Optical Triggers for Protein Folding

by

Lilia Milanesi

Department of Chemistry
University of Sheffield

Submitted to the University of Sheffield as partial fulfilment for the degree of Doctor of Philosophy

August 2002

Dedicato a Francesco, Vittoria,
Salva e Silvia

Summary

Lilia Milanesi

Optical triggers for protein folding

A better understanding of the mechanism by which protein folding proceeds requires the elucidation of processes that occur on the nanosecond time scale. Advances in laser technology have led to the development of new methods which can initiate and probe protein folding on short time scales. However, as discussed in chapter 1, the current optical methods are applicable to a narrow range of proteins or are limited to the microsecond timescale, which is insufficient to follow many protein rearrangements events.

This thesis describes the development of novel optical triggers that have the potential to be more generally applicable than previous methods for the initiation of protein folding in the submicrosecond time scale.

Chapter 2 outlines two strategies used for the development of our optical trigger mechanisms. Both strategies make use of chemical methods to introduce photocleavable aromatic disulfides at key points of the target protein structure.

Chapter 3 details the synthesis and photolysis of a series of aromatic disulfides that have been used as a model system to test the efficiency of the designed optical trigger mechanism. Our results show that the disulfide bond is cleaved in less than a few picoseconds, thus providing a fast trigger for protein folding.

Chapter 4 details the procedure used to introduce a photolabile aromatic disulfide into the target protein. Preliminary results indicate that the aromatic disulfide crosslink forces the protein in a non-native structure, thus providing a good system for the study of fast folding kinetics.

Chapter 5 details a second procedure that I have used to functionalise the target protein with a photolabile aromatic disulfide attached to a single amino acid residue. Preliminary results indicated that the modification of a single residue does not perturb the protein native structure thus this system is not useful for the study of fast folding kinetics.

ACKOWLEDGEMENTS

Firstly I want to thank my supervisor Chris Hunter who has supported my work with advice, patience (….a lot!!!!) and enthusiasm. Secondly I am grateful to the people who have worked on some aspects of the project (listed in order of their contribution):
the proteins used in this thesis have been prepared and characterised by Clare Jelinska under the supervision of J. Waltho in the department of Molecular Biology, the photolysis experiments have been carried out in collaboration with G. Reid and G. Beddard at the University of Leeds.

I am also thankful to Liam McDonnel who carried out the mass spectroscopy analyses of my proteins at the University of Warwick. Many thanks to Paul Thaw for spending some of his time on reprocessing my NMR spectra. In terms of theoretical and practical lessons for the use of the NMR spectrophotometer, I want to thank Lazlo Hosszu and Andrea Hounslow. Rosie Staniforth and Silva Giannini have been great for their valuable advice on work-related matters and for their general optimistic attitude that has cheered me up in several occasions.

I am HUGELY grateful to Salva for his help throughout my PhD, I couldn't have done it without him…

If I enjoyed my life in Sheffield, it is thanks to several people:
From the Hunter group: Adam, Andrea S., Andrew (thanks for reading my thesis) Claudio, Cristiano, Dina, Julie, Gianni (double thanks: for his friendship and for his VIVA celebration…), Liam, Lisa, Miroslav, Paolo, Pablo, Rafel. From the *Italian community*: Andrea M., Andrea T., Alessandro, Alessio, Luca, Luisa, Manuela, Marco, Mauro.

Declaration

The work described in this thesis was carried out in the Department of Chemistry at the University of Sheffield, Sheffield from September 1998 to September 2001. The thesis is the result of my own work unless otherwise is stated in the acknowledgements.

Lilia Milanesi, August 2002.

Contents

Chapter 1

1. The Protein Folding Problem

1.1 Introduction

Experimental and theoretical investigations of protein folding are a central problem in modern structural biology, and work in this area is undergoing an explosive growth. Most natural proteins in solution are much smaller in dimension compared with polypeptides with random or repetitive sequences. Hence they are generally referred to as globular.[1] The folded structure of globular proteins is remarkably compact. In the interior, non-polar side chains (valine, isoleucine, phenylalanine, and alanine) predominate, and polar groups are paired in hydrogen bonds. Most ionised groups (aspartic acid, glutamic acid, and lysine) are on the surface of the molecule, exposed to the solvent. Proteins adopt well defined conformations determined by non-covalent interactions between atoms that are remote in the covalent structure. Types of non-covalent interactions that take place are Van der Waals' interactions, hydrogen bonds, electrostatic interactions, and hydrophobic interactions.[1-3] It is customary to discuss proteins in terms of four levels of structure:

- Primary structure, which is the covalent backbone, defined by amino acid sequence
- Secondary structure, which is the local conformation of the polypeptide backbone
- Tertiary structure, which is the overall topology of the folded polypeptide chain (domains)
- Quaternary structure, which is the aggregation of two or more polypeptide chains, either identical or different.

The amino acid sequence that defines a particular protein must fold to a specific conformation to be biologically active. The unknown relationship between primary and tertiary structure is the protein folding problem.

The process of protein folding has been known for about 60 years, but experimental approaches to study this problem started after the work of Wu and Anfinsen. Wu discovered the process of denaturation, and he proposed that chaotropic agents such urea and guanidine hydrochloride, or changes in pH or temperature, break the non covalent interactions that stabilise the folded state. As a consequence the protein assumes a more diffuse, flexible structure.[4] In 1973 Anfisen was awarded the Nobel Prize for discovering that unfolding transitions are reversible: upon removal of the denaturant, the protein reacquires the native structure.[5] The observation that the equilibrium between folded and unfolded state may be perturbed under denaturing conditions is the basis of the modern experimental approach to the study of protein folding. The evidence that proteins refold spontaneously *in vitro* demonstrates that the information needed to specify the 3D structure of a protein is contained in its amino acid sequence. It has therefore been suggested that, *in vivo*, the presence of specific enzymes, such as chaperones, disulphide and proline isomerases, assist the folding of a polypeptide chain but do not play a key role as the driving forces that define the fold.[2, 6] It is believed that chaperones prevent competing processes such as the aggregation of the nascent unfolded polypeptide chain, whilst disulphide and proline isomerases accelerate slow steps of the folding reaction such the *cis -trans* isomerisazion of the peptides bonds preceding proline residues or the rearrangement of disulfide bonds.[1, 7, 8] Thus the role of these enzymes is to ensure that proteins fold correctly in the minimum time required. In *E.coli* it has been estimated that a 350 amino acid residue protein should fold in less than 1 minute in order to be biologically active.[6] However, rate constants as short as milliseconds or even microseconds have been calculated for the folding of small proteins *in vitro*.[9-14]

This finding suggests that factors other than enzymes control the folding rate. The observation that folding can occur on such short timescales has another implication: a polypeptide chain cannot randomly sample all possible conformations to find the native state, because this mechanism would require a long time. The following example gives an idea of the timescale for such a scenario: if a polypeptide is made of 100 amino acid residues and only 2 conformations are available for each residue, there are about 10^{30} possible conformations. Assuming that only 10^{-11} s are required

to convert one conformation into another, a random search will require 10^{11} years.[15] Cyrus Levinthal proposed that proteins may fold on short timescales, because only conformations that meet specific kinetic and thermodynamic requirements populate the folding pathway.[15-17] The aim of experimental and theoretical approaches is to elucidate the structures of these conformational states and the dynamics of their formation. The greater the amount of experimental data, the easier it will be to test the accuracy of computational models which attempt to simulate folding pathways.[18-20] The ultimate goal is to develop computational methods able to predict the 3D structure of a protein from its amino acid sequence.[21] The availability of such a tool together with the current availability of DNA sequence information will enable researchers to correlate structure and function of proteins directly with the genes that encoded them.[22] Medical practice, biotechnology-related industry and phylogenetic study will be radically altered by these developments.[23, 24]

Misfolding of globular proteins in bacterial expression systems is a major problem in biotechnology and limits the production of protein-based therapeutics such as interleukin, interferons, insulin, antibodies and their fragments.[25, 26] The deposit of protein aggregates is a feature of a family of disease which includes spongiform encephalopathies and Alzheimer's disease.[27-30] The infectious agent of spongiform encephalopathies has been called *prion.* A prion has been defined as a small proteinaceus infectious particle, which resists inactivation by procedures that modify nucleic acids.[30-32] The discovery that proteins alone can transmit an infectious disease is surprising because all other known pathogens, such as viruses, bacteria and fungi, need nucleic acid to transmit an infection. Experimental data suggest that prions propagate by contacting normal cellular proteins and causing them to flip from a predominant α-helical conformation to a β-sheet conformation.[34-37]

Once a complete understanding of folding mechanisms is achieved it will be possible to identify which cellular processes cause misfolding of proteins and how nature has selected sequences that fold fast enough to prevent aggregation. This knowledge may be used to develop therapeutic strategies for misfolding-related illnesses and to improve our understanding of protein folding *in vivo*.

Table 1.1

Method	**Information**
Hydrogen exchange labelling and NMR	Formation of persistent hydrogen bonds, burial from solvent at individual sites
Far UV CD	Formation of secondary structure
Near UV CD	Immobilisation of aromatic residues in tertiary structure
IR	Formation of secondary structure
Intrinsic fluorescence	Environment of aromatic residues
Fluorescence quenching	Solvent accessibility of fluorophores
UV Absorbance	Environment of aromatic residues

Control over protein structure will allow the design of proteins with tailored functions or the use of natural proteins as components of new materials or systems.[38-42] Some attempts in this area have already been successful:

- De Grado and co-workers have used the folding motif of cytochrome b as guideline for the design and syntheses of a new multiheme protein.[43, 44]
- Synthetic manipulation of natural metalloproteins has been used to design biosensors that can detect traces of metals. These proteins are unfolded in absence of zinc, and a zinc dependent transition has been used to measure binding by observing changes in the fluorescence of fluorophores incorporated in a tolerant region of the protein.[45, 46]

1.2 Experimental approaches for the study of protein folding

Experimental methods to study protein folding are based on perturbing the equilibrium between the folded and unfolded states under denaturing conditions. Once initiated, the reaction is monitored by techniques that are sensitive to changes in the protein conformation. As shown in Table 1.1, each technique provides information on conformational changes at different levels of the protein structure, and therefore a range of them is used to obtain a complete picture of the refolding reaction.

Two experimental approaches are used to investigate folding reactions: one is based on equilibrium measurements and the second on kinetic measurements. In the first approach the structural changes of the protein in presence of denaturant are measured after the system has reached equilibrium. Measurements of unfolding transitions usually give a curve like the one shown in Figure 1.1: the protein unfolds completely within a narrow range of denaturant concentrations. This behaviour is the result of the cooperativity of the folded state: the interactions that stabilise the native state are interdependent and they are stronger than the sum of each individual interaction.[1, 47]

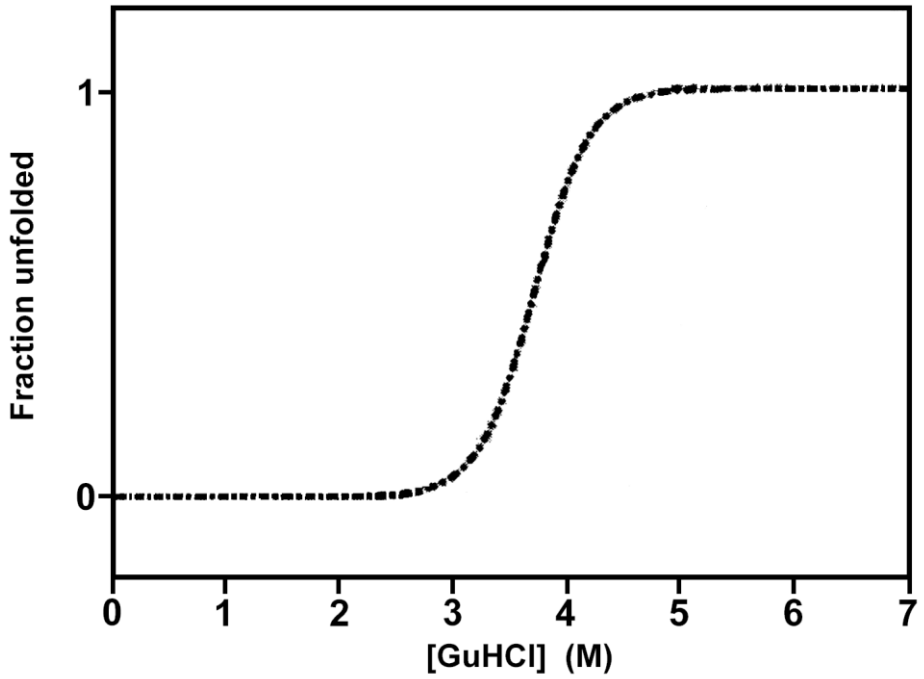


Figure 1.1. Guanidine Hydrochloride unfolding curve for cytochrome c (taken from reference 135). The fraction of unfolded molecule is calculated from fluorescence emission measurements.

Kinetic experiments are based on initiating the unfolding or refolding reaction by mixing the protein solution and the appropriate buffer with the use of a stopped-flow instrument. The time course of the folding reaction is then monitored to calculate the rate constant for folding and unfolding.

On the basis of experimental data, proteins may be classified into two groups:

- two-state folding proteins
- multistate folding proteins

In the first case, only the native state (N) and unfolded states (U) are populated in the refolding reaction, which can be described by the following equilibrium (Figure 1.2):

$$U \rightleftharpoons N$$

Figure 1.2

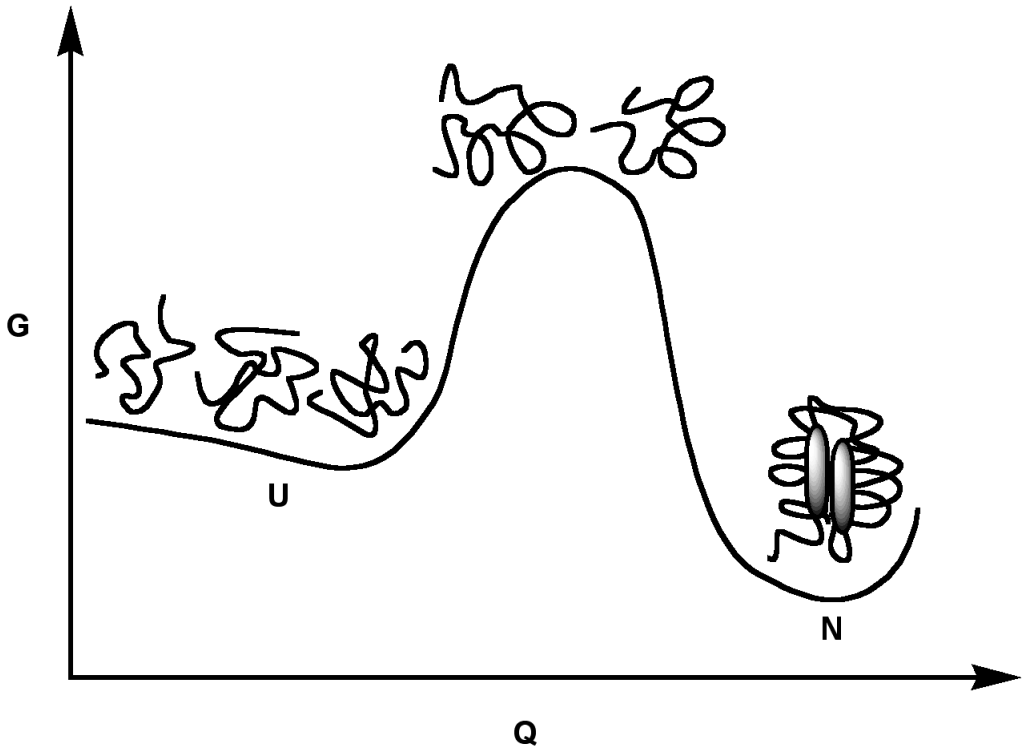


Figure 1.3. Free Energy (G) profile for the refolding of a 2 state protein. The conformational heterogeneity of the unfolded state and the predicted heterogeneity of the transition state is shown as a distribution of structures with a broad G value.

The energy profile for this kind of kinetics is depicted in Figure 1.3, where the free energy (G) is plotted as a function of a reaction co-ordinate, the fraction of native contacts (Q).[13, 48, 49]

In a two-state model, kinetic and equilibrium parameters are related according to equation 1.1:[50,51]

$$K_{(U \rightarrow N)} = \frac{[N]}{[U]} = \frac{k_f}{k_u} \qquad \text{Equ. 1.1}$$

where K is the equilibrium constant for the unfolding transition, and k_u and k_f are the rate constants for the unfolding and refolding reaction respectively. [50, 51] Thus equilibrium and rate constants may be used to obtain information on the relative stabilities of the folded and unfolded states. In a two-state model, the refolding rates and the stabilities of unfolded and folded states vary linearly with the concentration of denaturant.[52, 53] If an unfolding equilibrium measurement is considered, then equation 1.2 describes the effect of denaturant on the thermodynamic parameters:

$$\Delta G_{(U \rightarrow N)} = -RT\ln K_{(U \rightarrow N)} - m[\text{den.}] \qquad \text{Equ. 1.2}$$

where [den.] is the denaturant concentration and m is a measure of the solvent exposure of the protein surface during the unfolding transition.[13, 54, 55] Equations similar to equation 1.2 may be written using rate constants instead of the equilibrium constant. The values of m or ΔG obtained from kinetic and equilibrium measurements are equivalent.[51, 56, 57] Experimental data show that a two-state mechanism is usually a feature of single domain proteins, not more than one hundred residues long.[58]

The vast majority of proteins show a multistate folding mechanism in which one or more intermediates (I) are populated between the unfolded and folded states. [52, 53, 59] In the simplest case, only one intermediate is populated as depicted in Figure 1.4:

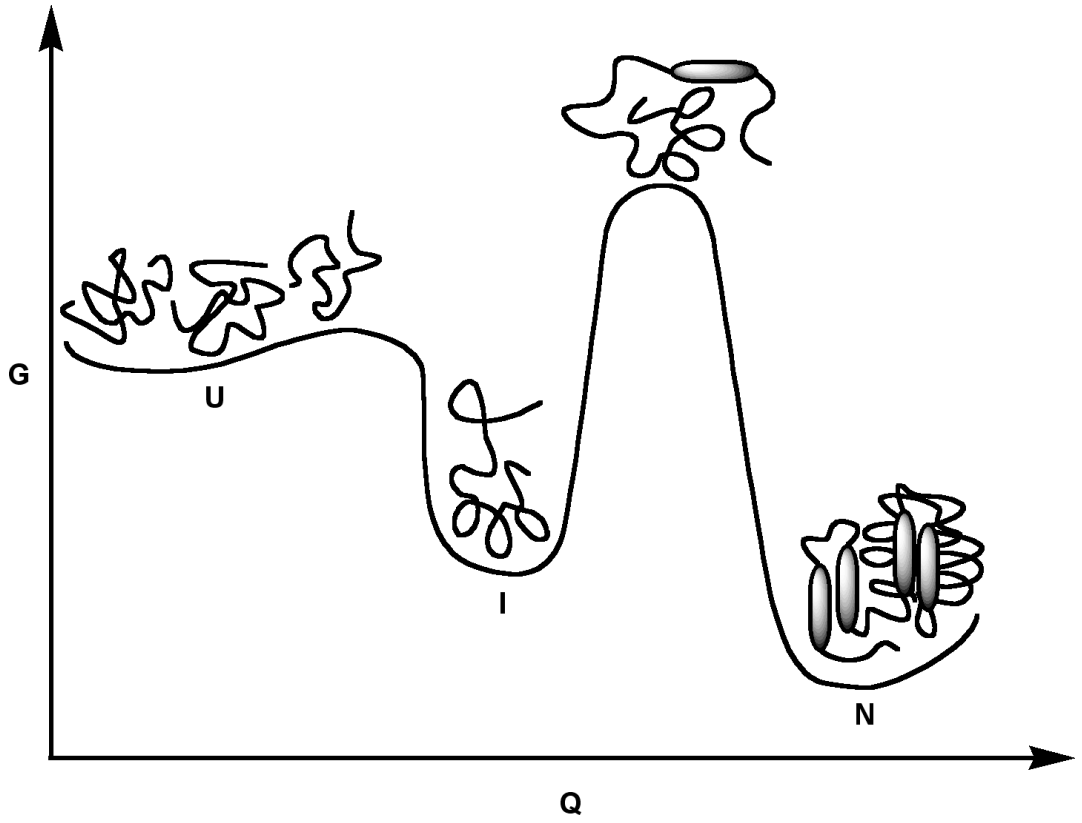


Figure 1.4. Free Energy (G) profile for a refolding reaction in which one intermediate (I) is populated.

This system has been described by the following mechanism (Figure 1.5), where the unfolded (U) and intermediate (I) states equilibrate faster than the intermediate and native state (N): [53]

$$U \rightleftharpoons I \rightleftharpoons N$$

Figure 1.5

Kinetic analyses of a multistate folding protein show that the rate of the refolding reaction does not vary linearly with the concentration of denaturant as observed in the case of a two-state folding protein. More complex equations must be used to calculate the refolding rates because in this case the refolding reaction takes place from I, and not directly from U.[57, 60, 61]

1.3 The role of intermediate states in protein folding

In the classical view of protein folding, most intermediate states are seen as an obligatory step to direct the unstructured polypeptide chain to the native state.[15, 53, 62-65] Thus the characterisation of the structure and stability of these intermediates is believed to be crucial in the understanding of the folding mechanism. The application of NMR spectroscopy methods such as hydrogen exchange labelling, to the study of folding reactions has provided structural details on these intermediate states.[53, 65, 66]

In a typical hydrogen exchange experiment, a protein is unfolded in a solution of guanidine hydrochloride in deuterated water. The refolding reaction starts by dilution of this solution with water. After different times, a short pulse to high pH is applied to the solution to promote fast deuterium to hydrogen exchange. The protein completes the folding reaction trapping the hydrogen/deuterium profile which can be analysed by 2D NMR. The hydrogen exchange for the amide groups that are not involved in hydrogen bonding occurs faster than the exchange of amide groups that are part of α-helices, β-sheet, or hydrophobic cores.[1, 67, 68] Thus it is possible to detect if intermediate states are populated as folding proceeds, and which residues are involved in the formation of hydrogen bonds.[69]

For some proteins, physical states different from the native and unfolded states have also been found under equilibrium unfolding conditions.[70-74] These states have been called *molten globule*. They have a high content of secondary structure, low tertiary contacts, and side chain packing is almost absent.[75-77] It has been found that for some proteins the conformational properties of a kinetic intermediate resembles that of an equilibrium molten globule state.[78-80] For the vast majority of proteins, however the study of kinetic intermediates is difficult because they form very quickly and are populated only for a short time.[53, 65]

One result from NMR methods is that the amide hydrogens of some proteins are protected from exchange in the initial phase of the refolding reaction.[81-83] Complementary methods such as circular dichroism, absorbance and fluorescence

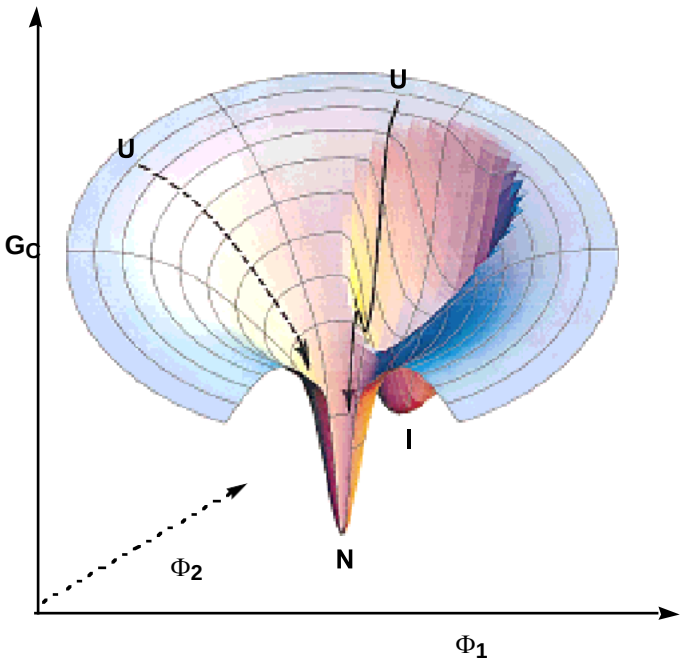


Figure 1.6. A folding scenario predicted by the energy landscape theory (adapted from reference 96).The free energy of a chain conformation (Gc) is plotted as a function of parameters which describe the degrees of freedom of the chain (Φ). The picture is an oversimplified representation because many degrees of freedom, and thus many lateral coordinates, exist for a real protein. For the same protein a two-state (dotted line) or a multistate (continuous line) pathway is possible under refolding condition.

spectroscopy show that a significant degree of secondary structure is found in early intermediates states.[84-87] These observations suggest that the polypeptide chain acquires part of the native topology within the time required to initiate the refolding reaction with traditional stopped-flow apparatus. This dead time is on the order of milliseconds, so any process taking place on this timescale or faster cannot be directly studied.

To solve the problem new methods have been developed which can trigger and observe a folding reaction on timescales as short as nanoseconds.[6, 48, 88] These fast techniques have been used to investigate the dynamics of secondary structure formation and the folding of very fast proteins. The results obtained so far show that elements of secondary and tertiary structure may be formed in nanoseconds (α–helices) or a few microseconds (small loops and β-turns) and that an increasing number of proteins fold in hundreds of microseconds without populating any intermediates.[89-93]

These findings have been used to support a new view of the protein folding problem which is based on the energy landscape theory developed by Bringselson *et al*.[94, 95] The theory describes folding as the movement of protein conformers on a funnel-like energy surface. The model predicts that different folding mechanisms are possible because the energy surface is not smooth. One possible landscape is depicted in Figure 1.6. Since the unfolded state contains a large number of conformers, the same protein under different refolding conditions may follow a two-state or a multistate mechanism with formation of intermediates. The landscape theory suggests that intermediates are misfolded structures that reduce the folding rate, because the protein must repair a conformational error to reach the native state. Thus the main difference compared with the classical view is that intermediate states are not obligatory steps for efficient folding, but rather kinetic traps on the energy landscape.[94-96]

According to the classical view of the folding problem, the existence of misfolded states and two state folding proteins is not enough evidence to prove that intermediates are not relevant to the progress of the folding reaction. Intermediates

may be undetected if they are less stable than the unfolded state.[53, 65, 97, 98] In addition, formation of misfolded species and multiple folding pathways has been observed only for some proteins or under specific experimental conditions:

- proteins that have disulfide bonds or proline residues show fast and slow folding pathways due to the formation of misfolded species with the incorrect bonding arrangements. In these cases, the refolding reaction occurs via two-state and multistate folding mechanisms. [99-101]
- proteins that show different kinetic behaviour depending on the experimental conditions used to study the refolding reaction. Cytochrome c folds via a multistate mechanism, if the pH of the refolding solution is above 6, because a deprotonated histidine residue binds to the heme replacing the native methionine ligand.[102] The misligation causes the formation of a misfolded structure which accumulates like an intermediate. The refolding reaction is complete in seconds. Under acidic conditions, no intermediate is detected. The refolding reaction is complete in milliseconds and may be observed only by fast techniques. The result in this case is a two-state folding behaviour without intermediates.[13, 103] Some proteins fold via a two-state or a multistate mechanism depending on the concentration of protein in the refolding solution. The slow folding rate is attributed to aggregation of the protein which occurs on the submillisecond timescale. Therefore if a traditional stopped-flow apparatus is used to trigger refolding, a transient aggregate may be mistaken for a reaction intermediate. [56, 104]

The landscape theory is based on simplified models of the polypeptide chain which assume that the movement of each peptide unit is not largely affected by that of nearest chain neighbours (th*e isolated-pair* hypothesis). However this assumption has been re-evaluated by experiment and calculation.[105, 106]

It has been suggested that segments of the polypeptide chain may adopt non-native contacts with subsequent formation of misfolded structures, in the early phase of the refolding reaction. However, as folding progresses these initial segments rearrange their structure due to the global motion of the entire polypeptide.[64]

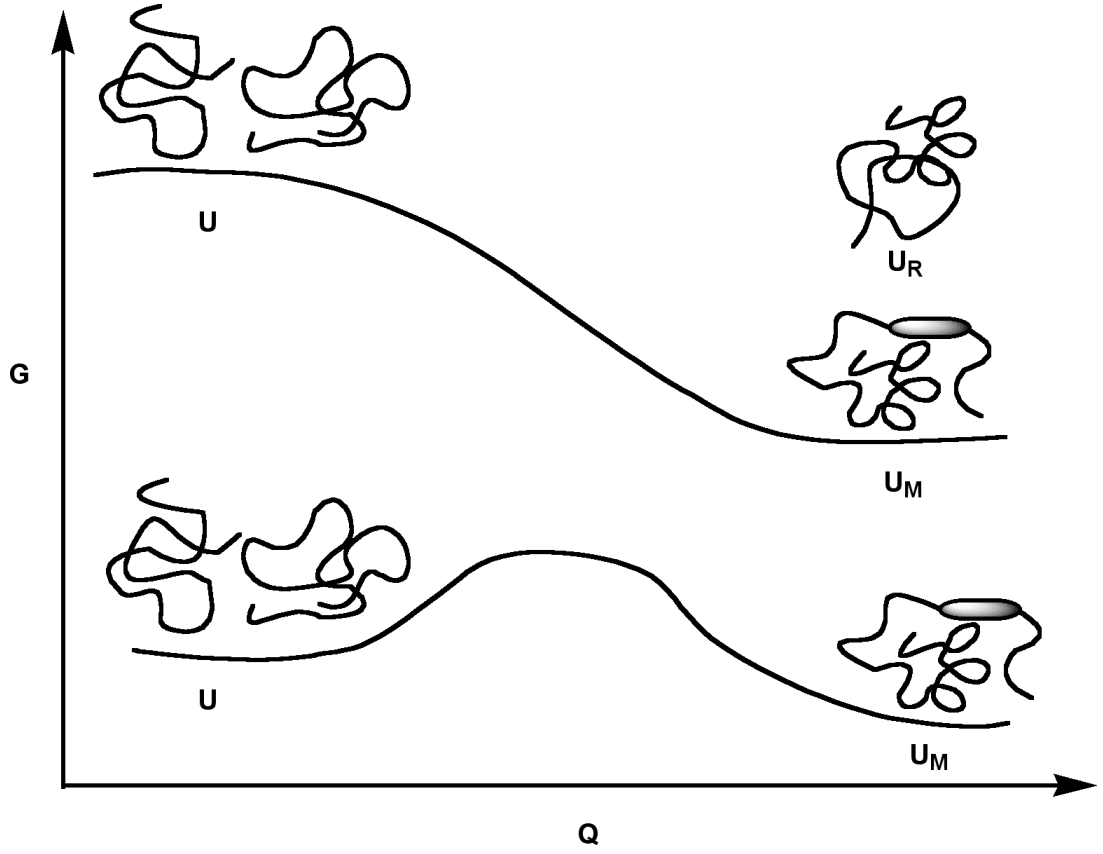


Figure 1.7. Collapse of the unfolded state under refolding conditions. Top curve: formation of a random globule (U_R) or molten globule-like state (U_M) is a process with no energy barriers. Bottom curve: collapse with an energy barrier between the unfolded state and the molten globule-like state.

