad (Bufo melanostictus) skin extract. *Toxicon*, 38(9):1267–1281, 2000.

[138] A. Teplyakov, K. Polyakov, G. Obmolova, B. Strokopytov, I. Kuranova, A. Osterman, N. Grishin, S. Smulevitch, O. Zagnitko, and O. Galperina. Crystal structure of carboxypeptidase T from Thermoactinomyces vulgaris. *Eur. J. Biochem.*, 208(2):281–288, 1992.

[139] H. N. Hunter, D. B. Fulton, T. Ganz, and H. J. Vogel. The solution structure of human hepcidin, a peptide hormone with antimicrobial activity that is involved in iron uptake and hereditary hemochromatosis. *J. Biol. Chem.*, 277(40):37597–37603, 2002.

[140] O. Carugo, M. Cemazar, S. Zahariev, I. Hudky, Z. Gspri, A. Perczel, and S. Pongor. Vicinal disulfide turns. *Protein Eng.*, in press, 2003.

[141] S. Capasso, C. Mattia, L. Mazzarella, and R. Puliti. Structure of a cis-peptide unit: molecular conformation of the cyclic disulfide L-cysteinyl-L-cysteine. *Acta Cryst.*, B33:2080–2083, 1977.

[142] Y. Hata, Y. Matsuura, N. Tanaka, T. Ashida, and M. Kakudo. Tert-butyloxycarbonyl-L-cysteinyl-L-cysteine disulfide methyl ester. *Acta Cryst.*, B33:3561–3564, 1977.

[143] S. Capasso, A. M. Garzillo, G. Marino, L. Mazzarella, P. Pucci, and G. Sannia. Mitochondrial bovine aspartate aminotransferase: Preliminary sequence and crystallographic data. *FEBS Lett.*, 101(2):351–354, 1979.

[144] S. Capasso, L. Mazzarella, F. Sica, and A. Zagari. Type II' beta-bend conformation of tert-butyloxycarbonyl-L-amino-succinyl-L-alanyl-glycine methyl ester in the solid state. *Int. J. Pept. Protein Res.*, 24(6):588–596, 1984.

[145] D. K. Sukumaran, M. Prorok, and D. S. Lawrence. A molecular contraint that generates a cis peptide bond. *J. Am. Chem. Soc.*, 113:706–707, 1991.

[146] D. Z. Avizonis, S. Farr-Jones, P. A. Kosen, and V. J. Basus. Conformations and dynamics of the essential cysteinyl-cysteine ring derived from the acetylcholine receptor. *J. Am. Chem. Soc.*, 118(51):13031–13039, 1996.

[147] C. C. Blake, M. Ghosh, K. Harlos, A. Avezoux, and C. Anthony. The active site of methanol dehydrogenase contains a disulphide bridge between adjacent cysteine residues. *Nat. Struct. Biol.*, 1(2):102–105, 1994.

[148] C. J. Creighton, C. H. Reynolds, D. H. Lee, G. C. Leo, and A. B. Reitz. Conformational analysis of the eight-membered ring of the oxidized cysteinyl-cysteine unit implicated in nicotinic acetylcholine receptor ligand recognition. *J. Am. Chem. Soc.*, 123(50):12664–12669, 2001.

[149] I. Hudaky, Z. Gspri, O. Carugo, M. Cemazar, S. Pongor, and A. Perczel. Vicinal disulfide bridge conformers by experimental methods and by ab initio and DFT molecular computations. *Protein Sci.*, in press, 2003.

[150] C. Park and R. T. Raines. Adjacent cysteine residues as a redox switch. *Protein Eng.*, 14(11):939–942, 2001.

[151] S. Engst and S. M. Miller. Alternative routes for entry of HgX_2 into the active site of mercuric ion reductase depend on the nature of the X ligands. *Biochemistry*, 38(12):3519–3529, 1999.

[152] M. Cemazar, S. Zahariev, J. J. Lopez, O. Carugo, J. A. Jones, P. J. Hore, and S. Pongor. Oxidative folding intermediates with nonnative disulfide bridges between adjacent cysteine residues. *Proc. Natl. Acad. Sci. USA*, 100(10):5754–5759, 2003.