Another problem is that computer simulations of the folding event, based on the energy landscape theory, have been able to predict only the first few nanoseconds of the refolding reaction.[107] So far, the longest time for the simulation of a 36 residue peptide has been 1 μs.[108] Thus it may be possible to better understand the role of intermediate states and the validity of theoretical models by monitoring the dynamics of the early stages of the refolding reaction, which are still not completely accessible to the fast techniques developed so far.

1.4 Submillisecond processes in protein folding

The unresolved initial phase of the refolding reaction has been identified with the collapse of the unfolded chain to a compact globule on the basis of fluorescence measurements, binding of hydrophobic dyes to the polypeptide chain and NMR studies.[6, 49, 53, 64, 87] It is believed that collapse is mainly driven by the hydrophobic effect: upon removal of the denaturant, the polypeptide chain buries hydrophobic groups in the interior, because their solubility in water is lower than in the presence of denaturants.[109-111] It has been suggested that hydrophobic collapse may lead to formation of a random globule state or a molten globule-like state.[6, 19, 90, 112] These two mechanisms are shown in Figure 1.7. If collapse occurs without formation of native like contacts, a random globule is expected. A molten globule-like state may result if collapse involves formation of native-like contacts. In this case, an energy barrier may exist between the unfolded and molten globule state and collapse may occur on the microsecond timescale.[8, 113] However if the formation of random or molten globule like states occur without energy barriers, collapse is estimated to be submicrosecond.[14, 114-116]

Since the kinetics of refolding on the submicrosecond timescale have not been resolved, the idea that hydrophobic collapse is the only mechanism of native structure acquisition is not universally accepted.[20, 117-119] In particular, it is unclear if collapse precedes secondary structure formation or *vice versa,* or if they both occur at the same time.[64, 76] It has been suggested that hydrophobic collapse may not be an efficient way to reach the native state: random globule states may be precursors of misfolded structures, and molten globule-like states, being more stable than the

unfolded state, may increase the energy barrier between the unfolded and native states.[117, 118]

Folding rates could be optimised if the unfolded chain acquires its native state via a mechanism different from hydrophobic collapse. The mechanisms proposed as theoretical models are:

- The framework model.[76, 120]
- The diffusion collision model.[121-123]
- The nuclear condensation model.[124-126]

The framework model and the diffusion-collision models suggest that secondary structure (α–helices and β-turns) is formed first, narrowing the conformational search space, so that the protein can fold in a finite time. According to the framework model, elements of secondary structure formed in the unfolded chain merge to an intermediate globular structure which evolves to a native conformation. In the diffusion-collision model, protein molecules consist of semistable isolated microdomains containing elements of secondary structure. The stability of the domains increases as a result of diffusion and collision events which put different domains in contact.

The nuclear condensation model predicts that the rate limiting step of the folding reaction is the formation of *nuclei*. A nucleation site is a short region of polypeptide chain that adopts native-like contacts and acts as an initiation core around which the rest of the structure may coalesce and grow.

1.5 Methods for fast initiation of protein folding

Development of ultrarapid mixing methods and advances in laser technology have improved the timescale on which a folding reaction may be initiated and monitored. Laser pulses as short as nanoseconds or picoseconds can be used to induce folding by different methods, such as triggering a photochemical reaction or a sudden change of the temperature of a protein solution.

1.5.1 Ultrarapid mixing

Although traditional rapid mixing methods have been limited to initiation of folding on the millisecond timescale, they have the advantage of being applicable to any protein which can be unfolded by chaotropic agents or pH jumps. Thus efforts have been made to reach mixing times of microseconds. Two such improvements are turbulent mixing and hydrodynamic focusing.[127-130]

The first method allows mixing of two solutions in 10 μs by generating a highly turbulent flow in the wake of a sphere placed at the junction of 2 concentric capillaries. Turbulence breaks the liquid into small volume elements, so that the solution diffuses over short distances, and the mixing time is reduced. However, the dead time of the instrument is 40-200 μs because, after mixing, the solution is not immediately accessible to the monitoring probe. This method provided evidence for a rapid collapse of cytochrome c in the first 50 μs of the refolding reaction.[128] In a later experiment, a mixing time of 1μs was achieved by focusing the solutions to diameters as short as 50 nm, so that the mixing time was further reduced.[130]

1.5.2 Laser induced temperature jump (laser T-jump)

In a laser T-jump experiment, folding or unfolding transitions are initiated by changing the temperature of a protein solution to a value above or below the range in which the native state is stable. If a protein undergoes cold denaturation above the freezing point of water, then a temperature increase of the cold solution, will induce folding, and the system will relax to a new equilibrium point. If the change in temperature is achieved more quickly than the time required for re-equilibration of the system, the kinetics of the folding reaction may be monitored. Since the temperature is increased by a laser pulse, time resolution is limited by the width of the pulse and by the time required to homogeneously heat the protein solution. The dead time of a typical T-jump apparatus is on the order of nanoseconds.[8, 88, 89] Data obtained with laser T-jump methods have shown that helices of small peptides are formed in hundred of nanoseconds. [90, 131] Laser T-jump experiments have been used

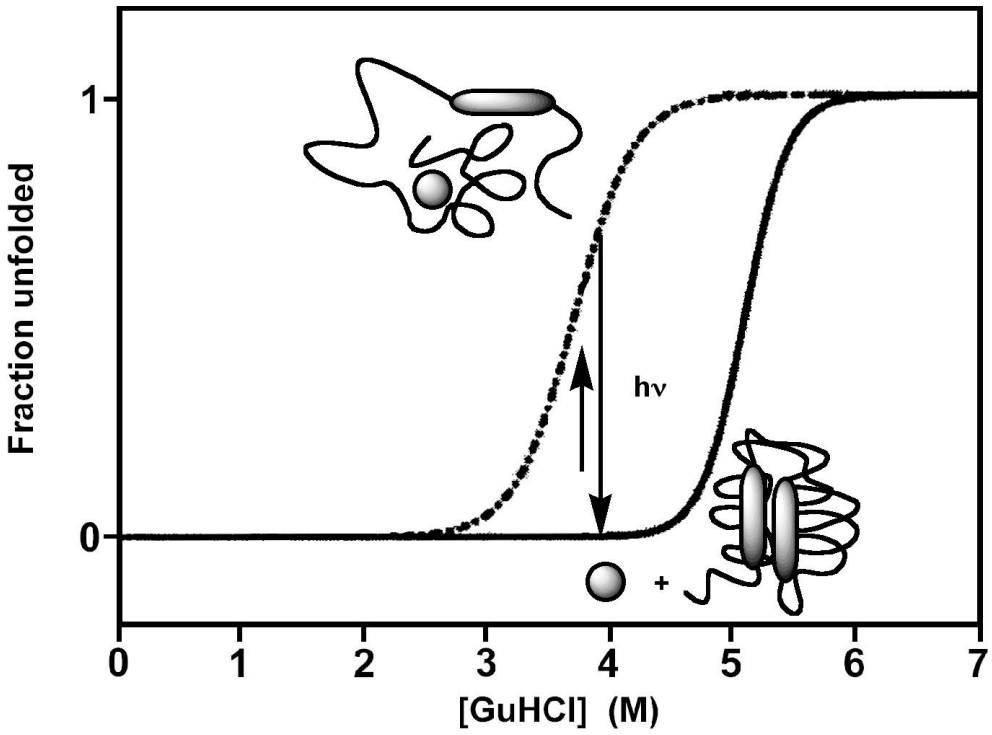


Figure 1.8. Guanidine hydrochloride unfolding curves of cytochrom c in the presence of carbon monoxide (depicted as a sphere). A light pulse initiates folding by photodissociating carbon monoxide from unfolded cytochrome c. The heme is omitted.

to monitor the collapse of unfolded apomyoglobin C and the unfolding of ribonuclease A on the nanosecond timescale.[14, 89, 115, 116, 132]

1.5.3 Photochemical methods

Photochemical methods make use of laser flash photolysis to initiate chemical reactions which can trigger folding or shift the equilibrium of the folding/unfolding transition. The time resolution of these methods can be on the submicrosecond timescale, depending on the experimental approach used.

1.5.3.1 Ligand photodissociation

This method is based on the observation that the binding of carbon monoxide to the heme of cytrochrome c is stronger in the unfolded state. The dependence of the binding event on protein conformation is illustrated by comparing the guanidine unfolding transition curves of Figure.1.8. Since the unfolding transition of cytochrome c in the presence of carbon monoxide occurs at a lower guanidine concentration than when no carbon monoxide is bound, photodissociation of carbon monoxide may initiate a folding reaction. The time resolution of this approach is limited only by the laser set up, because photodissociation is complete within picoseconds.

Jones *et al*. used a nanosecond laser to trigger refolding of cytochrome c from cytrochrome c bound to carbon monoxide.[133] They observed that in 40 μs a methionine binds the heme 50 residues away. This data and statistical mechanics were used to predict that a loop ten residues long is formed in 1 μs. It was estimated that 10 residues would be the length of the fastest forming loop because chain stiffness would slow down the formation of shorter loops.[134, 135] Since α helices and β turns are formed in nanoseconds and tens of microseconds respectively, it was suggested that formation of a small loop may be the rate limiting step of the protein folding reaction.[8, 11, 134]

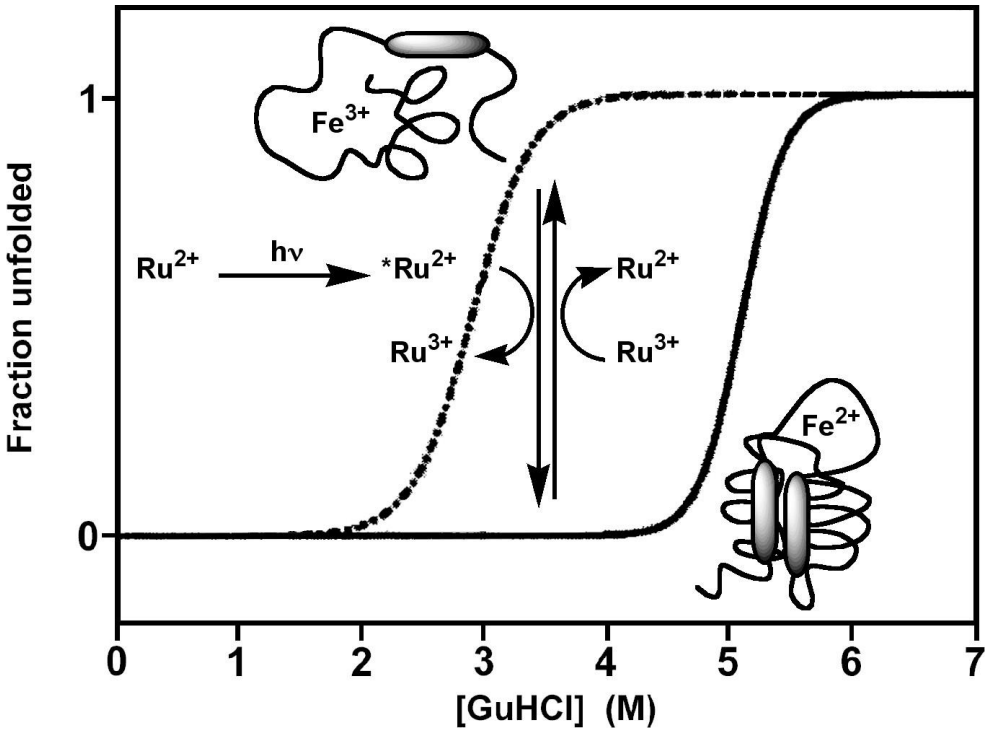


Figure 1.9. Guanidine hydrochloride unfolding curves for reduced and oxidised ferricytochrome c with a scheme of the photochemical method used to trigger folding. Ru represents a ruthenium salt. The heme is omitted

1.5.3.2 Photoinduced electron transfer

This method is based on the observation that some proteins may have different folding stability depending on their oxidation states. Therefore it is possible to find denaturant concentrations at which the protein in the oxidised state is more unfolded than in the reduced state or *vice versa*. This behaviour is shown in figure 1.9, where the guanidine unfolding transition for ferricytochrome c is depicted. Folding is initiated when photolysis induces rapid reduction of cytochrome c Fe^{III} unfolded to cytochrome c Fe^{II} folded. In Figure 1.9, a scheme of the events triggered by the laser pulse is shown: the first step is the photoactivation of a ruthenium salt which acts as an electron donor in its excited state. The next step is electron transfer to the ferro-heme causing reduction. The timescale for this process is less than 1 μs and sets the time limit for this method. In 1996 Pascher *et al*. using this method to initiate refolding of ferricytochrome c obtained data consistent with the CO photodissociation method of Jones *et al*.[135] In subsequent experiments, it was discovered that the use of NADH as a photochemical sensitizer instead of ruthenium salts, extended the timescale over which of the refolding reaction could be monitored from microseconds to seconds.[93, 136]

1.5.3.3 Photolabile crosslink

This approach makes use of solid phase synthesis to crosslink a synthetic peptide with a photolabile group. The crosslink constrains the peptide in a non-native conformation. Photolysis of the link releases the constraint and initiates the folding reaction. Since laser pulses as short as femtoseconds are available, the time resolution of this method is limited by the photochemistry of the ligand. Two experiments that use this approach have been described:

Disulfide crosslink.

In 1997, the relaxation of a peptide was triggered by laser flash photolysis.[137] (Figure 1.10). An aryl disulfide constrains the peptide in a non-α–helical conformation. Photolysis of the aryl disulfide cleaves the disulfide bond and liberates thiyl radicals. The radicals may undergo geminate recombination or diffuse apart. In the latter case,

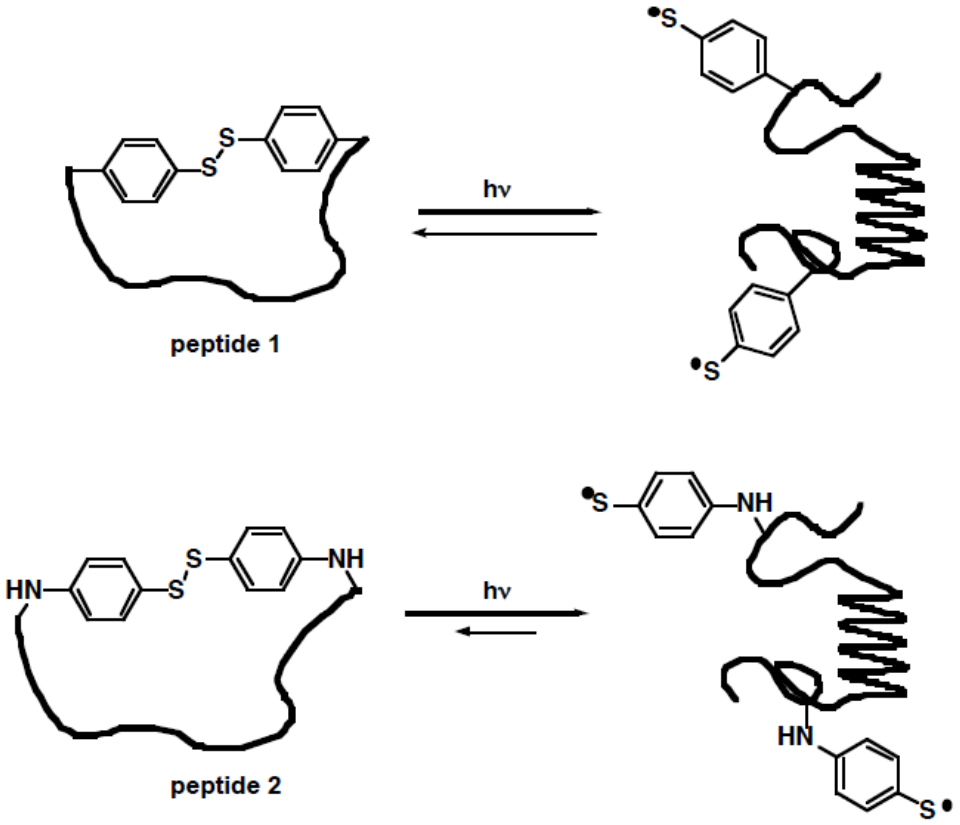


Figure 1.10. Disulfide crosslink. Folding is triggered by photolysis of an aromatic disulfide inserted by solid-phase synthesis. In peptide 2 recombination of thiyl radicals is slow enough to allow observation of secondary structure formation on the nanosecond timescale.

folding of the peptide may commence, because the disulfide constraint is released. Since aryl disulfides are cleaved in femtoseconds, the temporal resolution of this method is in the subpicosecond timescale.

Another advantage is the strong extinction coefficient of the resulting thiyl radical at 600 nm, a spectral region where peptides or proteins do not absorb. A potential problem is that disulfide photolysis is reversible: if the recombination of the radicals is faster than the relaxation of the peptide backbone, the folding reaction cannot be monitored.[138] This situation occurs in peptide 1 which uses a thiotyrosine moiety as an optical trigger. However on photolysis of peptide 2, where an aminothiotyrosine is used, the radical recombination is slower and it is possible to monitor the formation of the α-helix.[139] The different behaviour of the radicals from peptides 1 and 2 is consistent with previous studies on the parent disulfides which showed that thiyl radicals formed after photolysis of diphenyl disulfide recombine faster than thiyl radicals generated after photolysis of bis (*p*-aminophenyl) disulfide.[140, 141]

Benzoin crosslink

This approach, which is depicted in Figure 1.11, makes use of a benzoin linker to constrain a peptide in a cyclic form. Refolding of the peptide is observed on the nanosecond timescale after cleavage of the bezoin linker by laser photolysis.[142]

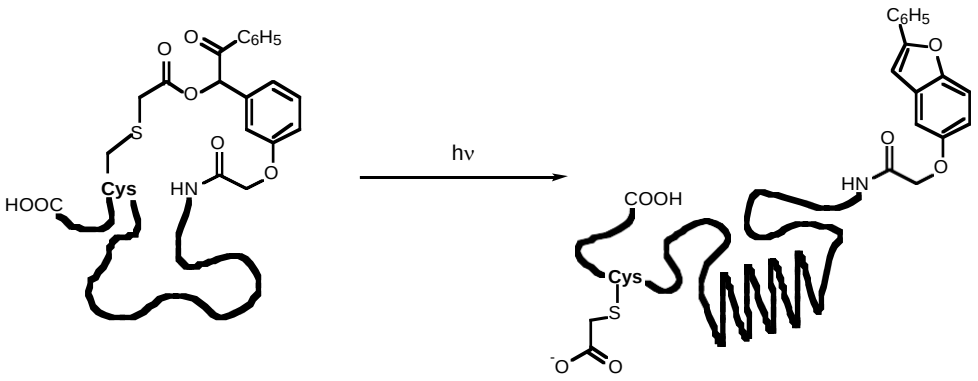


Figure 1.11. *Benzoin crosslink*. A small loop is formed in a short peptide by linking the N terminus to an internal cysteine with a benzoin moiety. Photolysis of the benzoin moiety yields the linear form of the peptide which forms secondary structure.

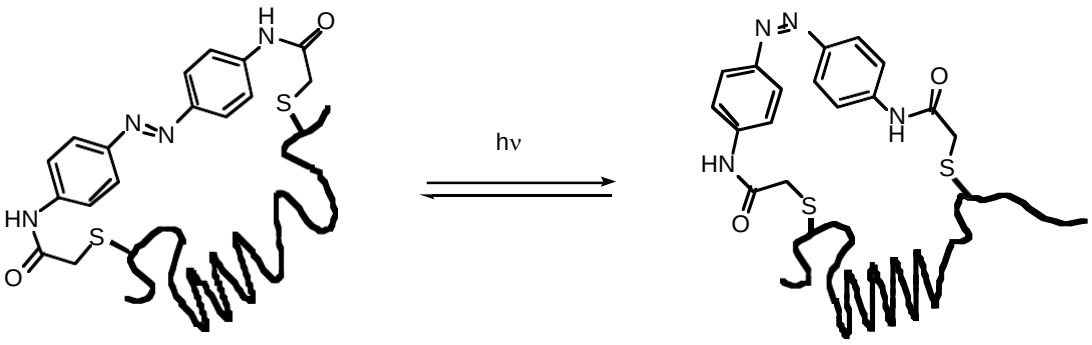


Figure 1.12. *Azobenzene crosslink.* A small loop is formed in a short peptide by linking two cysteine residues with an azobenzene cross-linker. Irradiation of the *trans* form of the crosslinked peptide induce isomerisation to the *cis* form which showed a substantial increase of α-helical structure**.**

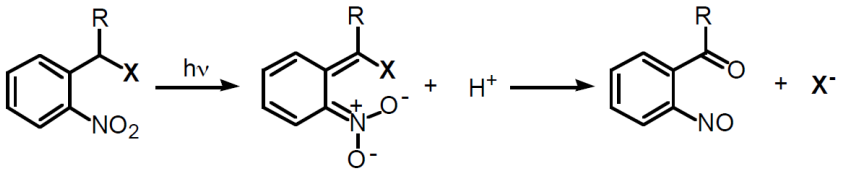
The main difference between this method and the previous one is that the triggering event is irreversible. This makes it possible to monitor the folding process over longer timescales, with no possibility of recombination.

Azobenzene crosslink

In this approach, which is depicted in Figure 1.12, a peptide is crosslinked at two cysteine residues with an azobenzene based ligand. Irradiation with 362 nm light triggers isomerisation of the *trans* form of the crosslinked peptide to the *cis* form which is more compatible with an α-helical structure.[143] Although α-helix formation was monitored five minutes after irradiation, this approach may be potentially used to trigger fast refolding reactions because azobenzene compounds photoisomerise in less than ten picoseconds.[144]

1.5.3.4 Photocaged ligand

The term photocaged is used to describe compounds which release an active species after photolysis. The active species may act as a substrate or trigger event for a range of biological processes.[145, 146] An example of a photocaged compound is depicted in Figure 1.13:



R = H, -CH3, -COOH, etc.

X = caged species

Figure 1.13

where the photolabile moiety is a 2-nitrophenyl derivative and the caged specis may be a secondary messenger such as ATP, cAMP, cGMP , a neurotransmitter such as acetylcholine or a cofactor such as calcium ions. In some of these systems, release of the caged molecule is complete in a few microseconds.[147-149]

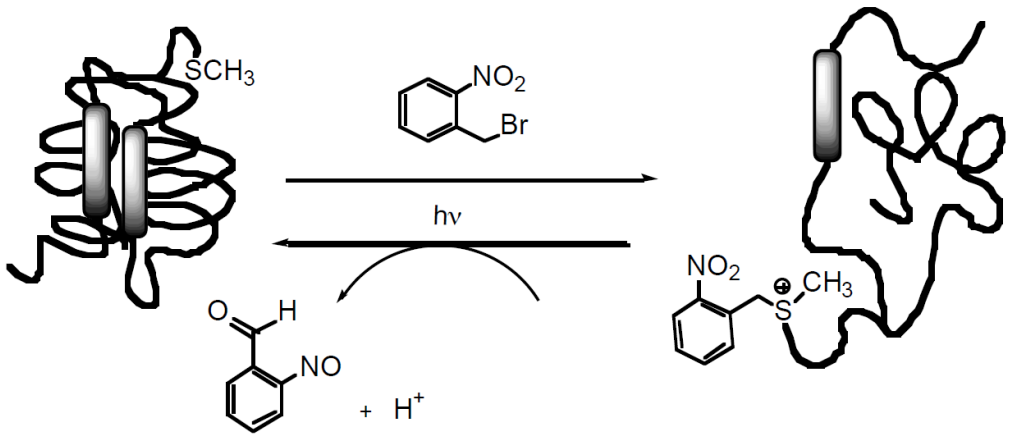


Figure 1.14. Photocaged protein. Chemical modification of a methionine with a photocleavable ligand destabilises the native state which can be recovered after photolysis.

Okuno *et al.* have designed a system in which the caged compound is a protein.[150] Figure 1.14 shows how this approach works. An *o*-nitrobenzyl group is introduced in a cytochrome protein via alkylation of a methionine residue. The modified protein is partially unfolded and starts to refold after removal of the photolabile moiety . The time limit of this method is the 3 μs required for the photorelease to be complete.

Photoinduced pH jumps.

This method makes use of a photocaged precursor which releases protons upon UV excitation. The subsequent change in pH may be used to trigger an unfolding reaction on a timescale limited by the kinetics of proton release. It has been shown that photolysis of *o*-nitrobenzaldehyde induces release of protons in nanoseconds.[151] Abbruzzetti *et al*. have photolysed a solution containing apomyoglobin and *o*-nitrobenzaldhyde with a nanosecond laser.[152] The approach is depicted in Figure 1.15. Acidification of the solution induces a transition from the native state to a *molten globule* state of the protein in a few nanoseconds. The same approach has been used to observe the kinetics of histidine deligation from the heme of cytochrome proteins.[153]

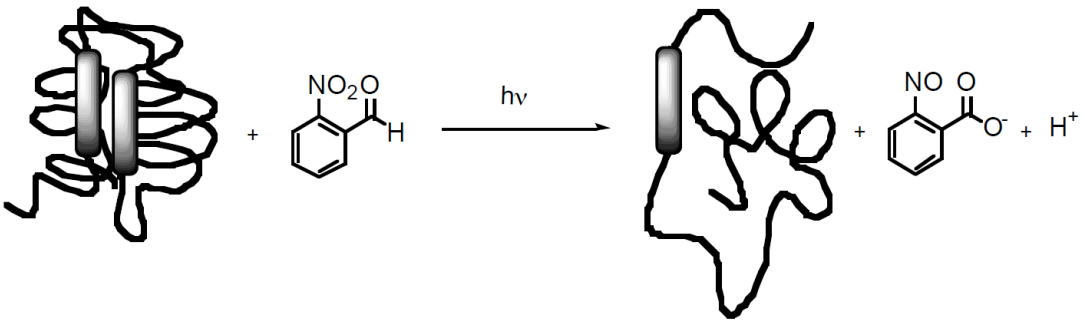


Figure 1.15. Photoinduced pH jump: photolysis of *o*-nitrobenzaldehyde induces a fast acidification of the solution which triggers formation of a *molten globule* state.

1.5.4 Conclusions

The main advantage of the approaches summarised in the previous section is the possibility to trigger folding reactions on a timescale of nanoseconds and in one case even femtoseconds. However all these methods have some disadvantages:

- Laser T-jumps are limited only to proteins which undergo cold denaturation. Since the properties of cold denatured states are not completely understood, the data obtained via this technique are subject to controversial interpretations.[48, 114]
- Photoinduced electron transfer and ligand photodissociation have been limited to heme proteins, and may be extended only to metalloproteins.
- Photochemical crosslink methods may have a wide applicability but until now they have been limited to the study of synthetic peptides.
- Photocaged methods and ultrarapid mixing have a wider applicability but a poorer time resolution when compared with the other methodologies.
- pH jump methods may be applied to trigger unfolding and folding reactions in the nanosecond timescale. However, although a wide range of proteins undergo pH-denaturation, for some of them the process is not reversible or induces formation of aggregates.[154, 155]

Since the fast folding methods developed so far have some limitations, new experimental approaches are required to trigger folding on short timescales, for a wider range of proteins.

1.6 Time-resolved techniques for the observation of folding processes

During a refolding reaction, an unfolded randomly arranged polypeptide undergoes structural changes before reaching the highly organised structure of the native state. Therefore folding reactions may be monitored by spectroscopic probes sensitive to changes in the protein conformation. Some of these techniques (listed in Table 1.1) such as circular dichroism, optical absorption, fluorescence and IR spectroscopy are available for time-resolved observation of protein dynamics on the nanosecond and

femtosecond timescale. A brief description of these technologies will be given in the following section.

UV/Vis. Absorbance

Time-resolved optical absorbance has been used to monitor folding of heme proteins. The method makes use of the changes in the Soret or visible bands of the prosthetic group, induced by changes in the polarity of the environment surrounding the heme. These changes are related to protein dynamics, because collapse of the unfolded chain transfers the heme from water to the hydrophobic protein interior.[11, 90, 93, 135]

Circular dichroism (CD)

The chiral properties of α-helices and β-sheets are different from those of random coils, and therefore their circular dichroism spectra have different shapes.[156] Although CD experiments have been widely used to assess the content of secondary structure in proteins, only very recently have time-resolved CD measurements been performed.[136, 157, 158] In one case secondary structure acquisition has been observed on the picosecond timescale.[159]

Infrared Absorbance (IR)

Infrared spectroscopy is used to monitor the stretching motions of the carbonyl group of the peptide backbone which absorbs between 1610 and 1680 cm^{-1}. This mode is sensitive to hydrogen bonding, dipole-dipole interactions and the geometry of the backbone. Therefore shifts in the frequency of the amide group IR bands are observed upon secondary structure formation or disruption. In a folded protein, α helices, β sheets, and turns shows bands centred at specific frequencies different from those of a random coil.[88, 156, 160] Thus the correlation between secondary structure and band position makes it possible to follow a folding reaction by infrared spectroscopy.[89, 132, 161, 162] An advantage of IR spectroscopy is that it allows conformational dynamics to be monitored with subpicosecond time resolution.[139, 163, 164]

Fluorescence

A correlation has been observed between the wavelength of fluorescent emission of aromatic residues and their exposure to water. Since the quantum yield of tryptophan

emission is high and its frequency of occurrence in protein sequence is low, the fluorescence of this residue is usually monitored. In the unfolded state tryptophan residues are exposed to water and fluoresce at λ_{max} = 350 nm, while in the folded state, they are buried in the protein core and the wavelength of maximum emission is centred at 330 nm. Thus, shifts of fluorescence emissions may be followed to monitor different stages of the refolding reaction, such as protein collapse.[88, 165, 166] Since emission occurs within nanoseconds of excitation, fluorescence methods have been used as time-resolved probes of the folding of a wide range of proteins.[103, 128, 167] Another reason for the extensive application of this technique is the availability of mutagenesis techniques, which allow insertion or deletion of tryptophan at specific positions of the protein sequence.

Another technique based on fluorescence is the time resolved measurement of fluorescence decay. This method is based on the observation that the fluorescent emission of tryptophan may be quenched by an intrinsic or extrinsic fluorophore. The efficiency of this process increases proportionally to the sixth power of the distance between the fluorophore and the tryptophan.[165] Therefore fluorescence decay measurements may provide data on the rate of contact formation between tryptophan and residues distant in the sequence.

Quenching of tryptophan by methionine, cysteine or heme have been used to investigate the folding reaction of various cytochromes and small peptides.[115, 127, 168, 169] An alternative approach is based on synthetic ligands that act as fluorescence quenchers when covalently attached to a protein or a peptide.[169-172]

1.7 References

1. T. E. Creighton, *Proteins: structure and molecular properties*, Chapter 6, 2nd Edition **1993**, W. H. Freeman and Company.

2. K. A. Dill, *Biochemistry* **1990**, *29*, 7134-7154.

3. W. Kauzmann, *Adv. Protein. Chem.* **1959**, *16*, 1-64.

4. H. Wu, Am. *J. Physiol.* **1929**, *90*, 562-566.

5. C. B. Anfisen, *Science* **1973**, *181*, 223-230.

6. W. A. Eaton, V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, J. Hofrichter, *Annu. Rev. Biophys. Biomol. Struct*. **2000**, *29*, 327-359.

7. M. J. Gething, J. Sambrook, *Nature* **1992**, *355*, 33-45.

8. F. U. Hartl, *Nature* **1996**, *381*, 571-580.

9. T. Schindler, M. Herrier, M. A. Marajel, F.X. Schmid, *Nat. Struct. Biol*. **1995**, *2*, 663-673.

10. G. S. Huang, T. G. Oas, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6878-6882.

11. S. J. Hagen, J. Hofrichter, A. Szabo, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12615-11617.

12. C. D. Waldburger, T. Jonsson, R.T. Sauer, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2629-2634.

13. T. R. Sosnick, L. Mayne, S. W. Englander, *Proteins: Structure, Function, and Genetics* **1996**, *24*, 413-426.

14. J. Sabelko, J. Ervin, M. Gruebele, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6031-6036.

15. C. Levinthal, *J. Chem. Phys.* **1968**, *65*, 44-45.

16. R. L. Baldwin, *Nature* **1994**, *369*, 183-184.

17. R. L. Baldwin, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2627-2628.

18. E. M. Boczko, C. L. Brooks III, *Science* **1995**, *269*, 393-396.

19. C. M. Dobson, A. Sali, M. Karplus, *Angew. Chem. Int. Ed.* **1998**, *37*, 868 –893.

20. T. Lazaridis, M. Karplus, *Science* **1997**, *278*, 1928-1931.

21. D. Baker, *Nature* **2000**, *405*, 39-42.

22. G. L. G. Miklos, G. M. Rubin, *Cell* **1996**, *86*, 521-529.

23. D. K. Pettit, W. R. Gombotz, *Trends in Biotechnology* **1998**, *16*, 343-349.

24. L. P. Encell, D. M. Landis, L. A. Loeb, *Nature Biotechnology* **1999**, *17*, 143-147.

25. M. M. Altamirano, C. Garcia, L. D. Possani, A. R. Fersht, *Nature Biotechnology* **1999**, *17*, 187-191.

26. J. W. Kelly, *Curr. Opin. Struct. Biol.* **1996**, *6*, 11-17.

27. J. P. Taylor, J. Hardy, K. H. Fischbeck*, Science* **2002**, *296*, 1991-1996.

28. D. J. Selkoe, *Science* **1997**, *275*, 630-631.

29. P. M. Harrison, P. Bamborough, V. Daggett, S. B. Prusiner, F. E. Cohen, *Curr. Opin. Struct. Biol*. **1997**, *7*, 53-59.

30. S. B. Prusiner, F. E. Cohen, *Annu. Rev. Biochem.* **1998**, *67*, 793-819.

31. T. Alper, W. A. Cramp, D. A. Haig, M. C. Clarke, *Nature* **1967**, *214*, 764-766.

32. S. B. Prusiner, *Science* **1982**, *216*, 136-144.

33. S. B. Prusiner, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13363-13383.

34. K. M. Pan, M. Baldwin, J. Nguyen, M. Gasset, S. B. Prusiner, F. E. Cohen, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10962-10966.

35. R. Riek, S. Hornemann, G. Wider, M. Billeter, R. Glockshuber, K. Wuthrich, *Nature* **1996**, *382*, 180-182.

36. G.S. Jackson, L. L. P. Hosszu, A. Power, A. F. Hill, J. Kenney, H. Saibil, C. J. Craven, J. P. Waltho, A. R. Clarke, J. Collinge, *Science* **1999**, *283*, 1935-1937.

37. F. L. Garcia, R. Zahn, R. Riek, K. Wuthrich, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8334-8339.

38. J. Skolnick, A. Kolinski, A. Godzik, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2099-2100.

39. N. Voyer, J. Lamothe, *Tetrahedron* **1995**, *51*, 9241-9284.

40. H. W. Hellinga, J. S. Marvin, *Trends in Biotechnology* **1998**, *16*, 183-189.

41. G. Tuchscherer, L. Scheibler, P. Dummy, M. Mutter, *Biopolymers* **1998**, *47*, 63-73.

42. S. V. Taylor, K. U. Walter, P. Kast, D. Hilvert, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10596-10601.

43. D. E. Robertson, R. S. Farid, C. C. Moser, J. L. Urbauer, S. E. Mulholland, R. Pidikiti, J. D. Lear, A. J. Wand, P. L. Dutton, W. F. De Grado, *Nature* **1994**, 368, 425-432.

44. C. T. Choma, J. D. Lear, M. J. Nelson, P. L. Dutton, D. E. Robertson, W. F. De Grado, *J. Am. Chem. Soc.* **1994**, *116*, 856-865.

45. G. K Walkup, B. Inperiali, *J. Am. Chem. Soc.* **1996**, *118*, 3053-3054.

46. G. K Walkup, B. Inperiali, *J. Am. Chem. Soc*. **1997**, *119*, 3443-3450.

47. M. H. Hao, H. A. Scheraga, *J. Mol. Biol.* **1998**, *277*, 973-983.

48. W. A. Eaton, P. A. Thompson, C. K. Chan, S. J. Hage, J. Hofrichter, *Structure* **1996**, *4*, 1133-1139.

49. W. A. Eaton, P. A. Thompson, C. K. Chan, V, Munoz, J. Hofrichter, *Curr. Opin. Struct. Biol.* **1997**, *7*, 10-14.

50. A. Matouschek, J.T. Kellis Jr., L. Serano, M. Bychroft, A. R. Fersht, *Nature* **1990**, *346*, 440-446.

51. S. E. Jackson, A. R. Fersht, *Biochemistry* **1991**, *30*, 10428-10435.

52. M. J. Parker, J. Spencer, A. R. Clarke, *J. Mol. Biol.* **1995**, *253*, 771-786.

53. A. R. Clarke, J. P. Waltho, *Curr. Opin. Struct. Biol*. **1997**, *8*, 400-410.

54. A. Matouschek, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7814-7818.

55. R. A. Staniforth, S. G. Burston, C. J. Smith, G. S. Jackson, I. G. badcoe, T. Atkinson, J. J. Holbrook, A. R. Clarke, *Biochemistry* **1993**, *32*, 3842-3851.

56. M. Oliveberg, *Acc. Chem. Res.* **1998**, *31*, 765-772.

57. N. Ferguson, A. P. Capaldi, R. James, C. Kleanthous, S. E. Radford, *J. Mol. Biol*. **1999**, *286*, 1597-1608.

58. S. E. Jackson, *Folding Design* **1998**, *3*, R81-R91.

59. Y. J. Tan, M. Oliveberg, A. R. Fersht, *J. Mol. Biol*. **1996**, *264*, 377-389.

60. M. J. Parker, J. Spencer, G. S. Jackson, S. G. Burston, L. L. P. Hosszu, C. J. Craven, J. P. Waltho, A. R. Clarke, *Biochemistry* **1996**, *35*, 15740-15752.

61. H. Roder, W. Colon, *Curr. Opin. Struct. Biol*. **1997**, *7*, 15-28.

62. A. P. Capaldi, C. Kleanthaus, S. E. Radford, *Nat. Struct. Biol*. **2002**, *3*, 209-216.

63. R. L. Baldwin, *Nat. Struct. Biol.* **1999**, *6*, 814-817.

64. C. R. Matthews, *Ann. Rev. Biochem.* **1993**, *62*, 653-683.

65. J. Rumbley, L. Hoang, L. Mayne, S. W. Englander, *Proc. Natl. Acad. Sci. USA* **2001**, 98, 105-112.

66. P. A. Evans, S. E. Radford, *Curr. Opin. Struct. Biol.* **1994**, *4*, 100-106.

67. H. Roder, *Methods Enzymol*. **1989**, *176*, 446-473.

68. P. Brandt, C. Woodward, *Biochemistry* **1987**, *26*, 3156-3162.

69. Y. Bai, T. R. Sosnick, L. Mayne, S. W. Englander, *Science* **1995**, *269*, 192-197.

70. K. P. Wong, L. M. Hamlin, *Biochemistry* **1974**, *13*, 2678-2683.

71. M. V. Jagannadham, D. Balasubramanian, *Febs Lett*. **1985**, 188, 326-330.

72. R. I. Gilmanshin, O. B. Ptisyn, *Febs Lett*. **1987**, *223*, 327-329.

73. Y.V. Griko, P. L. Privalov, S. Y. Venyaminov, V. P. Kutyshenko, *J. Mol. Biol.* **1988**, *202*, 127-138.

74. Y. Goto, A. L. Fink, *Biochemistry* **1989**, *28*, 945-952.

75. M. Ohgushi, A. Wada, *Febs Lett*. **1983**, *164*, 21-24.

76. O. B. Ptitsyn, *J. of Protein Chemistry* **1987**, *6*, 273-293.

77. K. Kuwajima, *Proteins: Structure, Function, and Genetics* **1989**, *6*, 87-103.

78. J. Baum, C. M. Dobson, P. A. Evans, C. Hanley, *Biochemistry* **1989**, 28, 7-13.

79. F. M. Hughson, P. E. Wright, R. L. Baldwin, *Science* **1990**, *249*, 1544-1548.

80. R. L. Baldwin, *Curr. Opin. Struct. Biol*. **1993**, *3*, 84-91.

81. H. Roder, G. A. Elove, S. W. Englander, *Nature* **1988**, *335*, 700-704.

82. A. Matouschek, L. Serrano, E. M. Meiering, M. Bycroft, A. R. Fersht, *J. Mol. Biol.* **1998**, *277*, 973-983.

83. S. T. Gladwin, P. A. Evans, *Fold. Des.* **1996**, *1*, 407-417.

84. K. Kuwajima, H. Yamaya, S. Miwa, S. Sugai, T. Nagamura, *Febs Lett*. **1987**, *221*, 115-118.

85. K. Kuwajima, S. Sugai, T. Sugawara, *Biochemistry* **1991**, *30*, 2698-2706.

86. G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky. A. F. Gripas, R. I. Gillmanshin, *Biopolymers* **1991**, *31*, 119-128.

87. A. F. Chaffotte, C. Cadieux, Y. Guillou, M. E. Goldberg, *Biochemistry* **1992**, *31*, 4303-4308.

88. R. H. Callender, R. B. Dyer, R. Gilmanshin, W. H. Woodruff, *Annu. Rev. Phys. Chem*. **1998**, *49*, 173-202.

89. R. B. Dyer, F. Gai, W. H. Woodruff, R. Gilmanshin, R. H. Callender, *Acc. Chem. Res.* **1998**, *31*, 709-716.

90. W. A. Eaton, W. Munoz, P. A. Thompson, E. R. Henry, J. Hofrichter, *Acc. Chem. Res.* **1998**, *31*, 745-753.

91. R. B. Dyer, W. H. Woodruff, R. Gilmanshin, R. H. Callender, S. Williams, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 3709-3713.

92. B Nolting, R. Golbik, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10668-10672.

93. J. R. Telford, P. Wittung-Stafshede, H. B. Graay, J. R. Winkler, *Acc. Chem. Res.* **1998**, *31*, 755-763.

94. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins: Structure, Function, and Genetics* **1995**, *21*, 167-195.

95. J. N. Onuchic, *Annu. Rev. Phys. Chem*. **1997**, *48*, 545-600.

96. H. S. Chan, K. A. Dill, *Proteins: Structure, Function, and Genetics* **1998**, *30*, 2-33.

97. R. L. Baldwin, *J. Biomol. NMR* **1995**, *5*, 103-109.

98. K. Inaba, N. Kobayashi, A. R. Fersht, *J. Mol. Biol.* **2000**, *302*, 219-223.

99. S. E. Radford, C. M. Dobson, P. A. Evans, *Nature* **1992**, *358*, 302-307.

100. W. A. Houry, H. A. Scheraga, *Biochemistry* **1996**, *35*, 11719-11733.

101. B. R. Rami, J. B. Udgaonkar, *Biochemistry* **2001**, *40*, 15267-15279.

102. S. R. Yeh, S. Takahashi, B. C. Fan, D. L. Rousseau, *Nat. Struct. Biol*. **1997**, *4*, 51-56.

103. S. W. Englander, T. R. Sosnick, L. C. Mayne, M. Shtilerman, P. X. Qi, Y. Bai, *Acc. Chem. Res.* **1998**, *31*, 737-744.

104. M. Silow, M. Oliveberg, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6084-6086.

105. S. Marqusee, V. H. Robbins, R. L. Baldwin, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5286-5290.

106. R. V. Pappu, R. Srinivasan, G. D. Rose, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12565-12570.

107. U. Mayor, C. M. Johnson, V. Daggett, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13518-13522.

108. Y. Duan, P. A. Kollman, *Science* **1998**, *282*, 740-744.

109. Dolgikh, R. I. Gillmanshin , E. V. Brazhnikov, V. E. Bychkova, Semisotnov, N. A., *Febs Lett*. **1981**, *136*, 311-315.

110. K. A. Dill, *Biochemistry* **1985**, 24, 1501-1509.

111. T. E. Chreighton, *Proteins: structure and molecular properties*, Chapter 7, 2nd Edition **1993**,W. H. Freeman and Company.

112. K. A. Dill, D. Shortle, *Annu. Rev. Biochem***. 1991**, *60*, 795-825.

113. S. J. Hagen, W. A. Eaton, *J. Mol. Biol*. **2000**, *301*, 1019-1027.

114. T. R. Sosnick, M. D. Shtilerman, L. Mayne, S. W. Englander, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 8545-8550.

115. R. M. Ballew, J. Sabelko, M. Gruebele, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5759-5764.

116. W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5897-5899.

117. A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10869-10874.

118. N. Schonbrunner, K. P. Koller, T. Kiefhaber, *J. Mol. Biol.* **1997**, *268*, 526-538.

119. J. K. Myers, T. G. Oas, *Nat. Struct. Biol.* **2001**, *8*, 552-558.

120. O.Ptisyn, *Protein. Eng*. **1994**, *7*, 593-596.

121. M. Karplus, D. L. Weaver, *Nature* **1976**, *260*, 404-406.

122. M. Karplus, D. L. Weaver, *Protein Science* **1994**, *3*, 650-668.

123. R. E. Burton, J. K. Myers, T. G. Oas, *Biochemistry* **1998**, *37*, 5337-5343.

124. D. E. Otzen, L. S. Itzhaki, N. F. Elmasry, S. E. Jackson, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10422-10425.

125. V. I. Abkevich, A. M. Gutin, E. I. Shakhnovic, *Biochemistry* **1994**, *33*, 10026-10036.

126. Z. Guo, D. Thirumalai*, Biopolymers* **1995**, 36, 83-102.

127. C-K. Chan, Y. Hu, S. Takahashi, D. L. Rousseau, W. A.Eaton, J. Hofrichter, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 1779-1784.

128. M. C. Ramachandra Shastry, J. M. Sauder, H. Roder, *Acc. Chem. Res.* **1998**, *31*, 717-725.

129. S. Takahashi, S. R. Yeh, T. K. Das, C. K. Chan, D. S. Gottfried, D. L. Rousseau, *Nat. Struct. Biol.* **1997**, *4*, 44-50.

130. L. Pollack, M. W. Tate, N. C. Darnton, J. B. Knight. S. Gruner, W. A. Eaton, R. H. Austin, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10115-10117.

131. P. A. Thompson, W. A. Eaton, J. Hofrichter, *Biochemistry* **1997**, *36*, 9200-9210.

132. C. M. Phillips, Y. Mizutani, R. M. Hoachstrasser, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7292-7296.

133. C. M. Jones, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11860-11864.

134. J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11426-11427.

135. T. Pascher, J. P. Chesick, J. R. Winkler, H. B. Gray, *Science* **1996**, *271*, 1558-1560.

136 E. Chen, P. Wittung-Stafshede, D. S. Kliger, *J. Am. Chem. Soc*. **1999**, *121*, 3811-3817.

137. H. S. M. Lu, M. Volk, Y. Kholodenko, E. A. Gooding, R. M. Hochstrasser, W. F. DeGrado, *J. Am. Chem. Soc.* **1997**, *119*, 7173-7180.

138. M. Volk, *Eur. J. Org. Chem*. **2001**, 2605-2621.

139. H. S. M. Lu, M. Volk, Y. Kholodenko, E. A.Gooding, R. M. Hochstrasser, W. F. DeGrado, *J. Phys. Chem. B* **1997**, *101*, 8607-8616.

140. T. W. Scott, S. N. Liu, *J. Phys. Chem.* **1989**, *93*, 1393-1396.

141. N. A. Borisevich, S. V. Mel' nichuk, S. A. Tikhomirov, G. B. Tolstorozev, *Ultrafast Phenomena in Spectroscopy*; Springer Series in Chemical Physics **1990**, *49*, 276 – 281. Editors: E. Klose, B. Wilhelmi

142. K. C. Hansen, R. S. Rock, R. W. Larsen, S I. Chan, *J. Am. Chem. Soc.* **2000**, *122*, 11567-11568.

143. J. R. Kumita, O. S. Smart, G. A. Wooley, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3803-3808.

144. J. Wachtveitl, T. Nagele, B. Puell, W. Zinth, M. Kruger, S. Rudolph-Bohner, D. Oesterhelt, L. Moroder, *J. Photochem. Photobiol. A Chem*. **1997**, *105*, 283-288.

145. A. M. Gurney, H. A. Lester, *Physiological Reviews*,**1987**, 583-617.

146. J. A. McCray, D. R. Trentham, *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 239-270.

147. A. P. Billington, N. Matsubara, W. W. Webb, G. P. Hess, *Techniques in Protein Chemistry III* **1992**, 417- 425, Ed. Ruth Hogue Angelett.

148. D. Ramesh, R. Wielboldt, L. Niu, B. K. Carpenter, G. P. Hess, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11074-11078.

149. T. Kuhn, H. Schwalbe, *J. Am. Chem. Soc*. **2000**, *122*, 6169-6174.

150. T. Okuno, S. Hirota, O. Yamauchi, *Biochemistry* **2000**, *39*, 7538-7545.

151. G. Bonetti, A. Vecli, C Viappiani, *Chem. Phys. Lett*. **1997**, *269*, 268-273.

152. S. Abbruzzetti, C. Viappiani, J. R. Small, L. J. Libertini, E. W. Small, *Biophysical Journal* **2000**, *78*, 405-415.

153. S. Abbruzzetti, C. Viappiani, J. R. Small, L. J. Libertini, E. W. Small, *J. Am. Chem. Soc.* **2001**, *123*, 6649-6653.

154. Y. Goto, L. J. Calciano, A. L. Fink, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 573-577.

155. D. E. Otzen, M. Oliveberg, *J. Mol. Biol.* **2001**, *313*, 479-483.

156. T. E. Chreighton, *Proteins: structure and molecular properties*, 190-191, 2nd Edition **1993**,W. H. Freeman and Company.

157. S. C. Biorling, R. A. Golbeck, S. J. Paquette, S. J. Milder, D. S. Kliger, *Biochemistry* **1996**, *35*, 8619-8627.

158. E. Chen, M. J. Wood, A. L. Fink, D.S. Kliger, *Biochemistry* **1998**, *37*, 5589-5598.

159. X.L. Xie, J. D. Simon, *J. Am. Chem. Soc.* **1990**, *112*, 7802-7803.

160. B. R. Singh, F. N. Fu, *Techniques in Protein Chemistry III* **1992**, 384- 397, Ed. Ruth Hogue Angelett.

161. D. Reinstadler, H. Fabian, J. Backmann, D. Naumann, *Biochemistry* **1996**, *35*, 15822-15830.

162. C. S. Colley, I. P. Clark, S. R. Griffiths-Jones, M. W. George, M. S. Searle, *Chem. Commun*. **2000**, 1493-1494.

163. P. Hamm, A. Lim, W. F. De Grado, R. M. Hoachstrasser, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2036-2041.

164. S. Woutersen, P. Hamm, *J. of Chem. Physics* **2001**, *115*, 7737-7743.

165. T. E. Creighton, *Proteins: structure and molecular properties*, 178-180 and 270-271, 2nd Edition **1993**,W. H. Freeman and Company.

166. M. Engelhard, P. A. Evans, *Folding & Design* **1996**, R31-R37.

167. J. B. Alexander Ross, C. J. Schmidt, L. Brand, *Biochemistry*, **1981**, *20*, 4369-4377.

168. C. K. Chan, J. Hofrichter, W. A. Eaton, *Science* **1996**, *274*, 628-629.

169. O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, T. Kiefhber, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9597-9601.

170. L. J. Lapidus, W.A. Eaton, J. Hofrichter, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7220-7225.

171. J. R. Telford, H. B. Gray, J. R. Winkler, F. A. Tezcan, *Biochemistry* **1999**, *38*, 1944-1949.

172. J. G. Lyubovitsky, J. R. Telford, H. B. Gray, J. R. Winkler, *J. Inorg. Biochem.* **1999**, *74*, 216-216.

Chapter 2

2. General Approach to Prepare Photochemically Caged Proteins for the Study of Fast Folding Kinetics

2.1 Aim of the project

As mentioned in the previous section, the current methodologies available for triggering protein folding on short timescales have limitations. The aim of this project is the development of optical triggers that may be applied to study folding of a wide range of proteins on timescales as short as picoseconds. Two approaches were planned:

1. Photocleavable crosslinking agents.
2. Photocleavable side chain functionalisation reagents.

The first approach is depicted in Figure 2.1. A bis-maleimide derivative **1** was designed as a new photochemical trigger. Since maleimides are specific reagents for thiols, reaction of **1** with a protein is expected to occur selectively at the cysteine residues.[1-10] The non-native crosslink so formed should constrain the protein in a conformation incompatible with a native-like folded structure. Thus cleavage of the disulfide bond by laser flash photolysis would trigger the refolding reaction, which may then be monitored by time-resolved fluorescence or IR spectroscopy. Previous studies on peptides have shown that reformation of the disulfide crosslink may compete with protein motions thus complicating the observation of the refolding reaction.[11] This problem can be solved by using a bis (*p*-aminophenyl) disulfide instead of diphenyldisulfide as the optical trigger.[12]

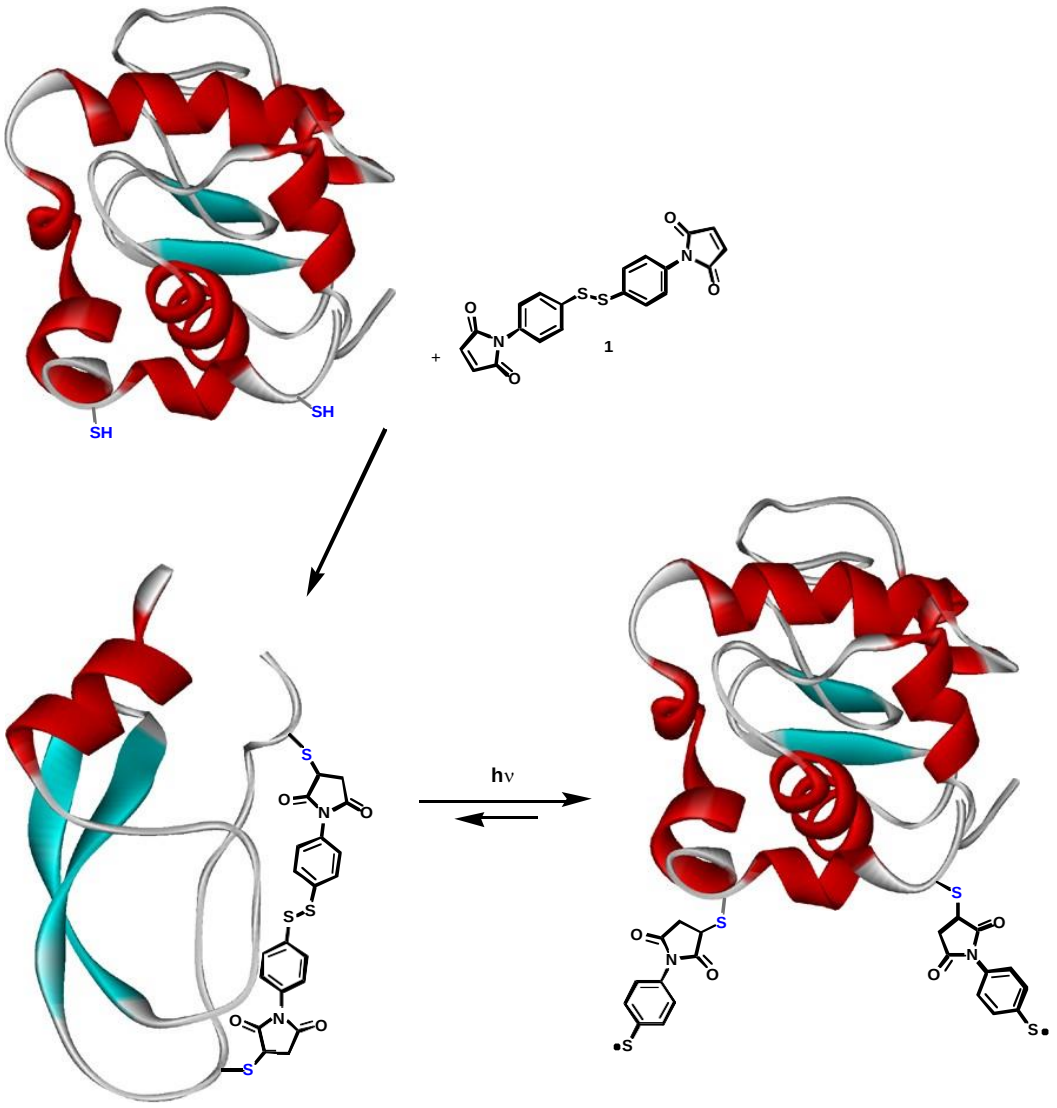


Figure 2.1. Schematic illustration of the strategy based on photocleavable crosslinking agents as optical triggers for the folding of proteins on a short timescale. The protein is forced into a non-native state by introducing an aromatic disulfide crosslink that is incompatible with the folded state. Cleavage of the aromatic disulfide by a femtosecond-UV laser pulse removes the non–native constrain and so initiates folding.

On the basis of these results, we have targeted a bis (*p*-aminophenyl) disulfide derivative as an optical trigger. The relative position of the disulfide and the maleimide moiety is a variable to be investigated, i.e. if maleimide and disulfide are *ortho* one may expect a lower recombination rate due to steric hindrance, whereas if they are *para*, reaction with the cysteine residues may be easier.