[153] S. J. Hubbard and J. M. Thornton. NACCESS. computer programme, 1996.

[154] D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient nmr techniques. *Biochemistry*, 38(50):16424–16431, 1999.

[155] H. Pan, G. Barany, and C. Woodward. Reduced BPTI is collapsed: A pulsed field gradient NMR study of unfolded and partially folded bovine pancreatic trypsin inhibitor. *Protein Sci.*, 6(9):1985–1992, 1997.

[156] S. J. Gibbs and C. S. Johnson. A PFG NMR experiment for accurate diffusion and flow studies in the presence of eddy currents. *J. Magn. Reson.*, 93:395–402, 1991.

[157] V. Sklenar, D. Torchia, and A. Bax. Measurement of ^{13}C longitudinal relaxation using ^{1}H detection. *J. Mag. Reson.*, 73:375–379, 1987.

[158] W. Peti, L. J. Smith, C. Redfield, and H. Schwalbe. Chemical shifts in denatured proteins: resonance assignments for denatured ubiquitin and comparisons with other denatured proteins. *J. Biomol. NMR*, 19(2):153–165, 2001.

[159] J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, and H. Schwalbe. Long-range interactions within a nonnative protein. *Science*, 295(5560):1719–1722, 2002.

[160] C. Redfield, B. A. Schulman, M. A. Milhollen, P. S. Kim, and C. M. Dobson. α-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat. Struct. Biol.*, 6(10):948–952, 1999.

[161] W. J. Wedemeyer, X. Xu, E. Welker, and H. A. Scheraga. Conformational propensities of protein folding intermediates: Distribution of species in the 1S, 2S, and 3S ensembles of the [C40A,C95A] mutant of bovine pancreatic ribonuclease A. *Biochemistry*, 41(5):1483–1491, 2002.

[162] W. Kauzmann. *Adv. Prot. Chem.*, 14:1–36, 1959.

[163] K. W. Plaxco, K. T. Simons, and D. Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.*, 277(4):985–994, 1998.

[164] H. F. Gilbert. Protein disulfide isomerase and assisted protein folding. *J. Biol. Chem.*, 272(47):29399–29402, 1997.

[165] K. Kainuma, T. Ookura, and Y. Kawamura. Purification and characterization of protein disulfide isomerase from soybean. *J. Biochem. (Tokyo)*, 117(1):208–215, 1995.

[166] M. Ciaffi, A. R. Paolacci, L. Dominici, O. A. Tanzarella, and E. Porceddu. Molecular characterization of gene sequences coding for protein disulfide isomerase (PDI) in durum wheat (Triticum turgidum ssp. durum). *Gene*, 265(1-2):147–156, 2001. ,

[167] Y. Shimoni, G. Segal, X. Z. Zhu, and G. Galili. Nucleotide sequence of a wheat cDNA encoding protein disulfide isomerase. *Plant Physiol.*, 107(1):281, 1995.

[168] C. P. Li and B. A. Larkins. Expression of protein disulfide isomerase is elevated in the endosperm of the maize floury-2 mutant. *Plant Mol. Biol.*, 30(5):873–882, 1996.

[169] Y. Takemoto, S. J. Coughlan, T. W. Okita, H. Satoh, M. Ogawa, and T. Kumamaru. The rice mutant esp2 greatly accumulates the glutelin precursor and deletes the protein disulfide isomerase. *Plant Physiol.*, 128(4):1212–1222, 2002.

[170] D. Shortle and M. S. Ackerman. Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, 293(5529):487–489, 2001.

[171] K. A. Muszkat and T. Wismontski-Knittel. Reactivities of tyrosine, histidine, tryptophan and methionine in radical pair formation in flavin triplet induced protein nuclear magnetic polarization. *Biochemistry*, 24:5416–5421, 1985.

[172] A. P. Capaldi, C. Kleanthous, and S. E. Radford. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.*, 9(3):209–216, 2002.