The second approach is depicted in Figure 2.2:

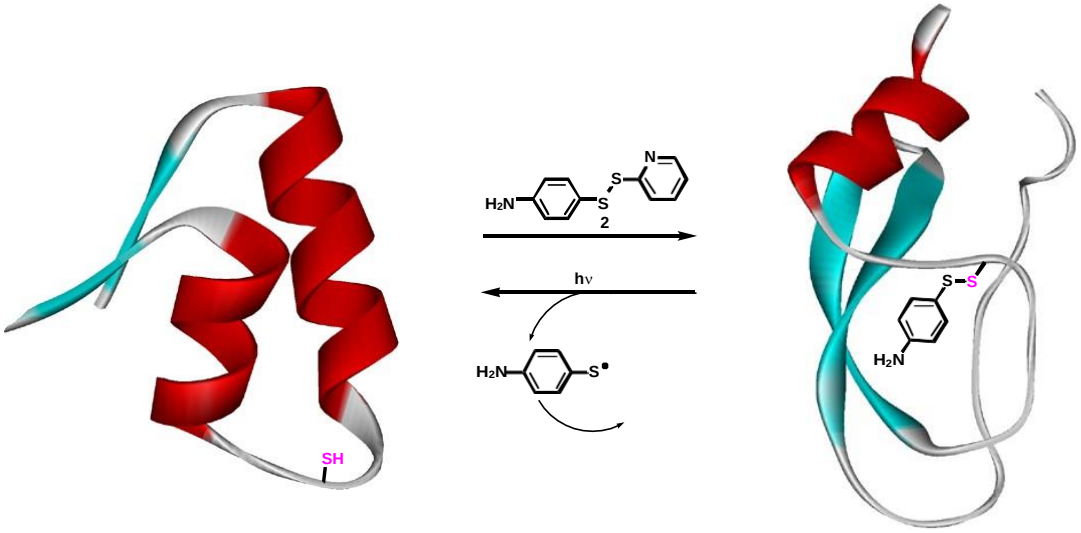


Figure 2.2. Schematic illustration of the strategy based on the introduction of a photocleavable group on a side chain. An aromatic disulfide is used to functionalise a single thiol of a protein. The functionalised form will have a less stable folded state than the parent protein and thus the refolding reaction may be triggered by cleavage of the aromatic disulfide with a femtosecond UV laser pulse.

An asymmetric disulfide, **2**, was designed as a new photochemical trigger. Reaction of **2** with the thiol group of a protein would result in the insertion of a photocleavable disulfide at the cysteine residue. The chemical modification may result in a perturbation of the protein stability, depending on the location of the cysteine residue. Removing the disulfide by laser flash photolysis would trigger relaxation of the protein to the native state which can be monitored by time-resolved spectroscopy. We expected that the thiol-disulfide exchange reaction would proceed as shown in Figure 2.2, because 2-thiopyridyl is a good leaving group.[13, 14]

We believed that the approaches described here would have two advantages:

- Subpicosecond time resolution.
- Application to any protein in which cysteine residues may be inserted by mutagenesis techniques.

We aimed to develop our method using the N-domain (1-174) of *Bacillus stearothermophilus* 3 phosphoglycerate kinase (NPGK) as the protein substrate.[15, 16] Previous studies indicated that at least one transient intermediate is populated in the refolding of NPGK.[17] Equilibrium amide exchange experiments showed that the intermediate acquires much of the native structure in the submillisecond phase of the refolding reaction.[18] These data suggest that NPGK is a good model system to study the early events in protein folding and thus to test the validity of our approach.

The structure of the isolated N terminal domain of PGK, determined by heteronuclear NMR spectroscopy, is very similar to that in the crystal structure of the entire protein (Figure 2.3).[16]

Since NPGK contains only one cysteine residue located in the hydrophobic core, the protein may be functionalised according to the approach depicted in Figure 2.2. However protein mutagenesis techniques will be used to insert cysteines at strategic positions in the protein sequence, in order to produce a substrate suitable for the approach shown in Figure 2.1.[19]

Figure 2.3. Three-dimensional structure of the N-terminal domain of 3-PGK (*B. Stearothermophilus*). The cysteine residue (C18) is coloured in yellow.

In this thesis, I will report the syntheses of disulphide reagents such as **1** and **2**, the result of laser flash photolysis experiments on these compounds and the attempts to incorporate these systems into *wt* (wild type) and mutants of NPGK.

2.2 References

1. G. E. Means, R. E. Feeney, *Chemical Modifications of Proteins* **1971**, Chapter 6, Holden – Day, San Francisco, California.

2. M. Machida, T. Sekine, Y. Kanaoka, *Chem. Pharm. Bull*. **1974**, *22*, 2642-2649.

3. Y. Kanaoka, *Angew. Chem. Int. Ed. Engl*. **1977**, *16*, 137-147.

4. H. J. Tae, *Methods in Enzymology* **1983**, *91*, 599-600.

5. K. Nakashima, K. Nishida, S. Nakatsuji, S. Akiyama, *Chem. Pharm. Bull*. **1986**, *34*, 1678-1684.

6. R. Aggeler, K. Chicas-Cruz, S. X. Cai, J. F. W. Keana, R. A. Capaldi, *Biochemistry* **1992**, *31*, 2956-2961.

7. J. T. Corrie, *J. Chem. Soc. Perkin Tans*. **1994**, *1*, 2975-2982.

8. D. T. McLachlin, S. D. Dunn, *Protein Expression and Purification* **1996**, *7*, 275-280.

9. S. K. Wright, R. E. Viola, *Analytical Biochemistry* **1998**, *265*, 8-14.

10. P. Schelte', C. Boeckler, B. Frisch, F. Schuber, *Bioconijugate Chem*. **2000**, *11*, 118-123.

11. H. S. M. Lu, M. Volk, Y. Kholodenko, E. A.Gooding, R. M. Hochstrasser, W. F. DeGrado, *J. Phys. Chem. B* **1997**, *101*, 8607-8616.

12. H. S. M. Lu, M. Volk, Y. Kholodenko, E. A. Gooding, R. M. Hochstrasser, W. F. DeGrado, *J. Am. Chem. Soc.* **1997**, *119*, 7173-7180.

13. G. E. Means, R. E. Feeney, *Chemical Modifications of Proteins* **1971**, 156-157, Holden – Day, San Francisco, California.

14. L. Ryden, J. Carlsson, *Protein Pufrification* **1989**, 253-258, Ed. J. C. Janson and L. Ryden.

15 . G. J. Davies, S. J. Gamblin, J. A. Littlechild, H. C. Watson, *Proteins: Struct. Funct. Genet*. **1993**, *15*, 283-289.

16. L. L. P. Hosszu , M. J. Parker, J. Spencer, M. Kelly, C. J. Craven, J. P. Waltho, A. R. Clarke, *Biochemistry* **1997**, *36*, 333-340.

17. M. J. Parker, J. Spencer, A. R. Clarke, *J. Mol. Biol*. **1995**, *253*, 771-786.

18. L. L. P. Hosszu, C.J. Craven, J. Spencer, M. J. Parker, M. Lorch, A. R. Clarke, *Nature Structural Biology* **1997**, *4*, 801-804.

19. C. Jelinska, *Equilibrium and kinetic studies of folding and unfolding of the N-terminal domain of phosphoglycerate kinase* **2002**, PhD thesis

Chapter 3

3. Design and Synthesis of a New Class of Disulfide Crosslinks for the Study of Protein Folding.

3.1 Introduction

Previous studies have shown that UV photolysis of aromatic disulfides cleaves the disulfide bond and results in a new absorption band in the visible spectrum which has been attributed to formation of aromatic thiyl radicals (Figure 3.1).[1-3] The dynamics of photodissociation have been studied by pump/probe spectroscopy experiments, which can generate light pulses at the wavelengths required to photolyse the disulfide bond and follow the temporal evolution of the thiyl radical absorption.[4-6]

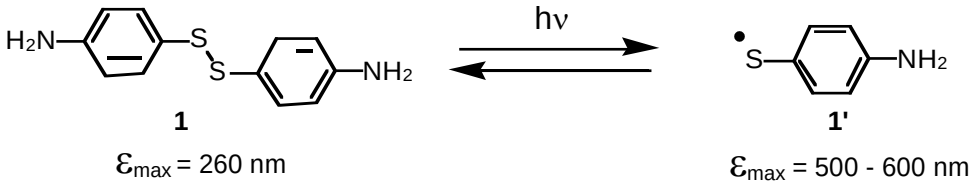


Figure 3.1

The main results of these studies are that formation of the thiyl radicals is complete in less than 1 picosecond, and their lifetimes range from picoseconds to microseconds depending on the structure of the radical and the properties of the solvent.[7-9] It is thought that the major factor which leads to a decrease in the radical population with time is the recombination of the radicals to generate the starting disulfide. Other decay channels are thought to be non-radiative processes, which compete with disulfide photocleavage.[8, 10] Evidence for recombination comes from the observation that the decay of the radical absorption is dependent on solvent viscosity and obeys second order kinetics.[7, 8, 10-13] The mechanism proposed by Noyes for radical recombination is depicted in Figure 3.2. The radicals generated after photolysis are subject to collision with the solvent that may lead to a recombination

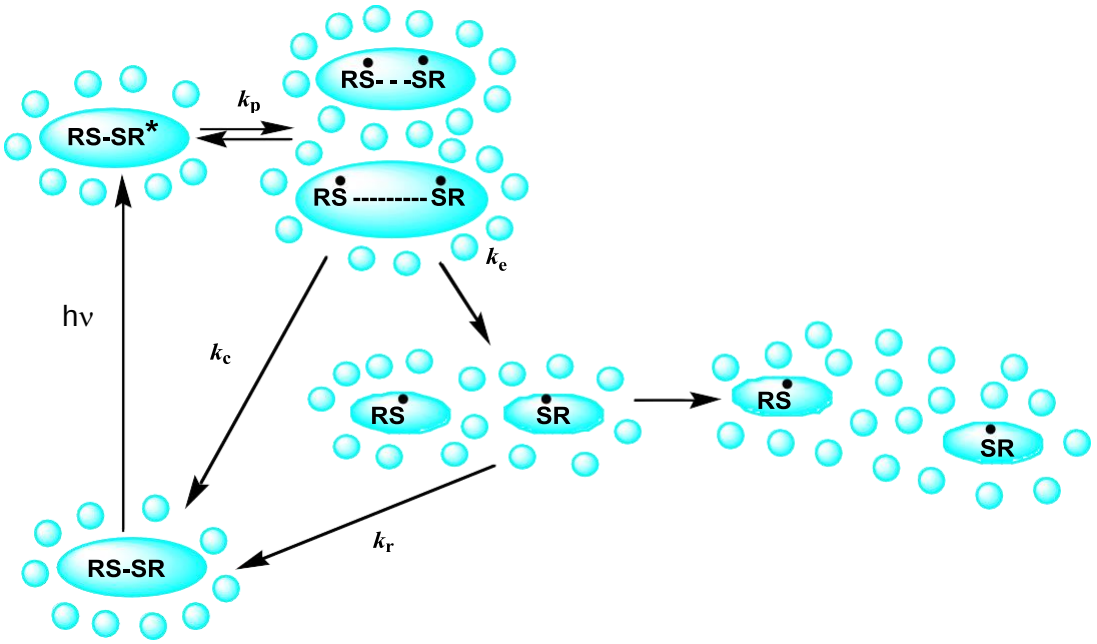


Figure 3.2. Simplified representation of the events following photolysis of a disulfide bond. Ellipsoids indicate the *solvent cage* effect. Solvent molecules are represented as small spheres. Rate constants for the processes following photolysis are also indicated:K*p* as the rate constant of photodissociation, K*e* as the rate constant of escape of the radical fragments from the solvent cage, K*c* as the rate constant of geminate recombination, K*r* as the recombination of radicals fragments that have escaped the solvent cage

of the radical fragments. The effect of the solvent in decreasing the quantum yield of photodissociation compared to the gas phase has been termed the *cage effect* , and it is depicted in Figure 3.2.[14, 15] Recombination may occur via different processes:

- From radical pairs which have never been fully separated or have been separated at a short distance and then encounter again. These processes are called *geminate* recombination and are associated with the rate constant k_c in Figure 3.2.
- From radical fragments which have escaped from the initial solvent cage and may recombine with the rate constant k_r in Figure 3.2.

Some of the radicals may diffuse apart and not undergo recombination. Since geminate recombination of thiyl radicals occurs on the picosecond timescale; this is the process of interest for the use of aromatic disulfides as optical triggers for protein folding.[16, 17]

The fraction of radicals which undergo geminate recombination is related to the efficiency of the cage effect and to the spin multiplicity of the radical according to:[18]

$$s = f_c f_e \qquad \text{Equ. 3.1}$$

The cage efficiency, f_c is a measure of the extent of recombination compared to escape from the solvent cage, and it is defined as:[19, 20]

$$f_c = \frac{k_c}{k_c + k_e} \qquad \text{Equ. 3.2}$$

The dependence of f_c on radical and solvent properties has been quantitatively described by Noyes:[21]

$$\left(\frac{1}{f_c} - 1\right) = \left(\frac{R_0 - 2b}{2b}\right) + \left(\frac{R_0}{2b}\right)\left[\left(\frac{A_T + \alpha A_E}{\alpha}\right)\left(\frac{1}{\eta}\right) + \left(\frac{A_T A_E}{\alpha}\right)\left(\frac{1}{\eta}\right)^2\right] \qquad \text{Equ. 3.3}$$

where
R_0 is the initial separation of the radicals
b is the radius of the radical
A_E is the translational energy of the separating radicals fragments after collision with the solvent
A_T is the kinetic energy of the separating radicals fragments after collision with the solvent.
η is the solvent viscosity
α is the radical reactivity

According to Equation 3.3, increasing the radical radius and solvent viscosity results in an increase in the cage efficiency. This behaviour is expected as a result of the reduced diffusional mobility (D) of the radical fragments according to the Stokes - Einstein Equation:

$$D = \frac{k_b T}{6\pi\eta b} \qquad \text{Equ. 3.4}$$

where k_B is the Boltzmann constant and T the absolute temperature.

The spin factor parameter, f_e in Equation 3.1, is introduced because the conservation of spin multiplicity requires that recombination takes place only from radical pairs in the singlet state. However, some of the excited radical pairs may populate triplet states during the time of photodissociation. Triplet radicals must overcome the spin barrier in order to recombine. If the rate of intersystem crossing from triplet to singlet state is slower than the cage lifetimes, triplet radical pairs will live longer than singlet pairs, which undergo cage recombination. Thus the higher the fraction of triplet radical pairs, the lower the yield of geminate recombination.[18, 22]

The time evolution of geminate recombination for phenyl thiyl radical pairs has been described by diffusion models based on several assumptions: there is no interaction between the solvent and the radicals; each radical is seen as a sphere without

rotational or vibrational degrees of freedom, and the evolution of spin multiplicity is neglected.[8, 1]

In the case of diphenyl disulfide, the diffusion models describe successfully the experimental decay traces of the corresponding thiyl radical absorption. In the case of bis (*p*-aminophenyl) disulfide, the diffusion model overestimates the fraction of radicals which undergo geminate recombination. The discrepancy between the model and the experimental data is more evident in polar than apolar solvents.[1, 8, 10, 23] To interpret these data, it has been suggested that a solvation effect may explain the extra stability of *p*-aminophenylthiyl radicals compared to phenylthiyl radicals. The description of the solvation effect is controversial and different mechanisms have been proposed. It is suggested that the relative contribution of the anionic form of the thiyl radical is increased in polar solvents (Figure 3.3).

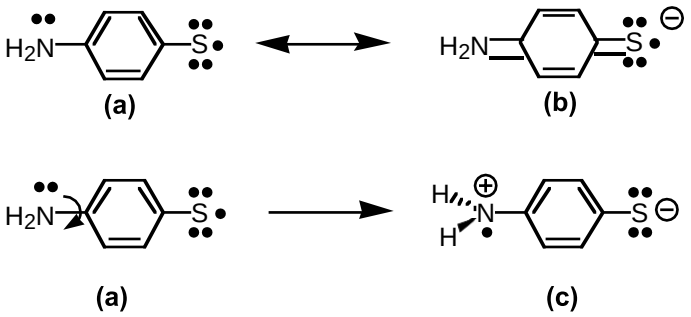


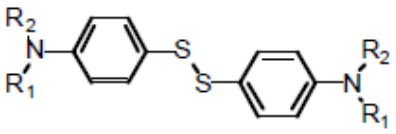
Figure 3.3. Neutral **(a)**, polar **(b)** and twisted form **(c)** of the *p*-aminophenylthiyl radical**.**

If the radical populates the polar form, then recombination may be reduced by the repulsion of the negative charges on the sulphur atoms.[12, 13] It has been suggested that the polar form may be further stabilised if charge separation occurs with nitrogen rotation, because in this case, reformation of the neutral form is blocked due to the orthogonal positioning of the π-system relative to the amino group.[7, 24, 25] The formation of the polar form is supported by the observation that the absorption spectra of the radical showed a band at 565 nm and a band at 610 nm, which increased with the polarity of the solvent.[7, 8, 24]

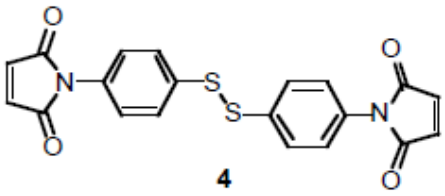
However, the existence of two different radical structures is in contradiction with more recent findings that shows no evidence of two distinct radical absorption bands in polar solvent on the subpicosecond and microsecond timescales.[4-6, 8] In these cases the authors explained the observed batochromic shift of the radical absorption as the result of solvent relaxation to the new equilibrium state, populated by polar forms of the radical which have an increased dipole moment compared to the ground state disulfide.[9, 26] According to this explanation, radical recombination is reduced if the dielectric relaxation of the solvent is faster than radical encounters. Since an analytical description of the effect of solvent polarity on geminate recombination is not available, it has been suggested that controversial findings for the transient absorption spectra of *p*-aminophenylthiyl radicals may arise from different experimental conditions.[27]

According to Equation 3.1, spin multiplicity factors should be taken into account to have a complete description of the recombination process. The studies of spin multiplicity effects on phenylthiyl radical recombination have shown that only in the case of bis (*p*-aminophenyl) disulfide can the triplet state be detected in time-resolved photolysis experiments. It has been suggested that spin inversion for these radicals is slower than in phenylthiyl radicals because the presence of an amino group may lift the degeneracy of the orbitals involved in the intersystem crossing from triplet to singlet states.[25, 28, 29] However the key result that emerges from different experiments is similar: *p*-aminophenylthiyl radicals are more stable than phenylthiyl radicals in polar solvents. It has also been shown that substitution of the amino group with another substituent (chloro, methoxy, *tert*-butoxy, hydroxy) reduces the stability of the corresponding thiyl radicals.[25]

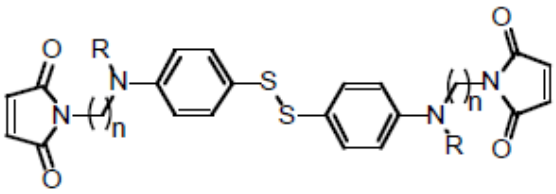
These photochemical properties make aromatic disulfides ideal systems for the study of fast biological processes. It has been shown that insertion of a bis (*p*-aminophenyl) disulfide moiety into the sequence of a small peptide allows triggering of α-helix formation on the nanosecond timescale.[16, 17] Here we set out to study the photochemistry of a series of bis (*p*-aminophenyl) disulfide derivatives that may be used to study fast processes in protein folding.



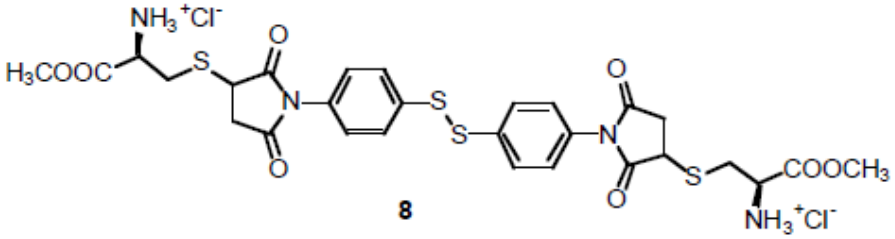
1, $R_1 = R_2 = H$
2, $R_1 = -CH_2CH_3$, $R_2 = H$
3, $R_1 = R_2 = -CH_2CH_3$



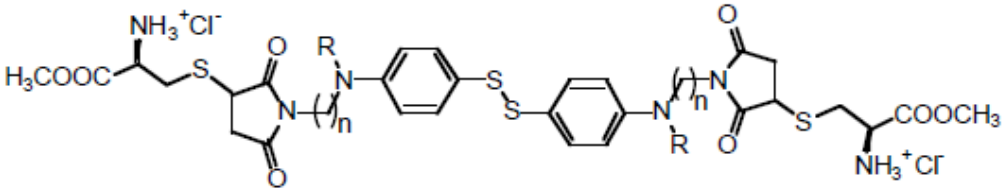
4



5, $R = -CH_2CH_3$, n = 2
6, R = H, n = 2
7, R = H, n = 1



8



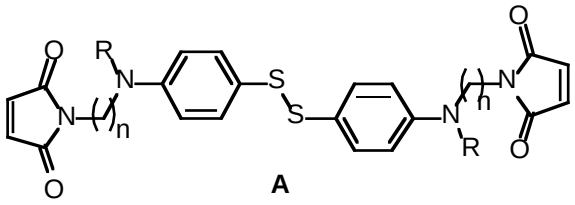
9, $R = -CH_2CH_3$, n = 2
10, R = H, n = 1

Figure 3.5

3.2 Results

3.2.1. Approach

The use of maleimide derivatives of general formula **A** (Figure 3.4) as optical triggers for protein folding require that the recombination of the corresponding thiyl radicals is slower, or competitive, compared to the motion of the protein to which the radical is attached.[16, 17, 22] Thus a number of synthetic strategies were explored to determine the effect of the positioning of the the maleimide moiety on the rate and efficiency of the photolysis reaction (Figure 3.5).



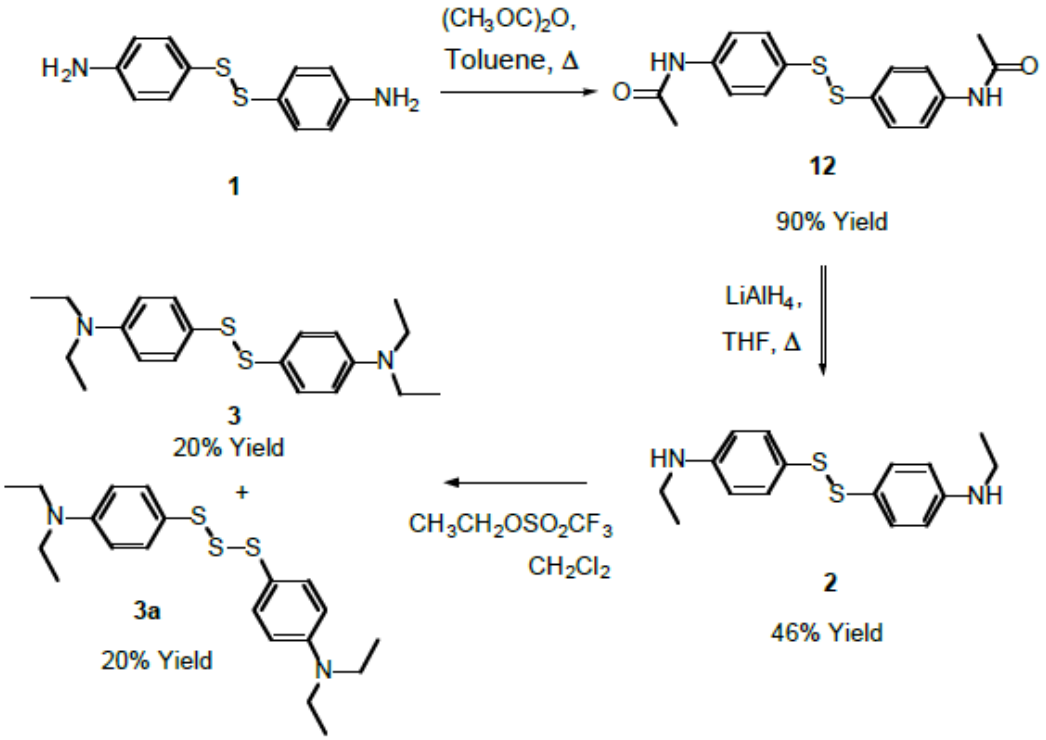
A

R = H , Alkyl

n = 1, 2

Figure 3.4

A series of model compounds such as **1 - 3** were prepared in order to test if functionalisation of the nitrogen of bis (*p*-aminophenyl) disulfide affects the recombination rate of the corresponding thiyl radicals. Then the maleimide derivatives **4 - 7** were designed on the basis of their synthetic accessibility. Compound **4** was prepared to test if changing the electronic properties of the aniline nitrogen significantly affects the efficiency of radical recombination. Derivatives **5 - 7** were designed to see if the influence of the maleimide unit on the properties of the disulfide chromophore can be reduced by moving it away from the aniline nitrogen. The maleimide derivatives (**4**, **5**, and **7**) were functionalised with cysteine methyl ester in order to study the photolysis reaction in water and to get insights into the problems that we may encounter when working with proteins. Except for compound **6**, the syntheses of the aromatic disulfides shown in Figure 3.5 was successful and their photochemistry was compared with that of bis (*p*-aminophenyl) disulfide (**1**).



Scheme 3.1

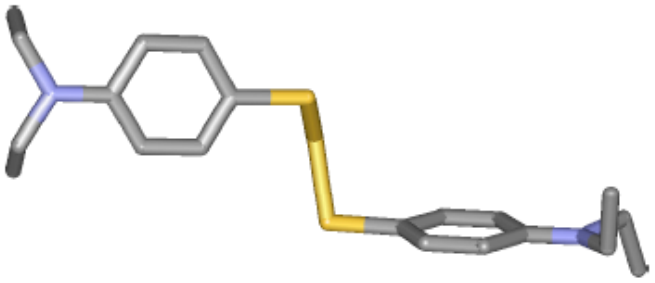
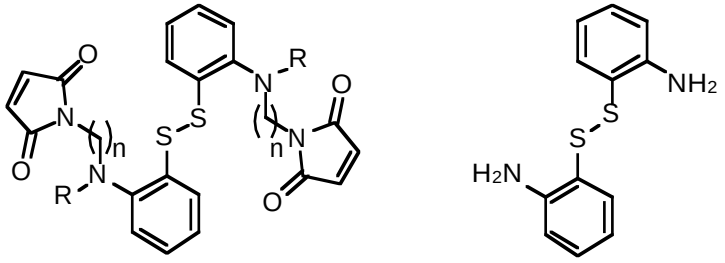


Figure 3.7. X-ray crystal structure of **3a**.

One approach to slow the rate of radical recombination is to introduce steric hindrance around the disulfide bond by placing substituents in the *ortho* position. In this case, maleimide derivatives of general formula **B** may be used as optical triggers for protein folding (Figure 3.6).



B **11**

R = H , Alkyl
n = 1, 2

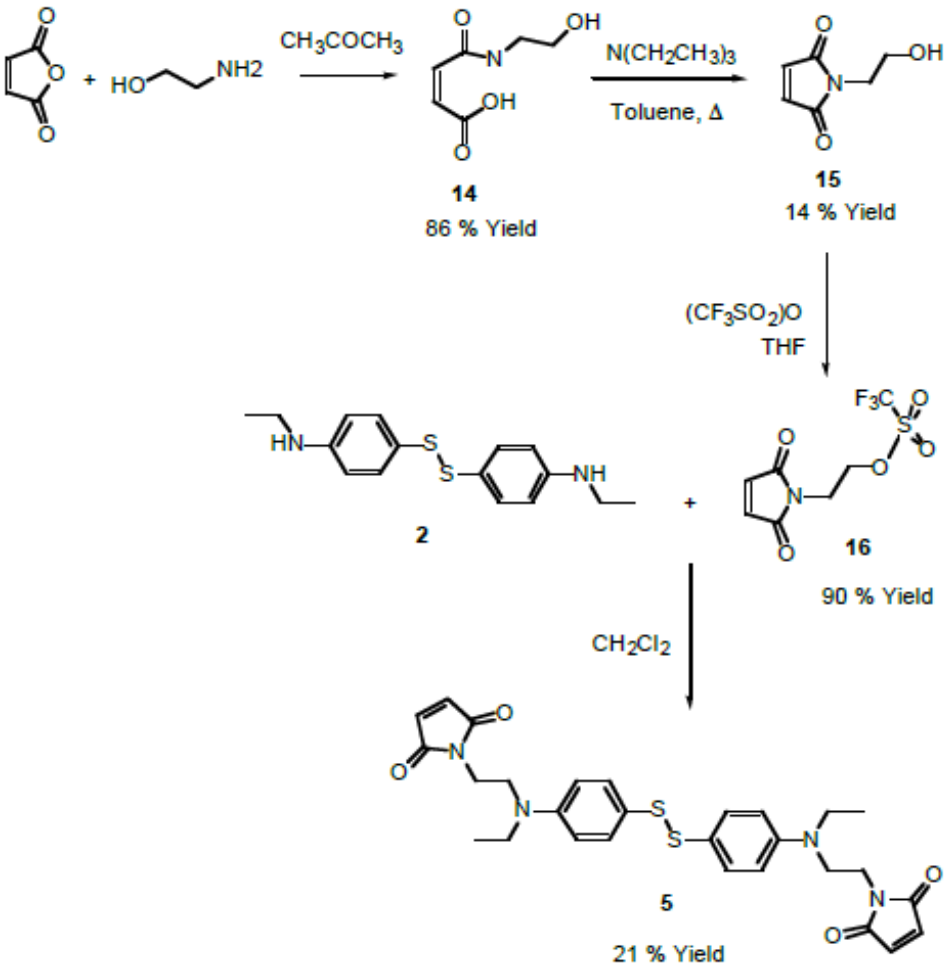
Figure 3.6

This approach will only work if the stability of *o*-aminophenylthiyl radicals is comparable with that of *p*-aminophenylthiyl radicals. Since there is no literature data on the stability of *o*-aminophenylthiyl radicals on the picosecond time scale, photolysis of bis (*o*-aminophenyl) disulfide (**11**, Figure 3.6) was also carried out.

3.2.2 Synthesis

Compound **2** was obtained by reaction of **1** with acetic anhydride followed by reduction with lithium aluminium hydride of the intermediate **12**. Compound **3** was obtained from **2** using trifluoro-ethanesulfonic acid ethyl ester as an alkylating agent (Scheme 3.1). The yield of **3** was lowered by the formation of the side product **3a** which was separated by column chromatography. The structure of **3a** was confirmed by X-ray crystallography (Figure 3.7).

Syntheses of **4** (Scheme 3.2) was carried out by mixing **1** with maleic anhydride in toluene at reflux followed by cyclization of the acid **13** using a mixture of acetic anhydride and sodium acetate as dehydrating agents. Initial attempts to synthesise **5**



Scheme 3.3

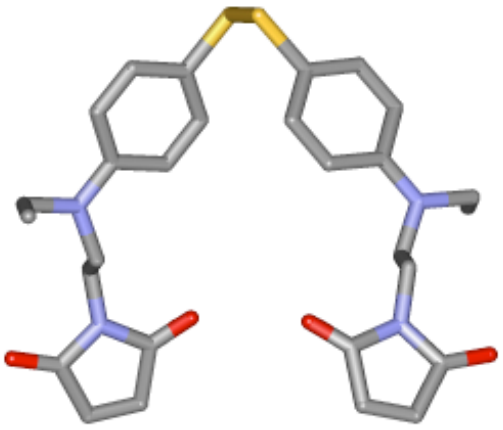
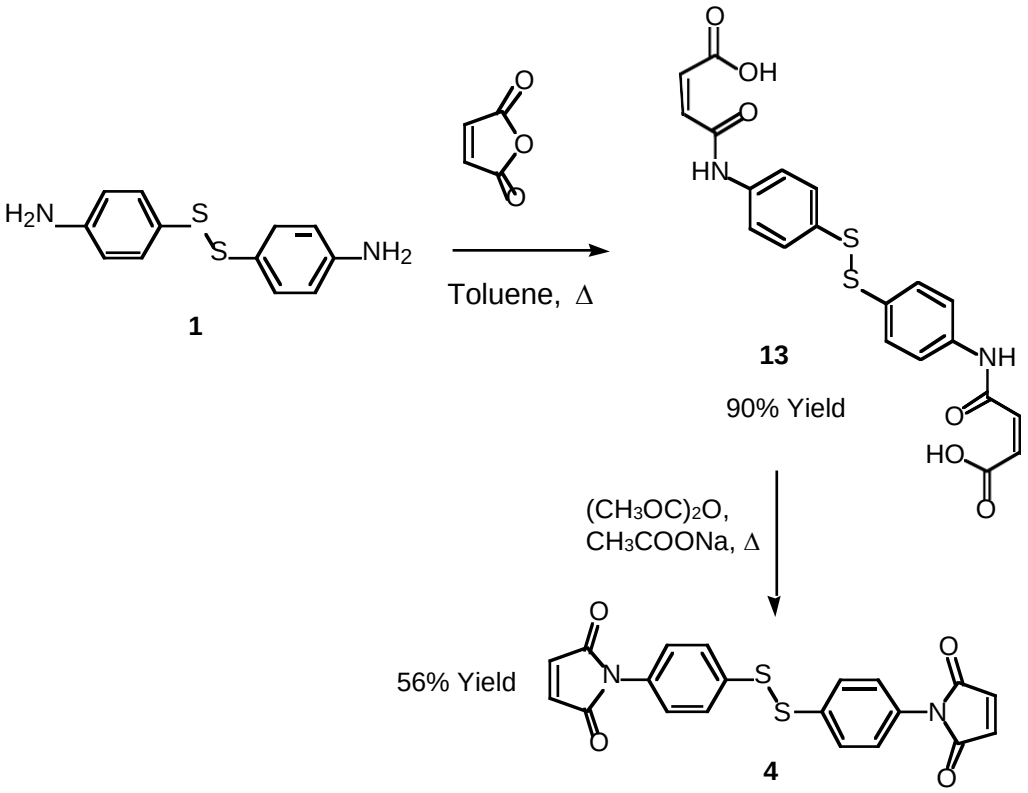


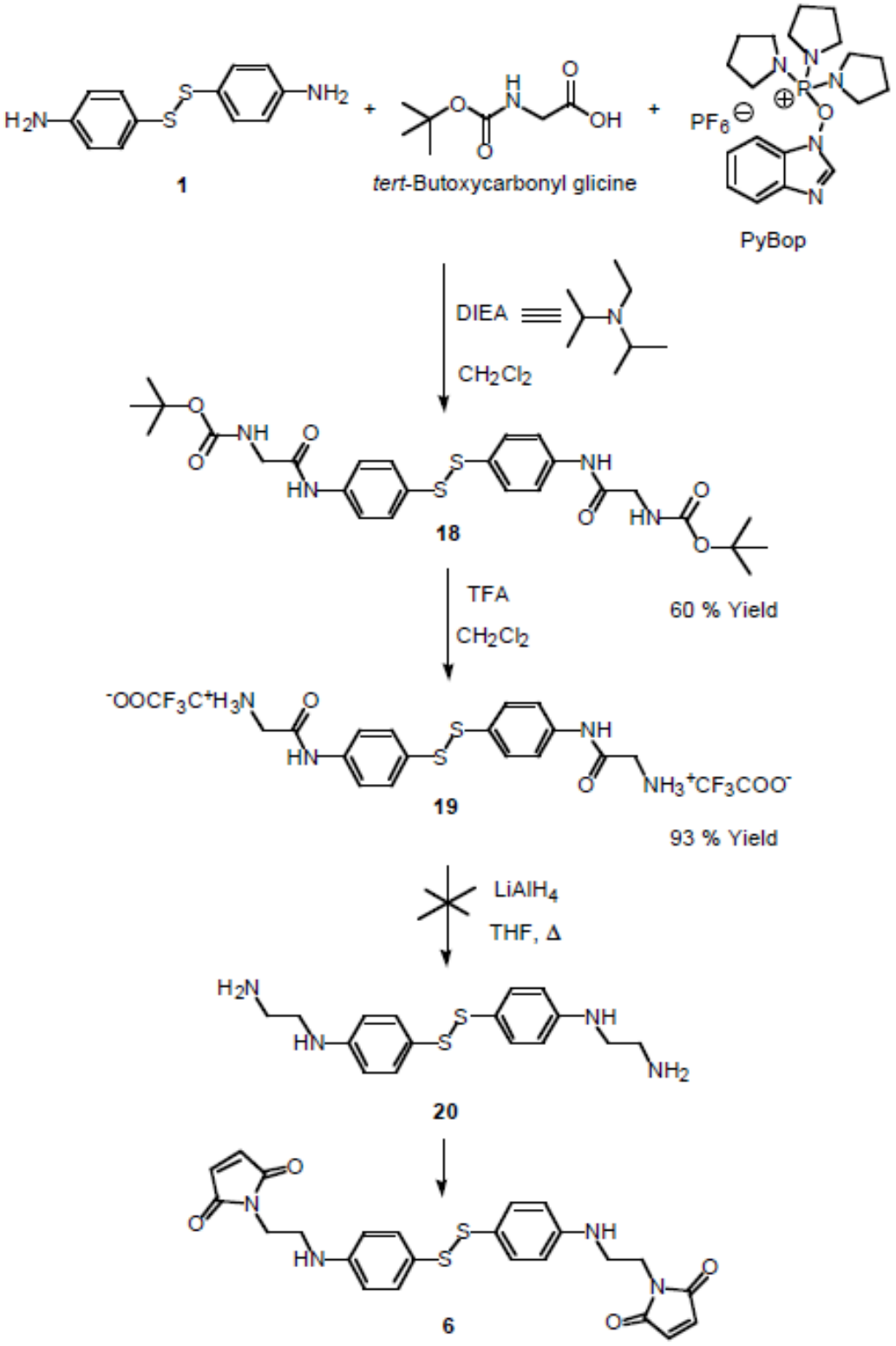
Figure 3.8. X-ray crystal structure of **5**.

by reaction of **1** with a tosylate derivative of N- (2-hydroxyethyl) maleimide failed to give the expected product probably due to the low nucleophilic character of the nitrogen of **1**. Instead, the use of **16,** which was synthesised following a literature procedure, afforded the expected product in 21 % yield (Scheme 3.3).[39] The crystal structure of **5** is shown in Figure 3.8.

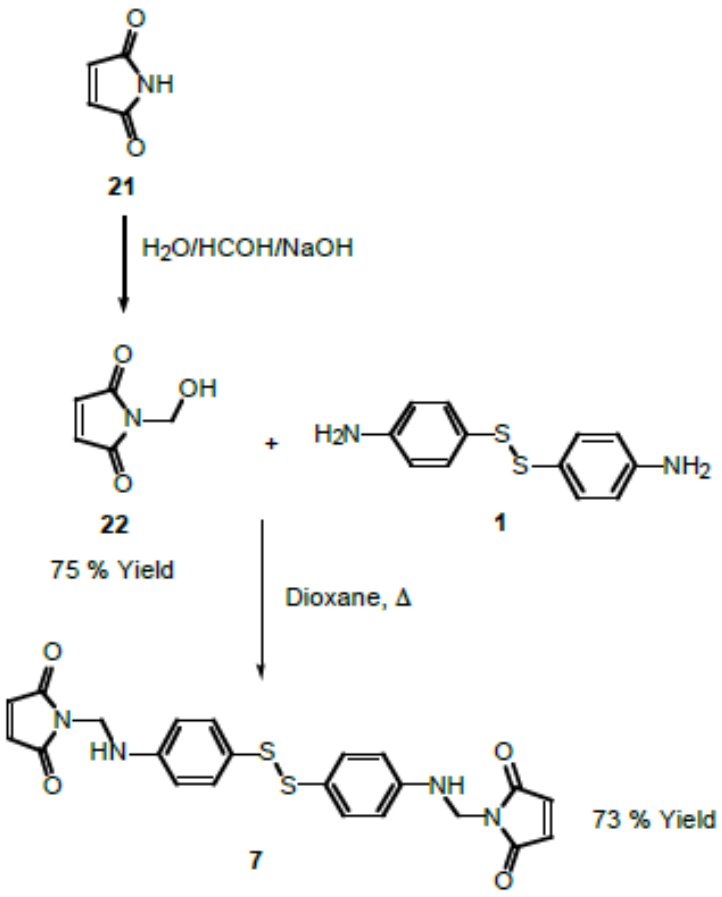


Scheme 3.2

The synthesis of **6** failed. One approach is depicted in Scheme 3.4. The strategy was to react **1** with the anhydride of a protected glycine to afford **18** followed by deprotection of the amino group and reduction of the amide **19** in order to obtain **20** which could be functionalised with a maleimide moiety. Synthesis of **18** and deprotection of the amino group were successful, but the reduction afforded **20** as mixture of products. Purification of **20** by column chromatography or extraction methods, was not successful.


1
tert-Butoxycarbonyl glicine
PyBop
DIEA
CH2Cl2
18
60 % Yield
TFA
CH2Cl2
19
93 % Yield
LiAlH4
THF, Δ
20
6

Scheme 3.4.



Scheme 3.5

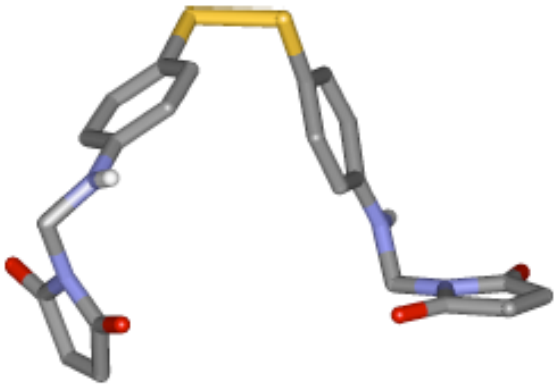


Figure 3.9. X- ray crystal structure of **7**.

Compound **7** was obtained in 60 % yield by reaction of **1** with **22** in refluxing dioxane (Scheme 3.5).[31] The crystal structure of **7** is shown in Figure 3.9.

A possible problem in the use of compound **7** may be its instability in mild acid conditions. Compound **7** was left overnight in solutions of tetrahydrofuran and water or acidified water, and the solutions were analysed by NMR spectroscopy after removing the water on a freeze - drier. Figure 3.10 shows that less than 10 % hydrolysis occurs when **7** is kept in a mixture THF/HCl 0.1 M 7:3 v/v, while no reactions occur when **7** is kept only in water/tetrahydrofuran. Since it is unlikely we will carry out protein functionalisation under strongly acidic conditions, compound **7** was considered to be stable enough to be used as an optical trigger for protein folding.

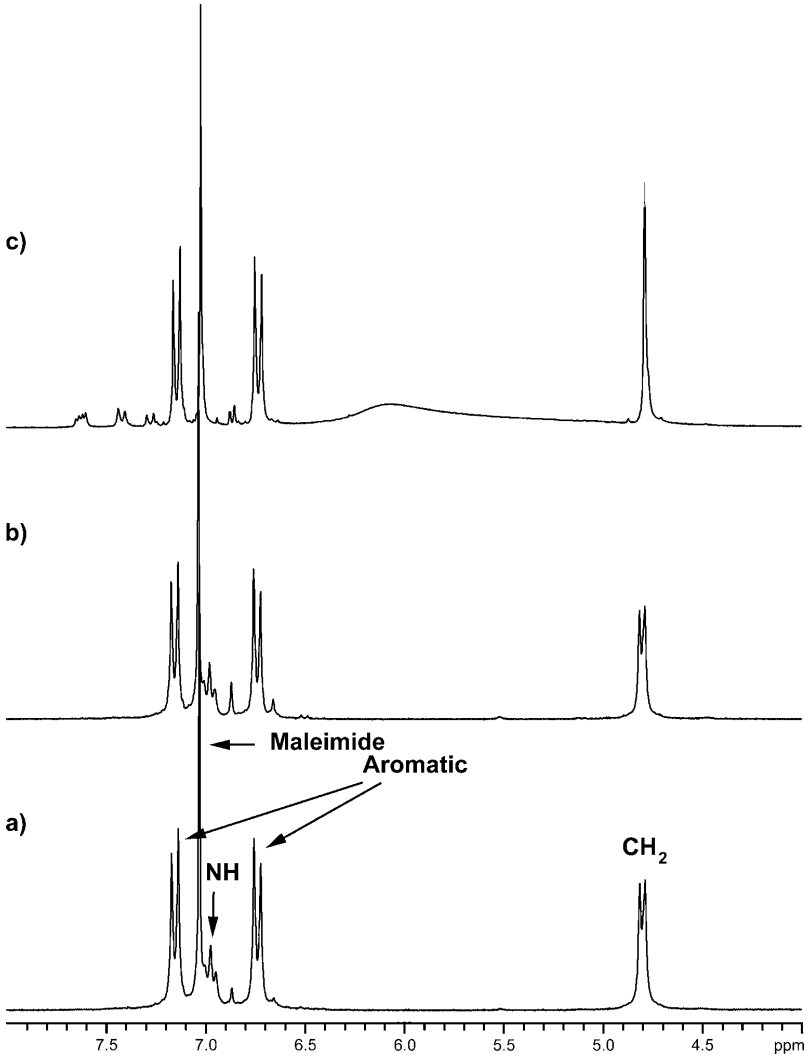
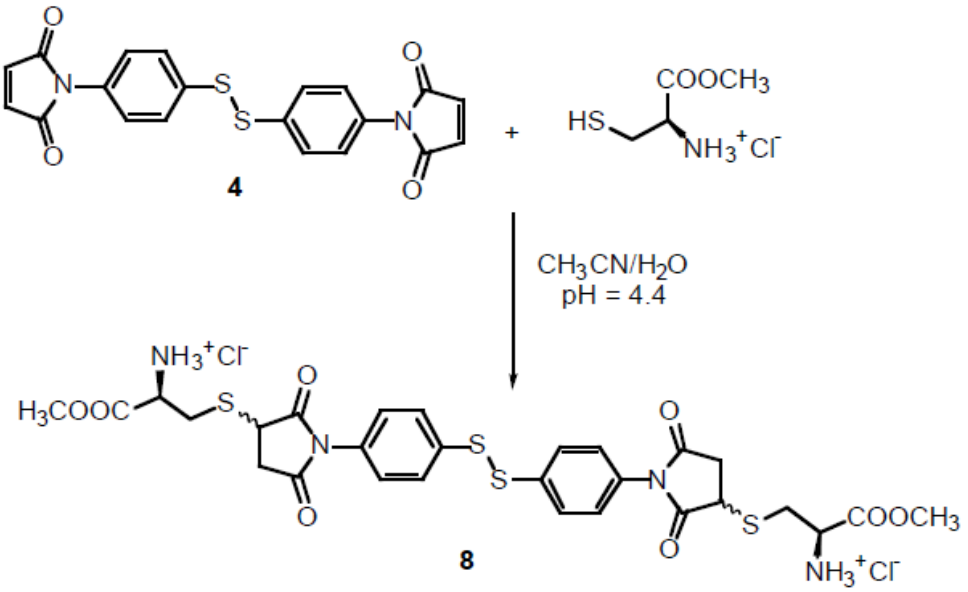


Figure 3.10. a) ^{1}H-NMR spectra of **7**. b) After 24 hours in THF/Water 6:4. c) After 24 hours in THF/HCl 0.1 M 7:3 v/v.



Scheme 3.6

The general method used for the reaction of the maleimide disulfides with cysteine was simply to mix cysteine methyl ester and the starting maleimide derivative in a solution of acetonitrile and water (Scheme 3.6).

Initial experiments were aimed to test the effect of pH and relative ratio of thiol to maleimide reagent on the extent of potential side reactions such as thiol-disulfide exchange. When the reagents were mixed in stoichiometric amounts the resulting pH was 4.4, due to the acidic properties of cysteine methyl ester hydrochloride. In these conditions, a mixture of compounds was formed as shown by the ^{1}H NMR spectrum in Figure 3.11.

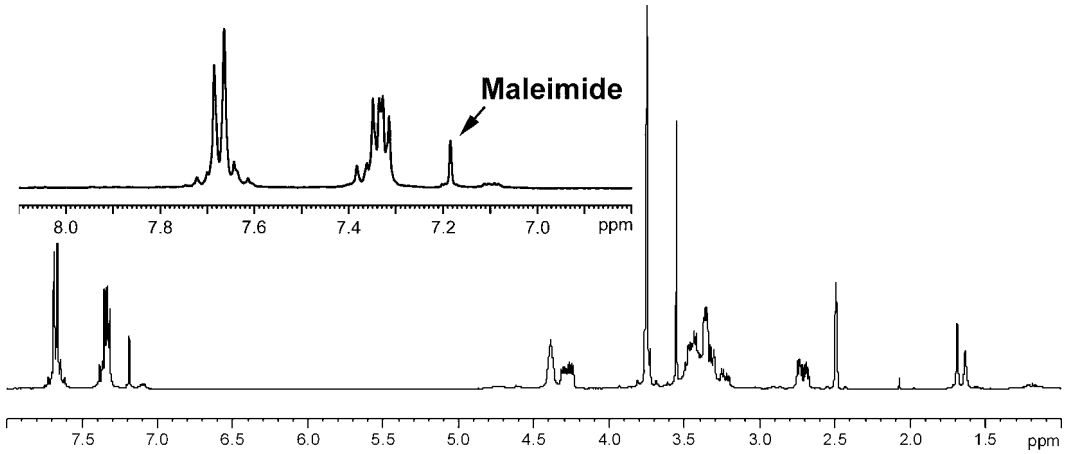
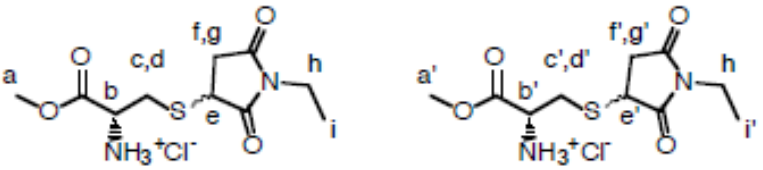


Figure 3.11. ^{1}H NMR spectra of the products of the reaction of disulfide **4** with stoichiometric cysteine methyl ester.

The presence of a maleimide signal in the spectrum, together with other minor signals in the aromatic region, suggests the formation of side products due to thiol-disulfide exchange. However the number of the aromatic and alkyl signals of the ^{1}H NMR spectrum could also be due to the formation of a mixture of three diastereoisomers resulting from the creation of a chiral center at the succinimide rings (Figure 3.12):



23, isomer A **23,** isomer B

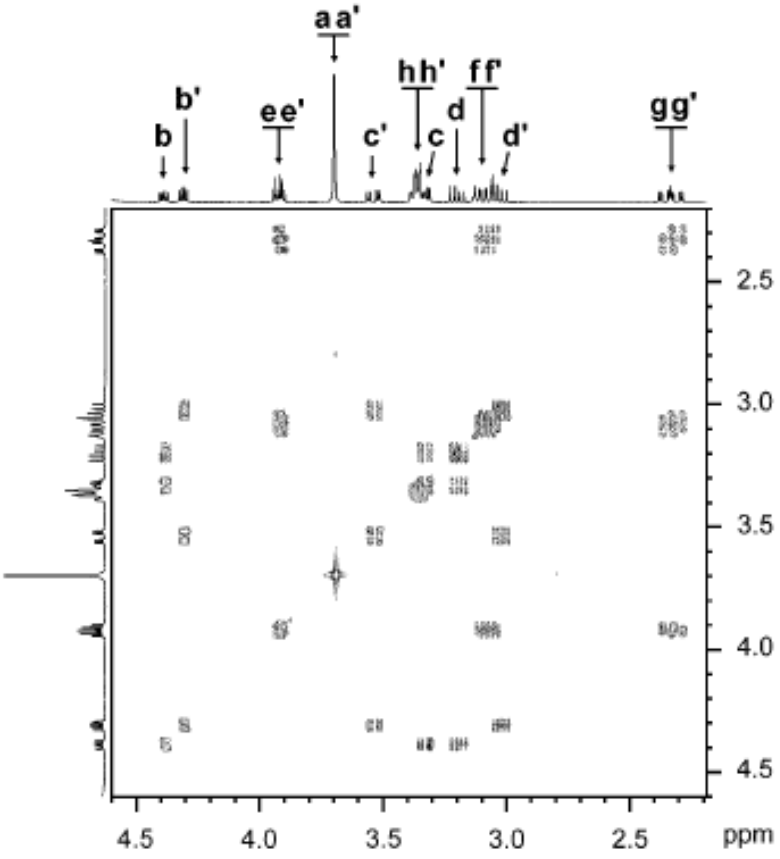


Figure 3.13. Section of a 2D NMR COSY spectrum of a solution 20 mM of **23** in CD_3OD.

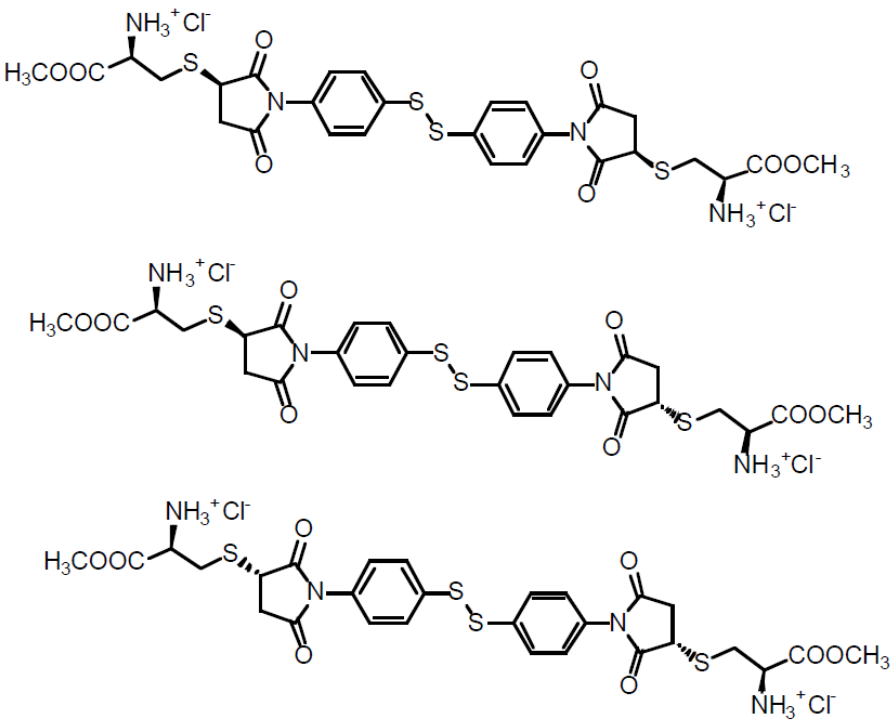


Figure 3.12. Possible stereoisomers of **8**.

In order to obtain information on the mixture of diastereoisomers **8**, a model compound **23** was synthesised by condensation of cysteine methyl ester with N-ethylmaleimide.[32] Since no disulfides are present, formation of **23** is not complicated by the thiol-disulfide exchange reaction. An assignment of the COSY spectrum shown in Figure 3.13 indicates that **23** is formed as a mixture 1:1, approximately, of two diastereosiomers. This result suggests that the minor signals observed in the aromatic area of the ^{1}H NMR spectrum of **8** are not due to the expected mixture of diastereoisomers, but to the presence of impurities that may come from competitive reactions. In fact, the mass spectrum of **8** showed peaks that correspond to the product of a thiol-disulfide exchange reaction. (Figure 3.14 and 3.15).

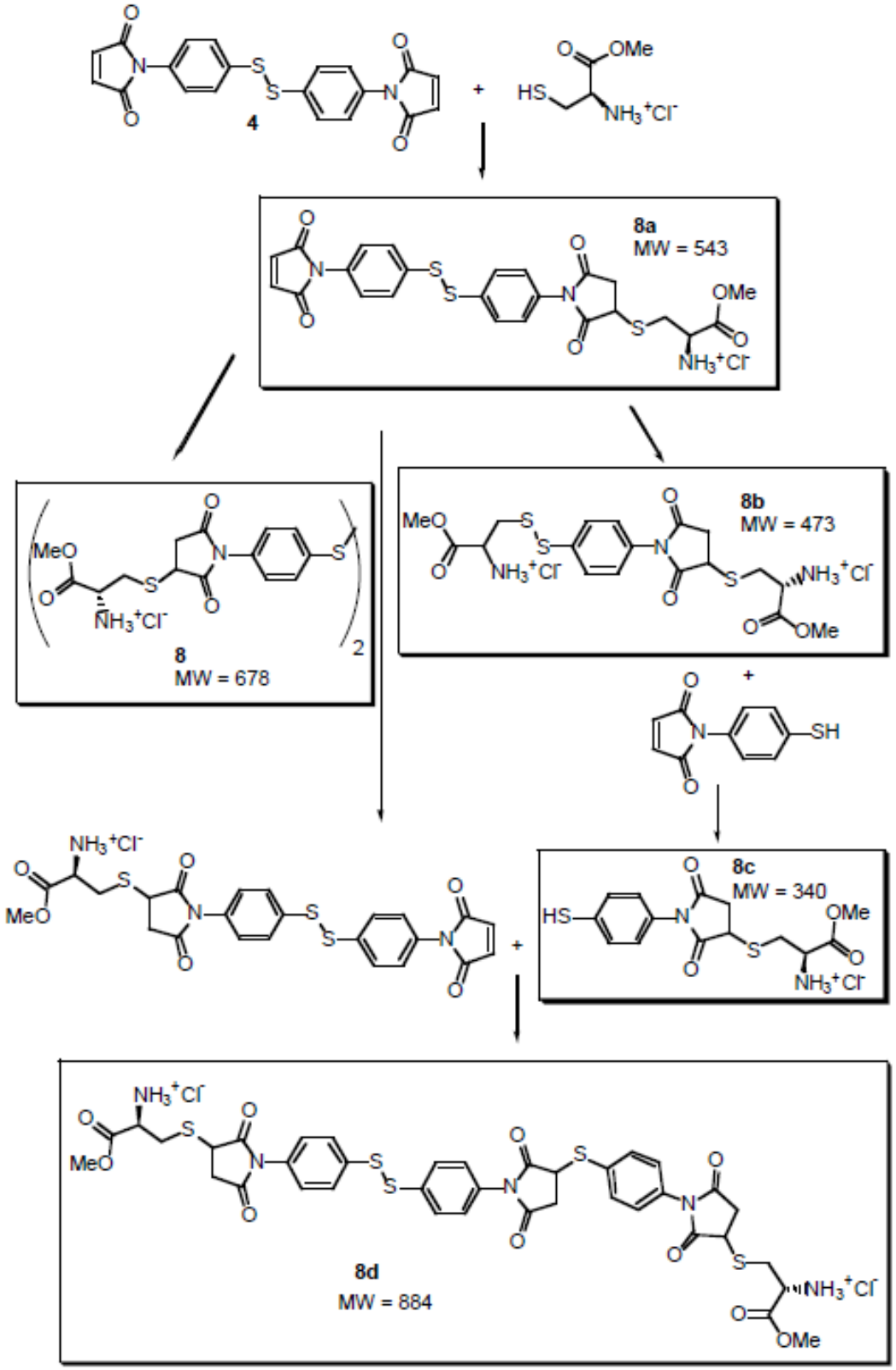


Figure 3.14. Possible mechanism for the reaction of **4** and a stoichiometric amount of cysteine methyl ester.

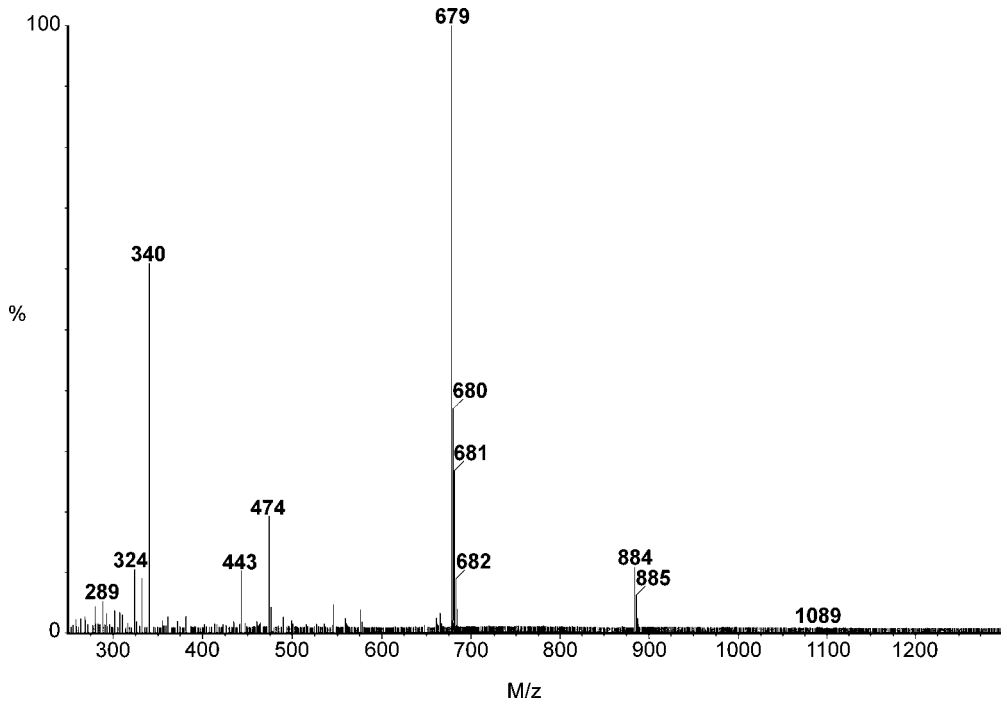


Figure 3.15. MS-ES$^+$ of the reaction mixture of **8** (MH^+ = 679) obtained by reaction of **4** (MW =408) with stoichiometric amount of cysteine methyl ester (MW = 135).

The major problem at this stage was to find out if the main product of the reaction in Scheme 3.6 was indeed **8**. To solve this problem, reaction of cysteine methyl ester with **4**, was carried with varying amounts of cysteine methyl ester. The NMR data, shown in Figure 3.16, indicated that when excess **4** is used, appear two sets of signals. A comparison with the ^{1}H NMR spectrum of the starting maleimide **4** suggests that one of the sets corresponds to unreacted maleimide moiety. Since **4** was removed from the sample before recording the spectrum, the signals were assigned to compounds **8** and **8a** (Figure 3.14). When a 2:1 mixture of cysteine methyl ester and starting disulfide is used new signals begin to emerge, but the maleimide singlet at 7.2 ppm does not disappear. This pattern suggests that some of the cysteine methyl ester is used in thiol-disulfide exchange reactions to give **8b**. When an excess of cysteine was used, the intensity of the signal assigned to the thiol-disulfide exchange product **8b** (Figure 3.14) increased dramatically, together with the appearance of new signals at 7.1 and 7.4 ppm. When this mixture was left standing for a few days and the NMR spectrum was recorded again, a change in the pattern of the signals is observed: the set of signals at 7.1 and 7.4 ppm disappeared and the intensity of the

doublets at 7.7 ppm strongly increased. Therefore we assigned the signal at 7.1 and 7.4 ppm to a product of the thiol-disulfide exchange reaction that may correspond to the structure **8c** (Figure 3.14). In presence of an excess of cysteine, thiol **8c** would eventually become mixed disulfide **8b** by oxidation.

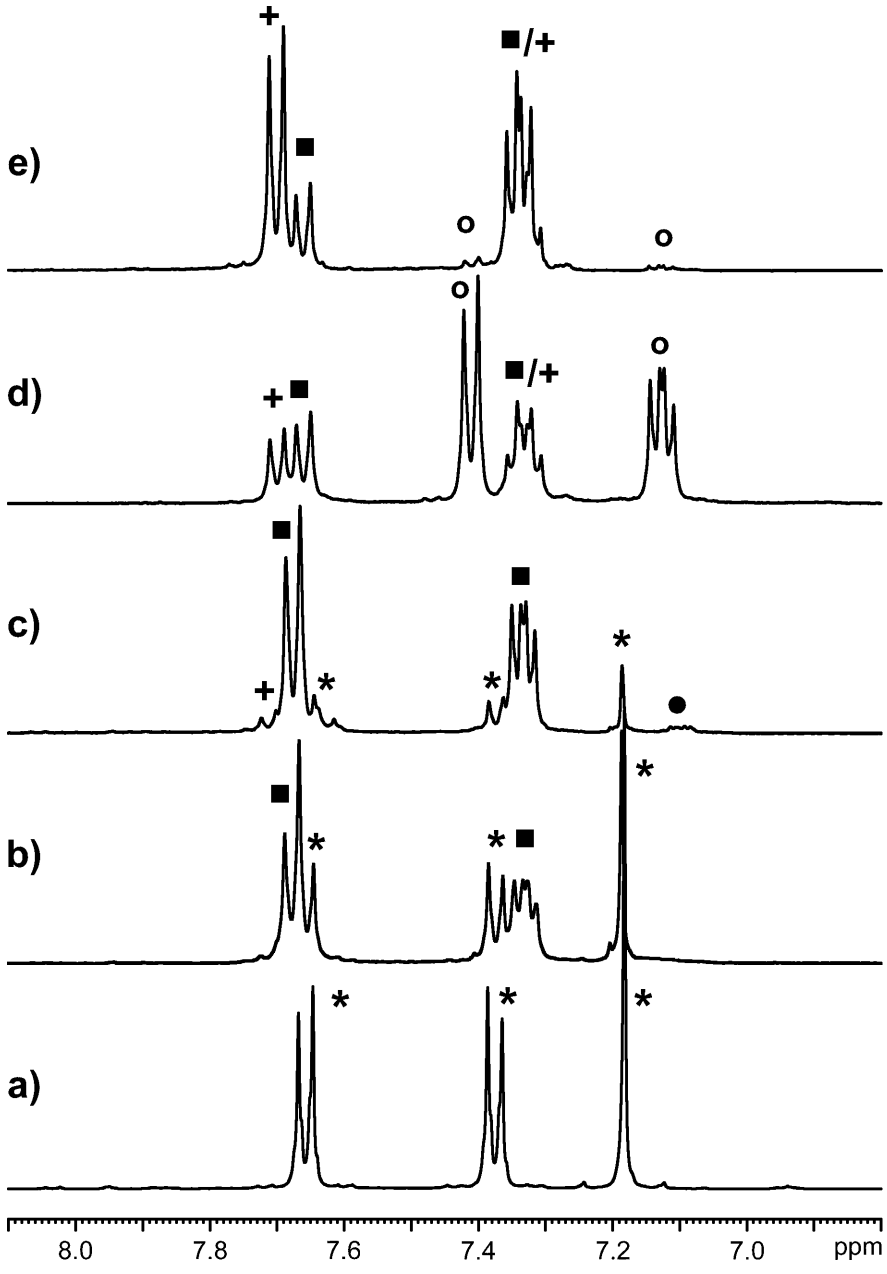


Figure 3.16. Aromatic region of the ^{1}H NMR spectra of disulfide **4** (a), the reaction of excess **4** with cysteine (b), stoichiometric cysteine (c) and excess cysteine (d) and excess cysteine after 7 days (e). Signals assigned to unreacted malemide (★), disulfide **8** (■), disulfide **8d** (●), thiol **8c** (**o**) and mixed disulfide **8b** (✚).

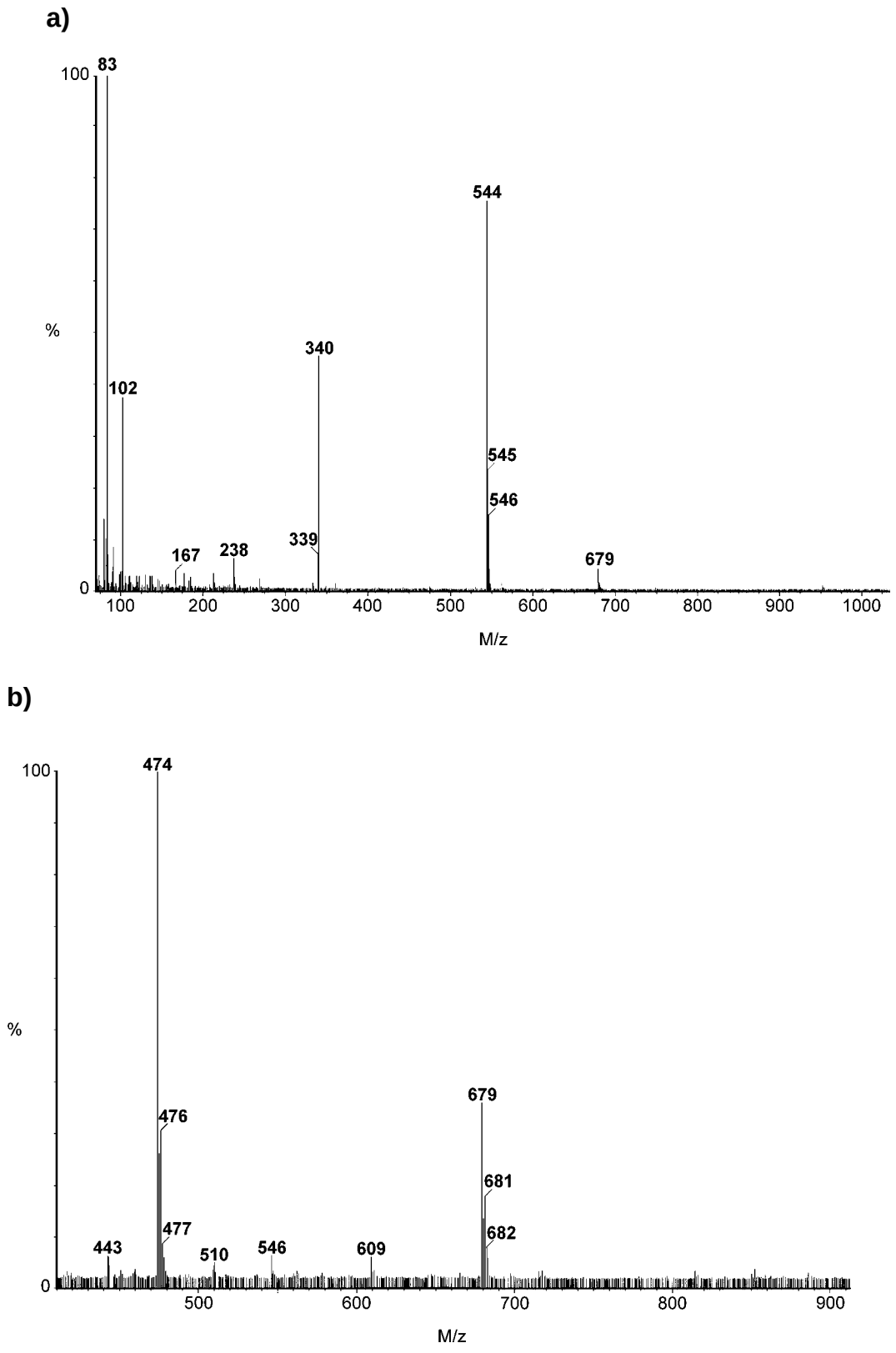


Figure 3.17. MS-ES$^+$ of the reaction of excess **4** with cysteine (a) and **4** with an excessof cysteine (b).

The interpretation of the ^{1}H NMR data is supported by mass spectrometry, (Figure 3.17). When the reaction is carried out in excess **4**, the thiol-disulfide exchange reaction is completely suppressed and only compounds **8** and **8a** are formed. In the presence of excess cysteine, the product from thiol-disulfide exchange, **8b** (Figure 3.14), is formed together with **8**. The mass spectra in Figure 3.15 and 3.17 show that high molecular weight products are formed only when cysteine methyl ester and maleimide **8** react in a 2:1 stoichiometry which suggests that the signals at 7.1 ppm in the NMR spectra of the corresponding product **8** may be assigned to the high molecular weight impurities. All these experiments suggest that **8** is the main product of the mixture obtained by reacting cysteine methyl ester and **4** in a 2:1 stoichiometry.

Since the presence of unreacted maleimide may complicate the interpretation of the photolysis experiments, we carried out the synthesis of **8** under conditions in which a slight excess of cysteine methyl ester was added. The ^{1}H NMR spectrum of the product is shown in Figure 3.18 and is in agreement with quantitative conversion of the starting maleimide **4** to the disuccinimide adduct **8.** This mixture was used without further purification for the photolysis experiments.

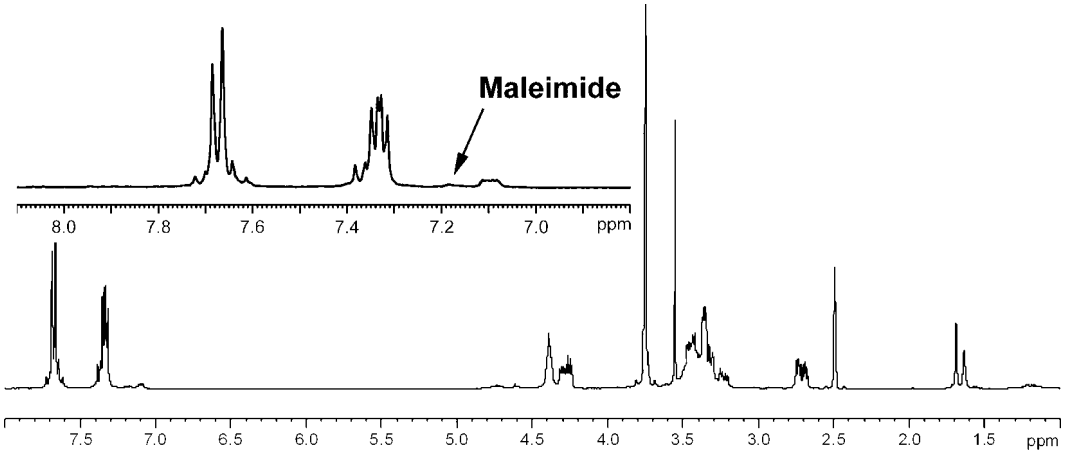
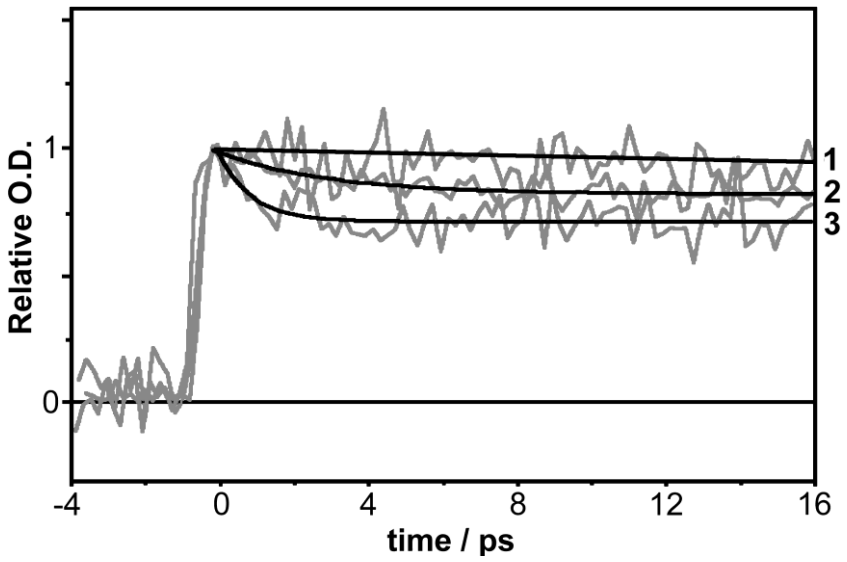


Figure 3.18. ^{1}H-NMR spectrum of the reaction of cysteine with disulfide **4** in a molar ratio 2.1:1.

Compounds **9** and **10** were obtained by reaction of two equivalents of cysteine methyl ester with **5** and **7** respectively. The ^{1}H NMR and MS spectra were in

a)



b)

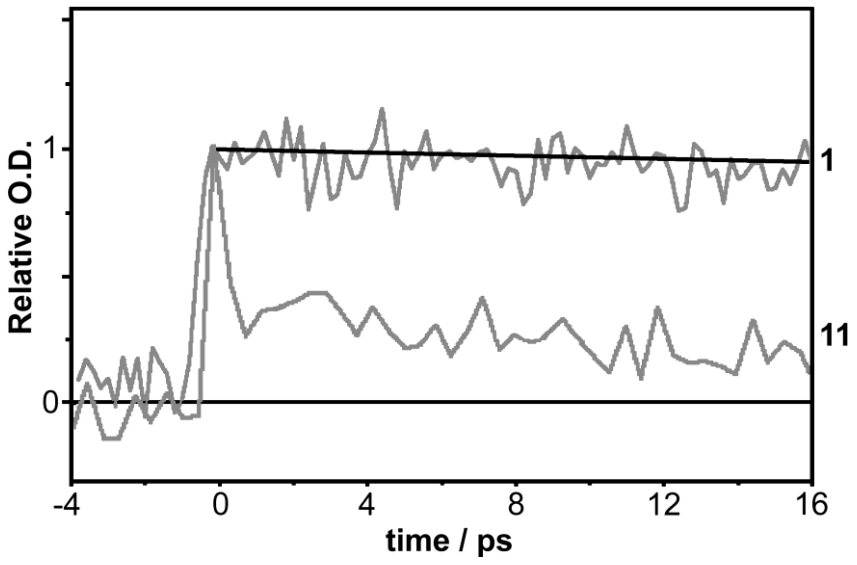


Figure 3.19. Evolution of the radical absorption signal at 550 nm after photolysis of **1, 2** and **3** (A) and **1, 11** (B) in acetonitrile. The fits to a monoexponential decay (Equation 3.5) are also shown for **1**, **2**, and**3**. The vertical axis is the optical density (OD) of the thiyl radical at 550nm relative to the OD of the ground state disulfide at the same wavelength.

agreement with the structure expected for **9** and **10** with a much lower amount of impurities due to the thiol-disulfide exchange (the ^{1}H NMR and MS spectra of **9** and **10** are shown in Appendix II). It is possible that in the case of **4** the thiol-disulfide exchange reaction competes with the addition of thiol to the maleimide because **4** provides a better leaving group than **5** or **7** due to differences in the electronic structure of the three disulfides.

Although we have shown that coupling of thiols with maleimide derivatives may be carried out under acid conditions, functionalisation of protein thiols is usually carried out under more basic conditions, because the reactivity of the thiolate is greater than that of the thiol and because proteins are more stable at a pH near to physiological values.[33, 34] Thus in order to test if the addition of thiols to maleimide derivatives **4**, **5**, and **7** may be successfully carried out under more realistic conditions, the reactions were repeated at pH 6. The extent of side product formation was not affected by the increase in the pH.

3.2.3 Photolysis

To study the photochemistry of the aryl disulfides a pump-probe experiment was performed using a laser apparatus that was set up to generate light pulses with a 100 fs width.[35] The wavelength of the pump and probe pulses were chosen using literature data: excitation at 260 nm corresponding to the energy necessary for disulfide bond dissociation (2.8 eV) and monitoring at 550 nm corresponding to the wavelength of the absorption maximum of the *p*-aminophenylthiyl radicals.[1, 2, 4, 7, 12, 13] Typically, each experiment was an average of 10-20 scans, to improve the signal to noise ratio. A sample of solvent was photolysed under the same experimental conditions used for the aromatic disulfides. No signal was found for the photolysis of the solvents probably due to the low intensity (~ 100 nJ) of the UV pulses. The time evolution of the radical absorption signal at 550 nm observed after photodissociation of disulfides **1**, **2**, **3** and **11** in acetonitrile is shown in Figure 3.19. In the case of bis (*o*-aminophenyl) disulfide the low extinction coefficient in acetonitrile and the limited solubility in acetonitrile gave a very low signal to noise compared to the other disulfides (see Appendix III).

In agreement with the literature, we found that the quantum yield of radicals formed by photodissociation of bis (*p*-aminophenyl) disulfide **1** in a polar solvent such as acetonitrile remained constant during the first picosecond after photodissociation.[4, 5, 9] In the case of **2**, **3** and **11,** a decrease in radical concentration was observed. This behaviour was found for all the disulfide derivatives under study, and it was quantified by fitting the experimental data to an exponential decay:

$$I(t) = Ae^{(-t/\tau)} + B \qquad \text{Equ. 3.5}$$

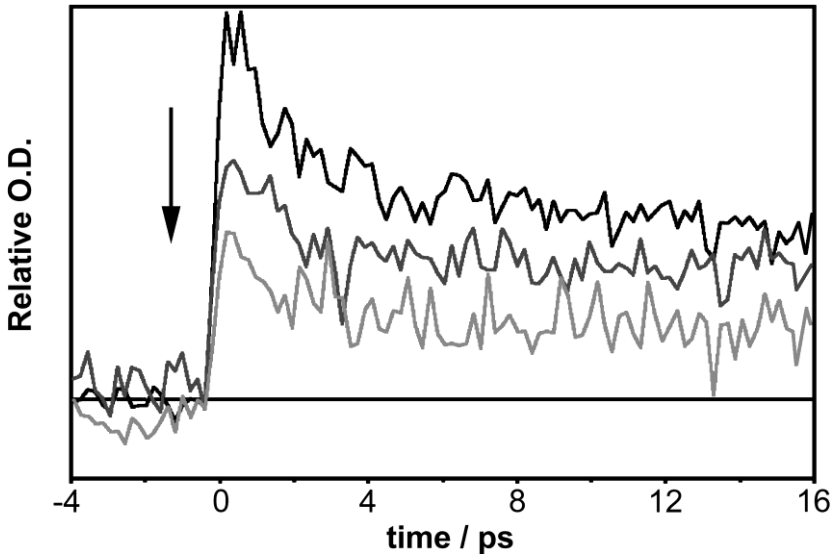
where I is the absorbance intensity, B is the absorbance intensity at t = ∞, (A + B) is the absorption at t = 0 and τ^{-1} is the observed rate constant for the decay in radical absorption. If the decay in radical population is due to geminate recombination, then τ^{-1} is the observed rate constant for geminate recombination.[6, 36] According to literature data the initial process of photodissociation (rate constant k_p in Figure 3.2) is complete in less than 100 fs for bis (*p*-aminophenyl) disulfide in polar solvents.[6, 8] Thus k_p is much faster than k_e and k_c, and the decay process may be described by $\tau^{-1} = (k_e + k_c)$. Then the fraction of radical absorption left after the initial fast decay, ϕ, is related to the kinetic parameters of the geminate recombination process according to Equation (3.6):

$$\phi = \frac{k_e}{k_c + k_e} \qquad \text{Equ. 3.6}$$

Since ϕ may be calculated from the experimental traces as the ratio of the absorbance at t = ∞ (B) and the absorbance at t = 0 (A+ B), k_c and k_e may be calculated.[36, 37] The parameters determined from the fitting process are listed in Table 3.1. Although analytical models more complex than Equation (3.5) may be used to obtain a detailed analysis of the transient absorption spectra, they do not explain the behaviour of *p*-aminothiyl radicals in polar solvents.[8] The theoretical and experimental work needed to modify these models is beyond the scope of the present work.

Photolysis of the maleimide derivatives was carried out in acetonitrile for **4** and **5** (Figure 3.21), and in tetrahydrofuran for **7** due to solubility problems (Figure 3.22).

a)



b)

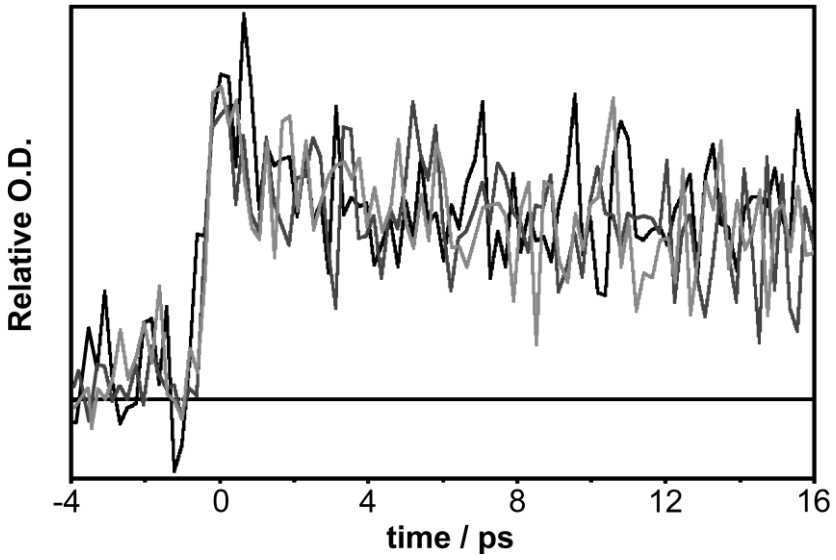


Figure 3.20. Successive photolysis traces acquired for disulfide **5** in acetonitrile at a concentration of 10^{-2} M (a) and 10^{-3} M (b).

In the case of **4** and **5,** we observed a decrease in the intensity of the signal after each scan as is shown in Figure 3.20 (a).

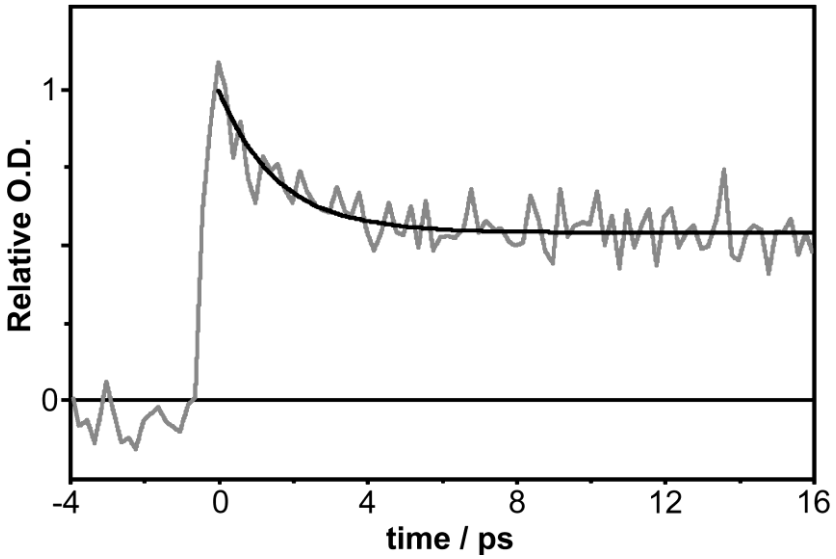


Figure 3.21. Evolution of the radical absorption signal at 550 nm after photolysis of disulfide **5** in acetonitrile. The fitting to a monoexponetial decay function is also shown.

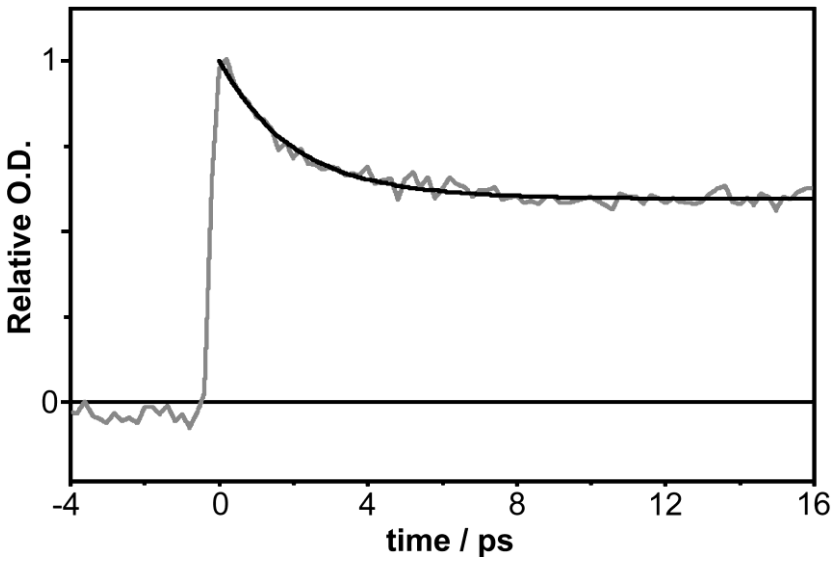


Figure 3.22. Evolution of the radical absorption signal at 550 nm after photolysis of disulfide **7** in THF. The fitting to a monoexponetial decay function is also shown.

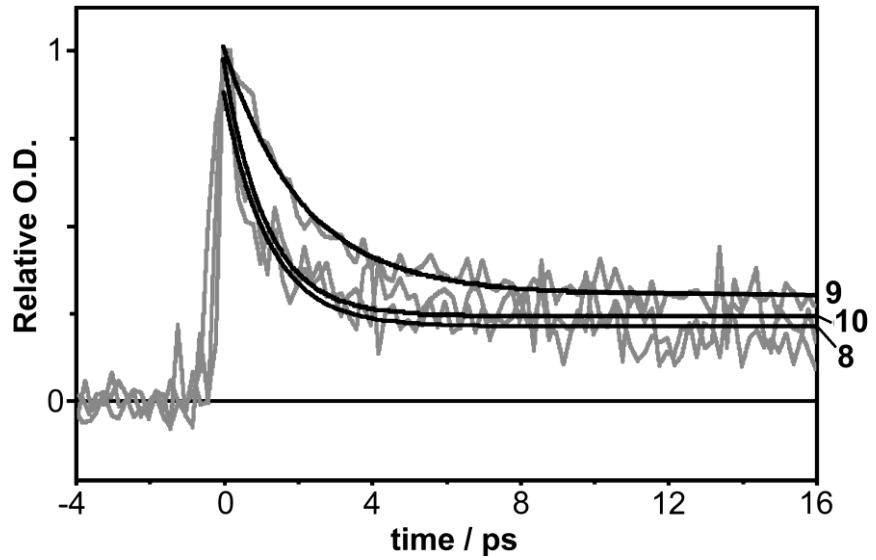


Figure 3.23. Evolution of the radical absorption signal at 550 nm after photolysis in water and the fitting to monoexponetial decay function for disulfides **8**, **9** and **10**.

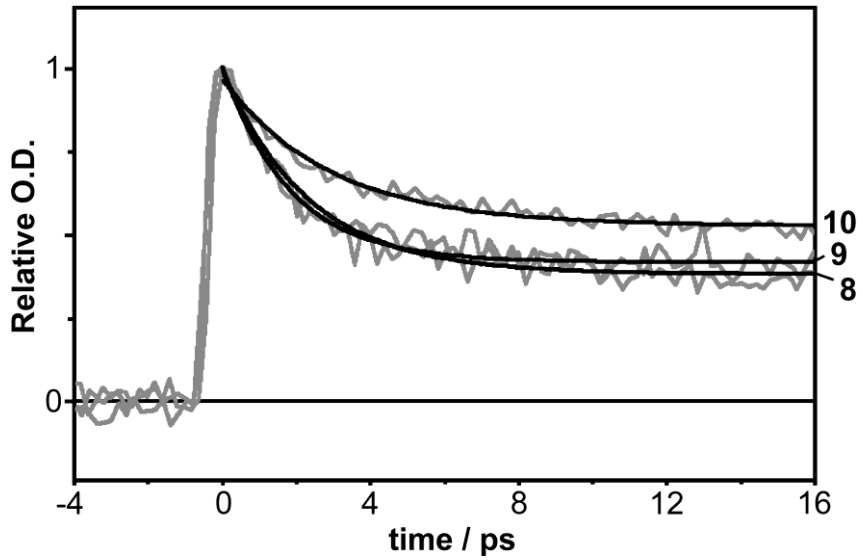


Figure 3.24. Evolution of the radical absorption signal at 550 nm after photolysis in methanol and the fitting to a monoexponetial decay function for disulfides **8**, **9** and **10**.

This behaviour suggests that the initial concentration of disulfide decreases with time. A possible explanation is that thiyl radicals generated after photolysis react with the double bond of the maleimide moiety. This hypothesis is supported by previous studies on the reactivity of aromatic thiyl radicals and by the fact that we didn't observe a decay in the concentration of the starting disulfide during photolysis of **2** and **3**.[12, 13] This effect was suppressed by reducing the concentration of the starting disulfide (Figure 3.20 (b)). However, in the case of **4**, the concentration of starting disulfide at which the side reaction could be suppressed proved to be too dilute for the observation of the transient absorption. This behaviour may be due to differences in the electronic structure of the maleimide derivatives.

The time evolution of the thiyl radical absorption observed after photolysis of water solutions of **8**, **9** and **10** is shown in Figure 3.23. We observed a 30% increase in the decay of thiyl radical absorbance in water compare to the decay of **5** and **7** in acetonitrile and tetrahydrofuran. In order to test if the solvation dynamic of thiyl radicals in water is different from that observed in acetonitrile, the photolysis of **8**, **9** and **10** should be carried out in acetonitrile. However, due to solubility problems, the experiments were carried out in methanol which has viscosity and dielectric properties similar to that of acetonitrile (Figure 3.24).[7, 8] Previous studies have shown that no geminate recombination is observed for *p*-aminophenylthiyl radicals in either solvent.[7, 8]

Table 3.1

Disulfide	Solvent (Viscosity)[a]	τ (ps)	k_e (ps^{-1})	k_r (ps^{-1})	A ($x10^5$)	B ($x10^5$)	ϕ^b
1	CH_3CN (0.38)	80	0.008	0.004	6.4	12.4	0.66 ± 0.33[c]
2	CH_3CN (0.38)	2.9	0.27	0.06	5.6	26.3	0.82 ± 0.25
3	CH_3CN (0.38)	0.9	0.76	0.35	5.6	11.7	0.68 ± 0.27
11[d]	CH_3CN (0.38)	-	-	-	11.1	3.7	0.25
5	CH_3CN (0.38)	1.6	0.33	0.28	10.8	12.5	0.54 ± 0.16
7	THF (0.55)	2.0	0.28	0.20	14.0	19.3	0.58 ± 0.06
8	MeOH (0.59)	2.3	0.15	0.21	17.8	9.9	0.36 ± 0.04
9	MeOH (0.59)	1.8	0.23	0.32	19.7	14.2	0.42 ± 0.05
10	MeOH (0.59)	2.9	0.19	0.40	8.5	10.6	0.56 ± 0.06
8	H_2O (1.00)	1.2	0.21	0.62	5.9	2.0	0.25 ± 0.05
9	H_2O (1.00)	2.1	0.14	0.32	14.1	6.1	0.30 ± 0.03
10	H_2O (1.00)	1.1	0.22	0.64	9.2	3.1	0.25 ± 0.05

[a]Viscosities are given in cP at 293 K and are taken from ref 38. [b] The error in the parameters from the fitting of the curves is taken as twice the standard deviation of the experimental data points calculated on the A value. [c] There is so little decay of the radical signal that the fitted parameters are very unreliable. The τ values show that this is a different process. [c] A and B were extimated by measuring the heights on the decay curve, the other data are not available.

3.3 Discussion

The behaviour of bis (*p*-aminophenyl) disulfide is in agreement with the literature, which attributes the absence of geminate recombination to the solvent stabilisation of polar radical form. Thus the increased decay observed for **2**, **3** and **11** might be the result of a less efficient solvation compared with *p*-aminophenylthiyl radicals.

In the case of **11'**, the faster decay of radical absorbance may be due the fact that in the corresponding polar form the charges are very close, so that the caged radical pairs may be kept in close proximity by the electrostatic interactions between the aniline nitrogen and the sulphur (Figure 3.25). This situation may better facilitate recombination, compared to the *p*-aminophenylthyil radical pair, which are kept distant by the electrostatic repulsion of the sulphur atoms when formation of the polar form occurs.

Another possibility is that the energy and distribution of excited states accessible to the thiyl radicals is different in the *ortho* isomer compared to the *para* isomer. According to MO theory, a change in the electronic charge distribution of the ground state disulfide may change the energy of the excited state radicals.[27, 39] This effect may account for the different stability of the *p* and *o*-aminophenylthiyl radicals. Indeed the batochromic shift and reduced intensity of the bis (*o*-aminophenyl) disulfide absorption spectrum compare with that of the bis (*p*-aminophenyl) disulfide (Appendix III) may be explained in terms of substituent effects on the energy of the excited state.[40, 41] A similar effect may occur after photolysis of the disulfides.

The fact that 50 % of the radicals recombine in the case of bis (*o*-aminophenyl) disulfide suggests that an approach based on *ortho* compounds is not promising.

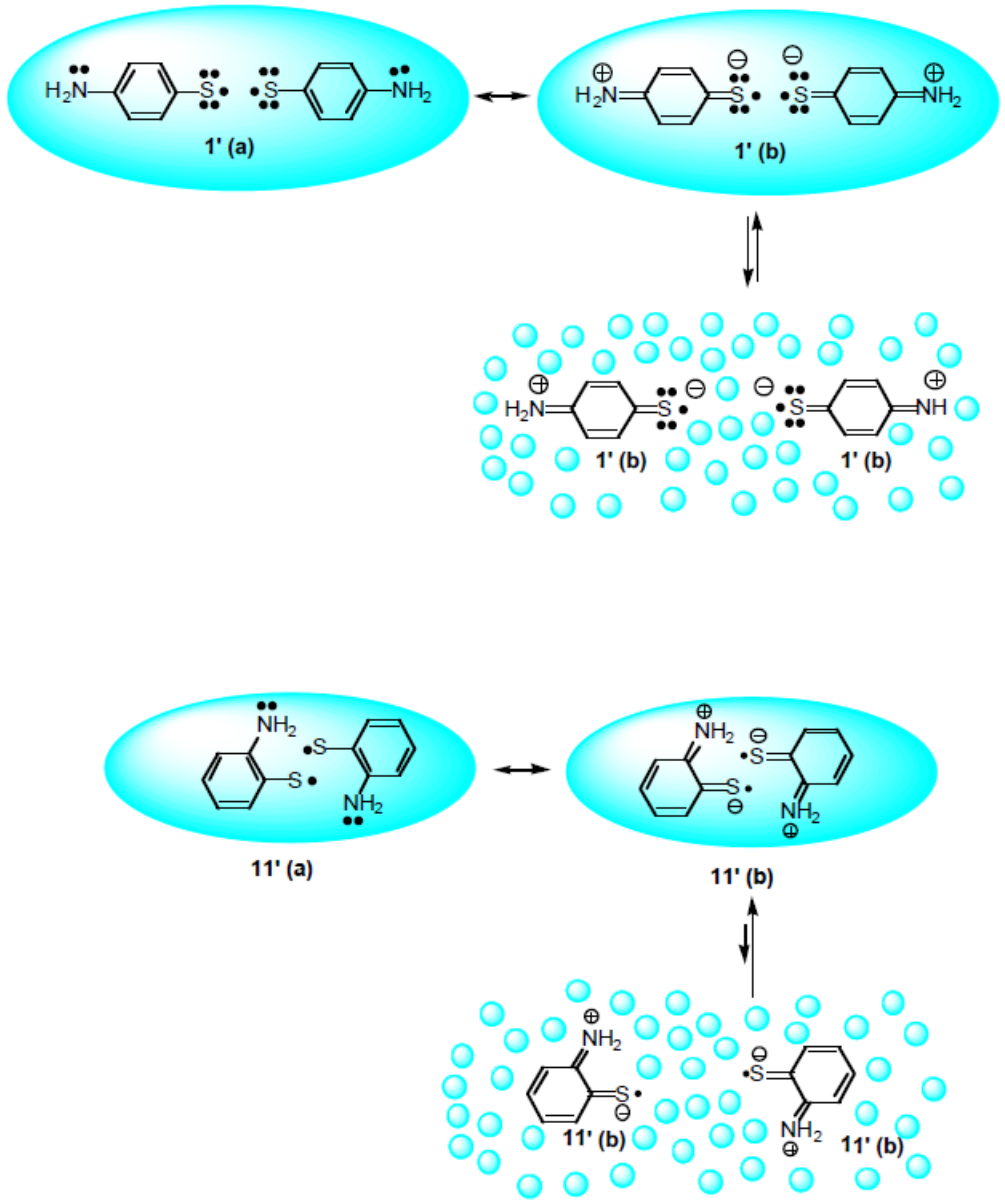


Figure 3.25 Two models for the formation of polar forms of the radicals pair generated after photolysis of *para* and *ortho* aminophenyl disulfide. Large ellipsoids indicate the solvent cage effect and small spheres indicate solvent molecules.

In the case of **1**, **2** and **3**, we observed that the decrease in radical population is proportional to the number of substituents on the aniline nitrogen. Solvent stabilisation of the radicals ionic form is possible for all the disulfides, but a possible

explanation of this behaviour is that radicals **1'** and **2'** may be more stable than radical **3'** due to direct hydrogen bonding interactions with the solvent (Figure 3.26).

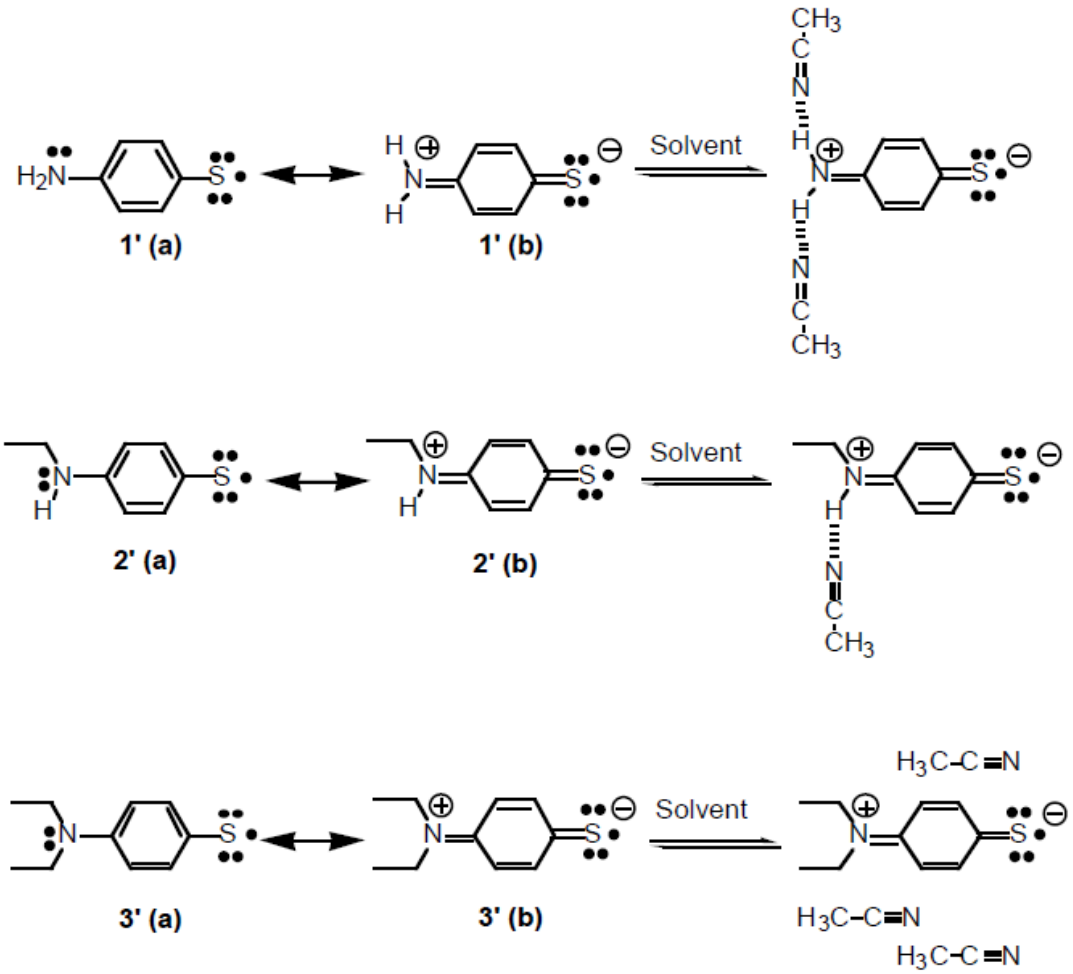


Figure 3.26. Model for the solvent stabilisation of radicals **1'**, **2'**, and **3'**: hydrogen bonding interactions with the solvent are shown for polar forms of **1'** and **2'**.

This additional solvent interaction may reduce the diffusional mobility of **2'** explaining the lower k_e observed for radical **2'** compared to radical **3'**. In addition, the timescale of solvent relaxation around thiyl radicals **2'**, **3'** and **1'** may be different because of differences in the dipole moments of the three radicals. An increase of the polar character of the thiyl radical is expected for **3'(b)** compared to **1' (b)** and **2' (b)** due to an increase in the number of electron donor substituents on the aniline nitrogen. The fact that a high percentage (60-80%) of radicals are left after photolysis of **2** and **3** shows that the functionalisation of the aniline nitrogen reduces, but does not alter significantly, the stability of the corresponding thiyl radicals.

Photolysis of **5** and **7** gives a yield of radical of about 50 %. Interpretation of the photolysis data of the maleimide derivatives is complicated by the reactivity of the thiyl radicals towards the double bond of the maleimide ring and by the insolubility of **7** in acetonitrile. However, a comparison of the photolysis data of radical **5** and **3** suggests that the increased decay observed for the maleimide derivatives may be due to differences in the rate of cage escape rather than the rate of recombination, since the k_c values for both the radicals is similar and thus they have similar stability. The value of k_e suggests that radical **5** escapes the solvent cage more slowly than radical **3**. This behaviour may be explained in terms of reduced diffusional mobility due to the increase in the size of radical **5** compared to radical **3**. Indeed according to Equation 3.3, the efficiency of geminate recombination increases with an increase in the size of the radical. The effect of size as driving force in the extent of radical recombination has also been reported in the literature.[19, 20, 37] Since the data for **7** were recorded in tetrahydrofuran, a comparison with the decay traces of the other disulfides is not possible. A complete study requires that the transient absorptions of all the disulfides are recorded in tetrahydrofuran, but the main focus of the present work is to test if the thiyl radicals are stable enough in polar solvents to be used as optical triggers for protein folding.

The cysteine derivatives in methanol showed a slight decrease in yield and k_e values relative to the maleimide derivatives. This behaviour could be attributed to the increase in the mass of the compounds or to a stronger interaction with the solvent due to the presence of the cysteine moiety. The trends are similar to the alkyl derivatives, and compound **10** gives the most stable radical. In water, the extent of decay of the radicals increases still further, and difference between the compounds can not be resolved due to the noise in the experimental traces. The increase in radical decay observed in water compared to methanol may be due to a even stronger interaction with the solvent and to the fact that the viscosity of water is nearly twice that of methanol (1.00 cP and 0.59 cP respectively).[38] According to the Noyes model, the efficiency of cage recombination is directly proportional to solvent viscosity.[8, 19]

The observation that nearly 30 % of radicals are left after photolysis of the cysteine derivatives in water suggests that a similar behaviour may be expected when working

with the protein, and thus the maleimide derivatives may be used to trigger fast folding reactions.

3.4 Conclusions

A new series of water soluble maleimide derivatives have been synthesised and their transient absorption spectra have been recorded using a pump-probe technique with a time resolution of 100 femtoseconds. The data obtained show that, although the stability of the radicals decreases, the photochemical properties of the maleimide reagents do not differ substantially from those of simple aromatic disulfide. Thus the approach based on functionalisation of aromatic disulfide for the preparation of new optical trigger for fast protein folding reactions is promising.

3.5 Experimental section

3.5.1 Materials

Chemicals were obtained from commercial sources and were used without further purification. HPLC grade solvents were used for UV analyses. TLC was carried out using Merck Silica Gel 60 (F_{254}) pre-coated aluminium sheets, and visualisation was under UV light or exposure to Iodine vapours. Flash chromatography was performed using BDH or MercK silica gel 60 (40-63 μm). Dichloromethane was dried over calcium hydride and distilled under nitrogen. Tetrahydrofuran (THF) was dried over sodium, benzophenone and distilled under nitrogen. Dioxane was refluxed overnight over lithium aluminium hydride to remove peroxides and then it was dried as described for THF. The solutions were deoxygenated using the following method: The sample container, provided with a magnetic stirrer was subjected to a partial vacuum that was broken with nitrogen. This process was repeated at least 5 times.

3.5.2 Instrumentation

^{1}H and ^{13}C NMR spectra were recorded on either a Bruker AC250 or AMX400 spectrometer. The 2D COSY spectrum was recorded on a Bruker AMX-400 spectrometer. All chemical shifts are quoted in ppm and are referenced to the solvent signal. The following abbreviations are used in the assignment of ^{1}H spectra: s singlet, d doublet, dd doublet of doublets, t triplet, q quintet , m multiplet, br broad signal. Positive fast atom bombardment (FAB^{+}) mass spectra and high-resolution mass spectra (HRMS/FAB^{+}) were recorded on a Fison VG ProSpec 8 using a matrix of *m*-nitrobenzyl alcohol (NBA) and a magnetic sector detection system. Positive electrospray (ES^{+}) mass spectra were recorded on a Fison VG Platform using a quadrupole detection system, samples were injected as a solution in acetonitrile/water or THF/water mixtures. UV/Visible spectra were recorded on a Varian Cary 3 Bio spectrophotometer. Melting points were recorded on a Reichter Kofler hot stage melting point apparatus and are not corrected. Microanalyses were performed using a Perkin–Elmer 2400 CHN elemental analyser operating at 975 ^{o}C.

3.5.3 Photolysis experiments

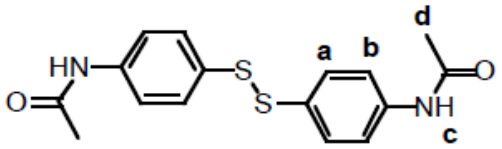
Method

The disulfides were dissolved in HPLC grade solvents (Merck). Concentrations of the samples were 1-2 mM for the measurements in THF, acetonitrile and methanol. For the measurements in water sample concentration was 0.8 – 1 mM. The cuvette path length was 1 mm for all the measurements. (UV spectra of all disulfides are shown in Appendix III).

Laser apparatus

The femtosecond laser and transient absorption spectrophotometer have been built and set up by Dr. Gavin Reid and Prof. Godfrey Beddard at the University of Leeds. A detailed description of this laser system is reported elsewhere.[3] For the present experiment the laser set up was modified as follows:
the pump beam was 520 nm and it was frequency doubled using a 100 micro-metre thick BBO crystal to 260 nm. Samples were excited with 100 fs of the 260 nm light (the energy was a maximum of 100 nJ). The zero time for the probe wavelength was determined by excitation of a solution of *trans* stilbene in acetonitrile. For each experiment the difference in absorbance between the excited and ground state species measured at different pump/probe delay times was less than 10^{-4}. The signal to noise ratio was improved by averaging 10-20 scans. The data were analysed with the same program used for data collection (ExptPPC1, version 3.0), which was kindly provided by Dr. Gavin Reid.

3.5.4 Synthesis of 12

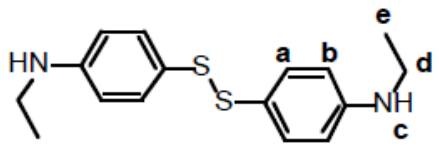


12

To a solution of 2.06 g (20 mmol) acetic anhydride in 150 ml toluene were added, dropwise, 2.48g (10 mmol) 4-aminophenyldisulfide dissolved in 20 ml of toluene. The solution was then stirred for an hour under reflux while the product slowly precipitated from the reaction mixture. After cooling, the amide **12** was collected by filtration (2.98 g, yield: 90%).

^{1}H-NMR (DMSO-d_6): 10.1 (s, 2H, **c**), 7.59 (d, 4H, **a**, J = 8.52), 7.42 (d, 4H, **b**, J = 8.55), (s, 3H, **d**).^{13} C-NMR (DMSO-d_6): 169.0, 140.0, 130.6, 129.8, 120.1, 24.5. MS (ES^+.) m/z: 333 ($M+H^+$), 355 ($M+Na^+$). $C_{16}H_{16}N_2O_2S_2$ requires 332.44. m.p.: 218-220^{o} C.

3.5.5 Synthesis of 2



2

A solution of **12** (2.37 g, 7.1 mmol) in 100 mL of dry THF was added dropwise to23.5 mL (21.4 mmol) of 1M lithium aluminium hydride solution in THF at 0 ^{o}C under nitrogen. The resulting solution was allowed to warm to room temperature and then it was heated under reflux overnight. It was then cooled to 5 ^{o} C and water was added until no more gas was produced. The mixture was filtered and the solid was washed with 100 mL of CH_2Cl_2. The organic phase was dried with anhydrous Na_2SO_4 and the solvent removed under reduced pressure. The resulting oil was purified by column chromatography using a mixture of CH_2Cl_2/hexane 95:5 v/v as eluent. Disulfide **2** (RF = 0.6) was obtained as a yellow oil that solidified slowly (1 g, yield: 46 %).

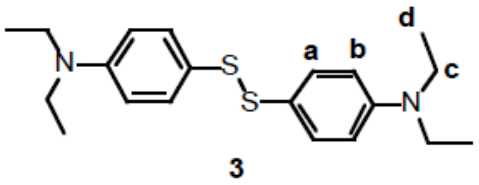
^{1}H-NMR (DMSO-d_6): 7.13 (d, 4H**, a**, J = 8.55), 6.5 (d, 4H, **b**, J = 8.85), 6.01 (t, 2H, **c**, J = 5.17), 3.04 (m, 4H, **d**), 1.16 (t, 6H, **e**, J = 7.3). ^{13} C-NMR ($CDCl_3$): 149.0, 134.6, 124.0,112.9, 38.3, 14.8 ppm. MS (ES^+) m/z: 152, 305 ($M+H^+$), $C_{16}H_{20}N_2S_2$ requires 304.48. CHN calculated for $C_{16}H_{20}N_2S_2$: C, 63.70; H, 7.55; N, 8.74; S, 20.01. Found: C, 63.46; H, 7.44; N, 8.61; S, 20.10. UV/Vis. (CH_3CN): ε_{260} = 16656.

*Crystal data for **2** ($C_{16}H_{20}N_2S_2$):* M = 304.46. Crystallises from ethanol as yellow blocks; crystal dimensions 0.16 x 0.16 x 0.05 mm^3. Monoclinic, a = 16.0946(12), b = 14.9471(12), c = 14.6262(12) Å, β = 115.7380(10)°, U = 3169.5(4) Å^3, Z = 8, D_c = 1.276 Mg/m^3, space group P2$_1$/c ($C^{5,2}_{h}$, No.14), Mo-Kα radiation (λ, $^{-}$ = 0.71073 Å), μ(Mo-Kα) = 0.328 mm^{-1}, F(000) = 1296. Data collected were measured on a Bruker Smart CCD area detector with Oxford Cryosystems low temperature system. Cell parameters were refined from the setting angles of 54 reflections (θ range 1.40 < 28.29°). Reflections were measured from a hemisphere of data collected of frames each covering 0.3 degrees in omega. Of the 20877 reflections measured, all of which were corrected for Lorentz and polarisation effects and for absorption by semi empirical methods based on symmetry-equivalent and repeated reflections (minimum and maximum transmission coefficients 0.9494 and 0.9838), 4695 independent reflections exceeded the significance level |F|/σ(|F|) > 4.0. The structure was solved by direct methods and refined by full matrix least squares methods on F^2. Hydrogen atoms were placed geometrically and refined with a riding model (including torsional freedom for methyl groups) and with U_{iso} constrained to be 1.2 (1.5 for methyl groups) times U_{eq} of the carrier atom. Refinement converged at a final R = 0.0628 (wR$_2$ = 0.1429, for all 7602 data, 361 parameters, mean and maximum δ/σ 0.000, 0.000) with allowance for the thermal anisotropy of all non-hydrogen atoms. Minimum and maximum final electron density -0.472 and 0.524 e.Å^{-3}. A weighting scheme w = 1/[σ^2(Fo^2) +(0.0572∗P)^2+0.00∗P] where P=(Fo^2+ 2 ∗ Fc^2)/3 was used in the latter stages of refinement. Complex scattering factors were taken from the program package SHELXTL^Y as implemented on the Viglen Pentium computer

Supplementary material (see Appendix I)

anisotropic thermal vibrational parameters with e.s.d.s
hydrogen atom position parameters
observed structure amplitudes and calculated structure factors

3.5.6 Synthesis of 3



3

To a solution of 0.3 g (1 mmol) of **2** in 10 ml of dry dicholorometane were added, dropwise, 0.5 ml (4 mmol) of trifluoro-ethanesulfonic acid ethyl ester under nitrogen, and the resulting solution was stirred overnight. The solvent was then removed under reduced pressure and the residue was suspended in 30 mL of saturated $NaHCO_3$ solution and washed with CH_2Cl_2 (3 x 30 mL). The organic phase was then dried over Na_2SO_4, and the solvent removed under reduced pressure. The residue was purified by a column chromatography using a mixture of hexane/CH_2Cl_2 70:30 v/v as eluent. The first fraction eluted was an impurity corresponding to trisufide **3a** (78 mg, yield: 20 %). Afterwards, compound **3** was obtained as a yellow solid (72 mg, yield: 20 %).

^{1}H-NMR ($CDCl_3$): 7.33 (d, 4H, **a**, J = 8.6), 6.58 (d, 4H, **b**, J = 8.8), 3.36 (q, 4H, **c**, J =7.0), 1.17 (t, 6H, **d**, J =7.0). ^{13} C-NMR ($CDCl_3$): 141.2, 130.1, 121.9, 113.3, 48.7, 13.3.HRMS (FAB$^+$) m/z: 361.177924 (M+H$^+$), $C_{20}H_{29}N_2S_2$ requires 361.1777218. UV/Vis. (CH_3CN): ε_{260}= 11990.

*Crystal data for **3a** ($C_{20}H_{28}N_2S_3$)*: M = 392.62; crystallises from dichloromethane as yellow blocks; crystal dimensions 0.32 x 0.18 x 0.06 mm^3. Monoclinic, *a* = 10.542(2), *b* = 10.2720(19), *c* = 19.133(3) Å, β = 94.801(4)°, *U* = 2064.5(7) Å^3, *Z* = 4, D_C = 1.263 Mg/m^3, space group P2$_1$/n (a non-standard setting of P2$_1$/c $C^{5,2}$ *h*, No.14), Mo-K$_\alpha$ radiation (λ,$^-$ = 0.71073 Å), μ(Mo-K$_\alpha$) = 0.365 mm^{-1}, F(000) = 840. Data collected were measured on a Bruker Smart CCD area detector with Oxford Cryosystems low temperature system. Cell parameters were refined from the

setting angles of 43 reflections (θ range 2.13 to 28.28°). Reflections were measured from a hemisphere of data collected of frames each covering 0.3 degrees in omega. Of the 12843 reflections measured, all of which were corrected for Lorentz and polarisation effects and for absorption by semi empirical methods based on symmetry-equivalent and repeated reflections (minimum and maximum transmission coefficients 0.8922 and 0.9784) 2708 independent reflections exceeded the significance level $|F|/\sigma(|F|) > 4.0$. The structure was solved by direct methods and refined by full matrix least squares methods on F^2. Hydrogen atoms were placed geometrically and refined with a riding model and with U_{iso} constrained to be 1.2 times U_{eq} of the carrier atom. Refinement converged at a final R = 0.0711 (wR_2 = 0.1958, for all 4984 data, 226 parameters, mean and maximum δ/σ 0.000, 0.000) with allowance for the thermal anisotropy of all non-hydrogen atoms. Minimum and maximum final electron density -0.378 and 0.630 e.$Å^{-3}$. A weighting scheme $w = 1/[\sigma^2(Fo^2) + (0.1025*P)^2 + 0.00*P]$ where $P=(Fo^2 + 2 * Fc^2)/3$ was used in the latter stages of refinement. Complex scattering factors were taken from the program package SHELXTL[Y] as implemented on the Viglen Pentium computer.

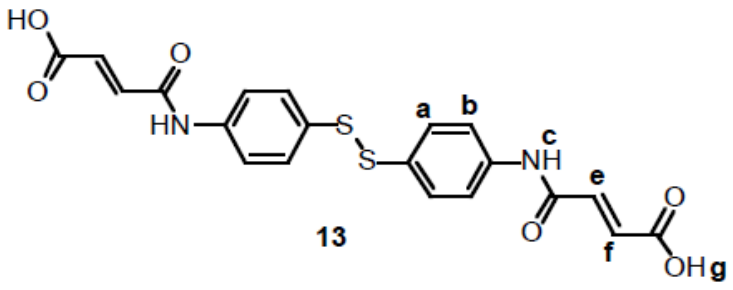
Supplementary material (see Appendix I)

anisotropic thermal vibrational parameters with e.s.d.s

hydrogen atom position parameters

observed structure amplitudes and calculated structure factors

3.5.7 Synthesis of 13

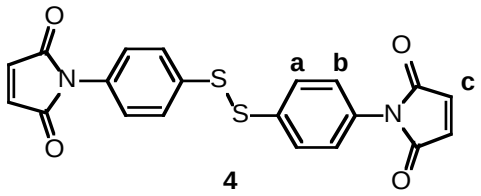


13

The procedure described for the synthesis of **12** was repeated using 2.48 g (10 mmol) of **1**, and 1.98 g (20 mmol) of maleic anhydride in 200 mL of toluene.

After one hour at reflux the product slowly precipitated from the reaction mixture. After cooling, theamide **13** was collected by filtration (4.00 g, yield: 90 %).
^{1}H-NMR (DMSO-d_6): 13.1 (s, 2H, **g**), 10. 50 (s, 2H, **c**), 7.65 (d, 4H, **a**, J = 8.8), 7. 50 (d, 4H, **b**, J = 6.7), 6.50 (d, 2H, **f**, J = 11.9), 6.34 (d, 2H, **e**, J = 11.9). ^{13} C-NMR (DMSO-d_6): 167.4, 163.8, 139.2, 130.7, 130.4, 132.1, 120.7. MS (ES^+) m/z: 112, 445 ($M+H^+$). $C_{20}H_{16}N_2O_8S_2$ requires 444. m.p. 191.0-193.0 ^{o}C.

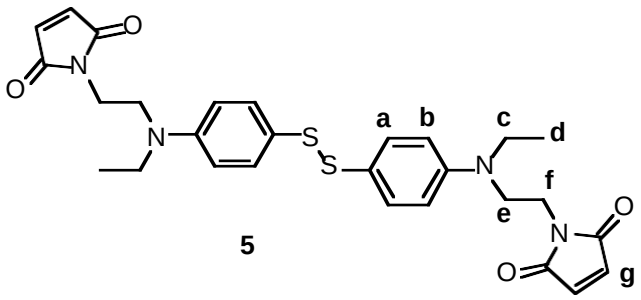
3.5.8 Synthesis of 4



4

7 mL of acetic anhydride and 0.65 g (8 mmol.) of anhydrous sodium acetate were added to 4.14 g (9.3 mmol) of **13** and the mixture was gently heated until complete dissolution of **13** (30 min). After cooling the solution was poured onto ice water (50 mL) and the product was extracted with CH_2Cl_2 (3 x 50 ml). The organic phase was washed with water (3 x 50 mL) and dried over Na_2SO_4 anhydrous. After removing the solvent under reduced pressure the product was recrystallized from acetone/cyclohexane to give a cream coloured solid (2.3 g, yield: 56%).
^{1}H-NMR ($CDCl_3$): 7.60 (d, 4H, **a**, J = 8.8), 7.33 (d, 4H, **b**, J= 8.8), 6.86 (s, 4H, **c**). ^{13}C-NMR ($CDCl_3$): 169.2, 136.4, 134.3, 130.4, 128.0, 126.5. MS (FAB $^{+}$) m/z: 204, 409 ($M+H^+$). $C_{20}H_{12}N_2O_4S_2$ requires 408. UV/Vis (CH_3CN): ε_{260}= 19613. CHN calculated for $C_{20}H_{12}N_2O_4S_2$: C, 58.81; H, 2.96; N, 6.86; S, 15.70. Found: C, 58.46; H, 2.84; N, 6.56; S, 15.90. m.p. 188-190 ^{o}C.

3.5.9 Synthesis of 5



5

1.16g (4.25 mmol) of **16** dissolved in 10 mL of CH_2Cl_2 were added dropwise to a solution of 0.634 g (2.08 mmol) of **2** in 25 ml of dry CH_2Cl_2 under nitrogen. The resulting solution was stirred for 48 hours at room temperature. The solvent was then removed under reduced pressure, and the residue was purified by flash chromatography, using a gradient mixture of solvent as eluent, ranging from CH_2Cl_2 /hexane 90:10 v/v, to eliminate the first impurities, pure CH_2Cl_2 and finishing with CH_2Cl_2 /ethylacetate 95:5 v/v to obtain the desired product **5** as a yellow solid (0.24 g, 21% yield).

^{1}H-NMR ($CDCl_3$): 7.31 (d, 4H, **b**, J = 7), 6.63 (d, 4H, **a**, J = 7), 6. 63 (s, 4H, **g**), 3.68 (t, 4H, **f**, J = 7.62), 3.46 (t, 4H, **e**, J = 6.7), 3.36 (q, 4H, **c**, J = 8.77), 1.15 (t, 6H, **d**, J = 7). ^{13} C-NMR ($CDCl_3$): 170.6, 147.8, 134.2, 134.1, 123.2, 112.2, 47.5, 44.5, 34.8, 12.2. HRMS (ES^+) m/z: 551.1807 ($M+H^+$). $C_{28}H_{31}N_4O_4S_2$ requires 551.1787. m.p. 120-121 ^{o}C. UV/Vis. (CH_3CN): ε_{260}= 15021.

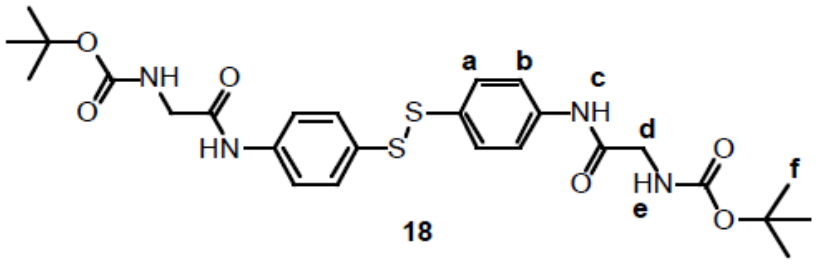
Crystal data for ***5*** *($C_{28}H_{30}N_4O_4S_2$)*: M = 550.68. Crystallises from methanol/dichloromethane as yellow plates; crystal dimensions 0.26 x 0.15 x 0.05 mm^3. Monoclinic, *a* = 25.908(3), *b* = 8.2753(12), *c* = 14.3757(19) Å, β = 117.428(2)°, *U* = 2735.7(6) $Å^3$, *Z* = 4, D_C = 1.337 Mg/m^3, space group C2/c , Mo-K_α radiation (λ, = 0.71073 Å), μ(Mo-K_α) = 0.236 mm^{-1}, F(000) = 1160. Data collected were measured on a Bruker Smart CCD area detector with Oxford Cryosystems low temperature system. Cell parameters were refined from the setting angles of 25 reflections (θ range 1.77 < 28.32°). Reflections were measured from a hemisphere of data collected of frames each covering 0.3 degrees in omega. Of the 8455 reflections measured, all of which were corrected for Lorentz and polarisation

effects and for absorption by semi empirical methods based on symmetry-equivalent and repeated reflections (minimum and maximum transmission coefficients 0.9412 and 0.9883) 2068 independent reflections exceeded the significance level $|F|/\sigma(|F|) > 4.0$. The structure was solved by direct methods and refined by full matrix least squares methods on F^2. Hydrogen atoms were placed geometrically and refined with a riding model (including torsional freedom for methyl groups) and with U_{iso} constrained to be 1.2 (1.5 for methyl groups) times U_{eq} of the carrier atom. Refinement converged at a final R = 0.0586 (wR_2 = 0.1308, for all 3247 data, 172 parameters, mean and maximum δ/σ 0.000, 0.000) with allowance for the thermal anisotropy of all non-hydrogen atoms. Minimum and maximum final electron density -0.616 and 0.338 e.Å^{-3}. A weighting scheme $w = 1/[\sigma^2(Fo^2) + (0.0612*P)^2 + 0.00*P]$ where $P=(Fo^2 + 2 * Fc^2)/3$ was used in the latter stages of refinement. Complex scattering factors were taken from the program package SHELXTL^Y as implemented on the Viglen Pentium computer.

Supplementary material (see Appendix I)

anisotropic thermal vibrational parameters with e.s.d.s
hydrogen atom position parameters
observed structure amplitudes and calculated structure factors

3.5.10 Synthesis of 18

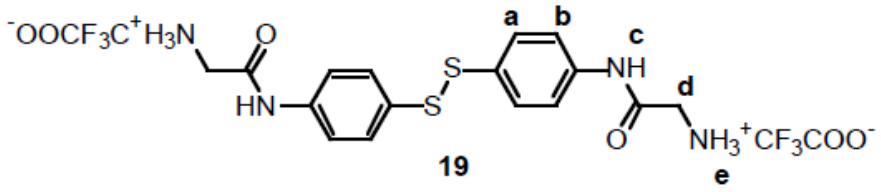


18

To a solution of **1** (1.5 g, 6 mmol) in CH_2Cl_2 (30 mL) were added (*tert*–Butyloxycarbonyl)–L-glycine (2.12 g, 12 mmol) and PyBOP (6.2 g, 12 mmol) followed by 2.1 ml of diisopropylethyl amine. The reaction mixture was stirred for 16 hours at room temperature. The solvent was then removed under reduced pressure and the residue was dissolved in ethyl acetate (200 mL) and washed with a saturated

$NaHCO_3$ solution (2 x 50 mL), water (2 x 50 mL), a 10 % citric acid solution solution (2 x 50 mL) and water (2 x 50 mL). The organic layer was dried over anhydrous Na_2SO_4 and then the solvent removed under reduced pressure. The residue was purified by column chromatography using a mixture of CH_2Cl_2/methanol 98:2 v/v as eluent. Compound **18** was obtained as a white solid (2.18 g, yield: 60 %).
^{1}H-NMR (DMSO-d_6): 10.11 (s, 2H, **c**), 7.44 (d, 4H, **b**, J = 8.05), 7.63 (d, 4H, **a**, J = 8.55), 7.08 (d, 2H, **e**, J = 10. 01), 3.72 (d, 4H, **d** J = 5.80), 1.4 (s, 18H, **f**). ^{13}C-NMR (DMSO-d_6): 172.1, 167.2, 140.9, 134.7, 132.6, 124.3, 80.5, 44.6, 27.4. MS (FAB$^+$) m/z: = 462, 563 ($M+H^+$), $C_{26}H_{34}N_4O_8S_2$ requires 562.19.

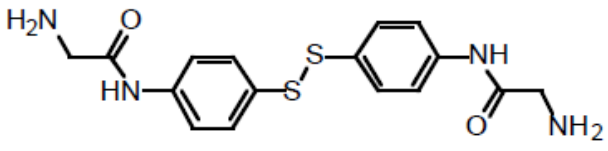
3.5.11 Synthesis of 19



19

A solution of **18** (1g, 1.77 mmol) in 5 mL of $CH_2Cl_2$was added dropwise to 5 mL of TFA at 0^o C. The solution was allowed to warm to room temperature for 30 minutes and then the solvent was removed under vacuum for 8 hours to provide crude **19** (0.980 g, yield: 93 %). Compound **19** was used without further purification for the following reaction.
^{1}H-NMR (DMSO-d_6): 10.69 (s, 2H, **c**), 8.25 (s, 6H, **e**), 7.62 (d, 4H, **a**, J = 8.55), 7.52 (d, 4H, **b**, J = 8.85), 3.80 (s, 4H, **d**). ^{13}C-NMR (DMSO-d_6): 165.4, 159.4, 158.9, 138.7, 130.8, 130.3, 120.5, 119.6, 114.9, 41.4. MS (ES$^+$) m/z: 182, 363 (MH^+). $C_{16}H_{18}N_4O_2S_2$ requires 362.09.

3.5.12 Procedure for the preparation of 20

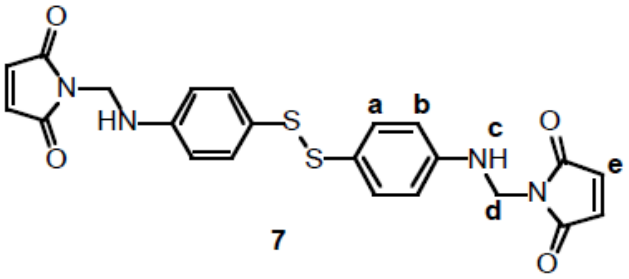


20

5 mL of 1M $LiAlH_4$ solution in THF (4.5 mmol) were added dropwise to a solution of 0.9 g of **19** (1.5 mmol) in 250 mL of dry THF at 0 °C under nitrogen. The resulting solution was allowed to warm to room temperature and then it was heated under reflux for 16 hours. It was then cooled to 5 °C and water was added until no more gas was formed. The organic phase was dried over anhydrous Na_2SO_4 and the solvent removed under reduced pressure to give a brown solid (100 mg). Attempts to purify **20** by column chromatography or extraction methods failed.

^{1}H-NMR (DMSO-d_6): 7.15 (m, broad), 6.55 (m, broad), 3.75 (m, broad), 3.08 (m, broad), 2.72 (m, broad), 1.75 (s). MS (ES^+) m/z: = 119, 124, 150, 167, 183, 225, 335 ($M+H^+$), 397. $C_{16}H_{22}N_4S_2$ requires 334.50.

3.5.13 Synthesis of 7



7

0.59 g (2.34 mmol) of **1** and 0.6 g (4.68 mmol) of **22** were dissolved in 10 mL of dry dioxane. The solution was stirred under reflux for 4 hours. After cooling, the solvent was removed by lyophilisation. The residue was washed with acetonitrile, filtered off, and dried under vacuum (0.8 g, yield: 73 %). Solutions to test for hydrolysis were prepared as follow: 10 mg of **7** (0.02 mmol) were dissolved in a mixture of THF/water 6:4 v/v, or in a mixture of THF/HCl 0.1M 7:3 v/v. The samples were

incubated for 24 hours and then the solvent was removed by lyophilisation. The residue was dissolved in DMSO-d_6 for the ^{1}H-NMR analysis (Figure 3.10).
^{1}H-NMR (DMSO-d_6): 7.17 (b, 4H, **b**, J = 8.55), 7.05 (s, 4H, **e**), 6.99 (t, 2H, **c**, J = 6.7), 6.76 (d, 4H, **a**, J = 8.55), 4.82 (d, 4H, **d**, J = 6.7). ^{13}C-NMR (DMSO-d_6): 171.2, 147.3, 134.8, 133.6, 122.9, 113.0, 46.2. HRMS (FAB$^+$) m/z: 466.078253 (M^+),

$C_{22}H_{18}N_4O_4S_2$ requires 466.076949. UV/Vis. (THF): ε_{260}= 18762.

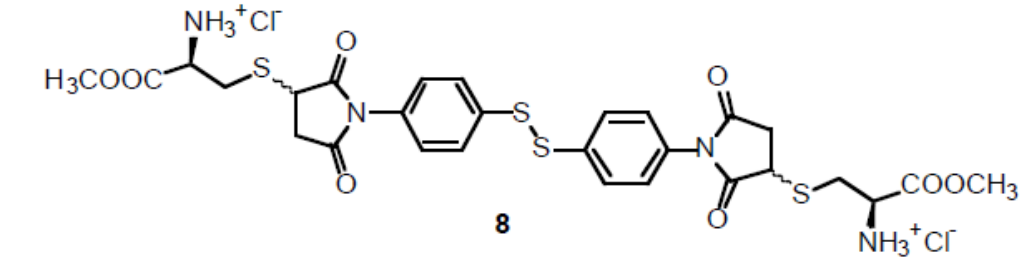
Crystal data for ***7****($C_{22}H_{18}N_4O_4S_2$):* M = 466.52; crystallises from THF/cyclohexane as yellow needles; crystal dimensions 0.12 x 0.06 x 0.03 mm^3. Monoclinic, a = 6.210(4), b = 23.380(15), c = 14.011(9) Å, β = 90.000(14)°, U = 2034(2) Å^3, Z = 4, D_c = 1.523 Mg/m^3, space group $P2_1/c$ ($C^{5,2}_h$, No.14), Mo-Kα radiation (λ, $^-$ = 0.71073 Å), μ(Mo-Kα) = 0.302 mm^{-1}, F(000) = 968. Data collected were measured on a Bruker Smart CCD area detector with Oxford Cryosystems low temperature system. Cell parameters were refined from the setting angles of 15 reflections (θ range 1.69 to 28.38°). Reflections were measured from a hemisphere of data collected of frames each covering 0.3 degrees in omega. Of the 10256 reflections measured, all of which were corrected for Lorentz and polarisation effects and for absorption by semi empirical methods based on symmetry-equivalent and repeated reflections (minimum and maximum transmission coefficients 0.9646 and 0.9910) 1412 independent reflections exceeded the significance level $|F|/\sigma(|F|) > 4.0$. The structure was solved by direct methods and refined by full matrix least squares methods on F^2. Hydrogen atoms were placed geometrically and refined with a riding model and with U_{iso} constrained to be 1.2 times U_{eq} of the carrier atom. Refinement converged at a final R = 0.0956 (wR_2 = 0.2329, for all 4685 data, 289 parameters, mean and maximum δ/σ 0.000, 0.000) with allowance for the thermal anisotropy of all non-hydrogen atoms. Minimum and maximum final electron density -0.481 and 0.475 e.Å^{-3}. A weighting scheme $w = 1/[\sigma^2(Fo^2) + (0.0403*P)^2 + 0.00*P]$ where $P=(Fo^2 + 2 * Fc^2)/3$ was used in the latter stages of refinement. Complex scattering factors were taken from the program package SHELXTL^Y as implemented on the Viglen Pentium computer

Supplementary material (see Appendix I)

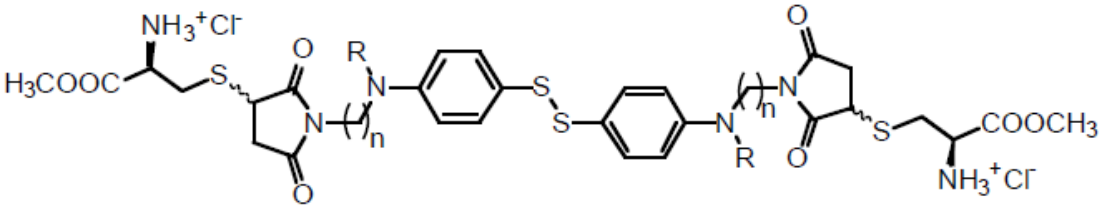
anisotropic thermal vibrational parameters with e.s.d.s

hydrogen atom position parameters
observed structure amplitudes and calculated structure factors

3.5.14 Procedure for the preparation of 8, 9 and 10



8



9, R = -CH_2CH_3, n = 2
10, R = H, n = 1

General method

The reaction of **4**, **5** and **7** with L - cysteine methyl ester hydrochloride was carried out according to the following general procedure:

A deoxygenated water solution (20 mL) of L-cysteine methyl ester hydrochloride (85.5 mg, 0.5 mmol·) was added to a deoxygenated solution of disulfide (0.25 mmol) in 20 mL of acetonitrile (THF for **7**), stirring under nitrogen. The yellow solution (pH 4.4) became colourless in the first 15 minutes †. After 2 hours the solution was concentrated under reduced pressure and the aqueous layer was lyophilised to give a white solid, which was used, without further purification, for the photolysis experiments. An approximated 93% conversion of starting disulfide to expected product was estimated from ^{1}H-NMR data of the reaction mixture (see Chapter 3 and Appendix II):

· For the preparation of **8**, the amount of cysteine used was 0.52 mmol.

† To carry out the reaction at pH 6, cysteine methyl ester (0.5 mmol) was dissolved in 20 mL of 0.013 M phosphate buffer (2.6 mL of 0.1 M Na_2HPO_4/Na H_2PO_4 pH 7, diluted to 20 mL).

Reaction mixture of **8**:
^{1}H-NMR (DMSO-d_6): 8.87 (s, broad, -$NH_3^+Cl^-$), 7.74 (d, phenyl), 7.68 (d, phenyl), 7.62 (d, phenyl), 7.35 (m, phenyl), 7.1 (m, phenyl), 4.61 (m), 4.36 (m), 4.30 (m), 3.76 (s), 3.75 (s), 3.74 (s), 3.72 (s), 3.36 (m), 2.68 (m). MS (ES^+) m/z: 474 (20%), 679 ($M+H^+$, 100%), 884 (10%), $C_{28}H_{30}N_4O_8S_4$ requires 678.82. UV/Vis. (H_2O): ε_{251}= 15216, ε_{260}= 13066, ε_{280}= 5777.

Reaction mixture of **9**
^{1}H-NMR (CD_3OD): 7.31(phenyl), 6.75 (phenyl), 4.5 (m), 3.98 (m), 3.91(s), 3.90 (s), 3.87 (s), 3.84 (q, J = 4.8), 3.69 (m), 3.66 (m), 3.57 (m), 3.45 (m,), 3.3 (d), 3.25 (d), 3.15 (m), 2.50 (m, 1H), 2.42 (m, 1H), 1.36 (t, J = 7.3), 1.19 (t, 6H, J = 7), 0.98 (t, J = 7). HRMS (FAB^+) m/z: 821.249671 ($M+H^+$), $C_{36}H_{48}N_6O_8S_4$ requires 821.249475. UV/Vis. (H_2O); ε_{260}= 13691.

Reaction mixture of **10**:
^{1}H-NMR (CD_3OD): 7.21 (d, 4H, phenyl), 6.8 (d, 4H, phenyl), 4.46 (m, 1H), 4.55 (m, 1H), 3.90 (s), 4.07 (m, 2H), 3.89 (s), 3.86 (s), 3.84 (s), 3.76 (m), 3.26 (m), 2.55 (2H, m). HRMS (FAB^+) m/z: 737.1565 ($M+H^+$), $C_{30}H_{36}N_6O_8S_4$ requires 737.1556. UV/Vis. (H_2O); ε_{260}= 13854.

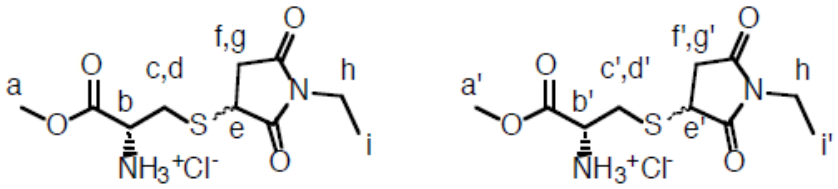
*Reaction of **4** with an excess of cysteine methyl ester*
The procedure was the same as described for the preparation of **8**, **9** and **10**, but the amounts of reagents were modified as follows: 25 mg (0.06mmol) of **4** dissolved in 7 ml of acetonitrile were added to 21.0 mg (0.6 mmol) of L-cysteine methyl ester hydrochloride dissolved in 7 mL of water. The excess of cysteine was not removed from the final mixture.
^{1}H-NMR (DMSO-d_6): 2.72 (m, 1H), 3.41 (m, 13H), 3.70 (s, 1H), 3.73 (s, 1H), 3.74 (s, 1H), 3.75 (s), 4.26 (m), 4.32 (m), 4.96 (t, 1H), 7.33 (m, phenyl), 7.66 (d, phenyl), 7.7 (d, phenyl), 9.02 (s, broad, -NH3$^+$Cl$^-$). MS (ES^+) m/z: 136 (cysteine methyl ester), 474 ($M+H^+$, 100%), 679 (40%). $C_{28}H_{30}N_4O_8S_4$ requires 678.82.

*Reaction of cysteine methyl ester with an excess of **4***

The procedure was the same as described for preparation of **8, 9** and **10**, but the amounts of reagents were modified as follows: 57 mg (0.14 mmol) of **4** dissolved in 12 mL of acetonitrile were reacted with 8 mg (0.04 mmol) of L-cysteine methyl ester hydrochloride dissolved in 12 mL of water. The excess of **4** was removed by filtration after evaporation of the acetonitrile from the reaction mixture.

^{1}H-NMR (DMSO-d_6): 8.85 (s, broad, -NH_3^+), 7.68 (t, phenyl), 7.4 (d, phenyl), 7.33 (m, phenyl), 7.18 (s, maleimide), 4.39 (m), 4.26 (m), 3.75 (s), 3.74 (s), 3.73 (s), 3.68 (s), 3.55 (s), 3.43 (m,), 3.18 (m), 3.03 (m), 2.69 (m). MS (ES^+) m/z: = 544 ($M+H^+$, 100%), 679 (40%). $C_{28}H_{30}N_4O_8S_4$ requires 678.82.

3.5.15 Synthesis of 23



23, isomer A **23**, isomer B

38 mg (0.3 mmol) of N-ethylmaleimide were added over a deoxygenated water solution (10 mL) of L-cysteine methyl ester hydrochloride (51 mg, 0.3 mmol) under stirring. After one hour, the solvent was lyophilised to give a white solid (89 mg, 100% yield).

^{1}H-NMR (CD_3OD): 4.67 (s, broad, $NH_3^+Cl^-$), 4.38 (dd, 0.4H, **b,** J = 4.40, J' = 8.22), 4.30 (dd, 0.6H, **b'**, J = 4.40, J' = 8.22), 3.92 (m, 1H, **e**, **e'**), 3.70 (s, 3H, **a**, **a'**), 3.54 (dd, 0.6H, **c'**, J = 4.40, J' = 14.97), 3.36 (m, 2H, **h**, **h'**), 3.32 (m, 0.4H, **c**), 3.19 (dd, 0.4H, **d**, J = 8.22, J' = 14.97), 3.09 (m, 1H, **f**, **f'**), 3.03 (m, 0.6H, **d'**), 2.33 (m, 1H, **g**, **g'**), 0.96 (t, 3H, **i**, **i'**, J = 7.34). MS (ES^+) m/z: 261 ($M+H^+$). $C_{10}H_{16}N_2O_4S$ requires 260.31.

3.6. References

1. F. C. Thyrion, *J. Phys. Chem*. **1973**, *77*, 1478–1482.
2. N. A. Borisevich, Ya. N. Malkin, Sh. Ruziev, S. V. Mel' nichuk, S. A. Tikhomirov, G. B. Tolstorozev, V. A. Kuz' min, *Bull. of the Academy of Science of the USSR division of chemical science* **1990**, *39*, 468–471.
3. O. Ito, M. Matsuda, *J. Am. Chem. Soc.* **1983**, *105*, 1937–1940.
4. N. P. Ernsting, *Chem. Phys. Lett*. **1990**, *166*, 221 – 226.
5. N. P. Ernsting, *Ultrafast Phenomena VIII*; Springer Series in Chemical Physics **1993**, *55*, 638–640, Editors: J. L. Martin, A. Migus, G. A. Mourou, A. H. Zewail.
6. Y. Hirata, T. Niga, T. Okada, *Chem. Phys. Lett*. **1994**, *221*, 283–288.
7. N. A. Borisevich, S. V. Mel' nichuk, S. A. Tikhomirov, G. B. Tolstorozev, *Ultrafast Phenomena in Spectroscopy*; Springer Series in Chemical Physics **1990**, *49*, 276 – 281. Editors: E. Klose, B. Wilhelmi
8. N. P. Ernsting, T. Bultmann, *J. Phys. Chem*. **1996**, *100*, 19417–19424.
9. Y. Hirata, Y. Niga, S. Makita, T. Okada, *J. Phys. Chem. A* **1997**, *101*, 561-565.
10. T. W. Scott, S. N. Liu, *J. Phys. Chem*. **1998**, *93*, 1393 – 1396.
11. Y. Schaafsma, A. F. Bickel, E. C. Kooyman, *Tetrahedron* **1960**, *10*, 76–82.
12. O. Ito, M. Matsuda, *J. Am. Chem. Soc*. **1982**, *104*, 568–572.
13. O. Ito, M. Matsuda, *J. Phys. Chem.* **1984**, *88*, 1002–1005.
14. E. Rabinowitsch, J. Franck., *Trans. Faraday Soc.* **1934**, *30*, 120–131.
15. T. Koenig, H. Fischer, *Free Radicals*, J. K. Kochi, Editor: Wiley **1973**, Vol. I, Chapter 5.
16. H. S. M. Lu, M. Volk, Y. Kholodenko, E. A. Gooding, R. M. Hochstrasser, W. F. DeGrado, *J. Am. Chem. Soc.* **1997**, *119*, 7173-7180.
17. H. S. M. Lu, M. Volk, Y. Kholodenko, E. A.Gooding, R. M. Hochstrasser, W. F. DeGrado, *J. Phys. Chem. B* **1997**, *101*, 8607-8616.
18. M. S. Alnajjar, M. S. Garossian, S. T. Autrey, K. F. Ferris, J. A. Franz, *J. Phys. Chem*. **1992**, *96*, 7037–7043.
19. J. L. Male, B. E. Lindfors, K. J. Covert, D. R. Tyler, *J. Am. Chem. Soc*. **1998**, *120*, 13176–13186.
20. D. A. Braden, E. E. Parrack, D. R. Tyler, *Coordination Chemistry Rev*. **2001**, *211*, 279-284.
21. H. R. Ward, *Free Radicals*, J. K. Kochi, Editor: Wiley **1973**, Vol. I, Chapter 6.
22. M. Volk, Eur. *J. Org. Chem.* **2001**, 2601–2621
23. Ya. V. Krivosheev, S. V. Mel' nichuk, V. G. Plotnikov, S. A. Tikhomirov, G. B. Tolstorozev, *Dokl. Akad. Nauk. SSSR* **1990**, *312*, 913–918.

24. N. A. Borisevich, S. V. Mel' nichuk, S. A. Tikhomirov, G. B. Tolstorozev, *Izv. Ross. Akad. Nauk. Ser. Fiz*. **1992**, *56*, 64-67.

25. G. Jeschke, M. Wasaka, Y. Sakaguchi, H. Hayashi, *J. Phys. Chem*. **1994**, *98*, 4069–4075.

26. G. H. Morine, R. R. Kuntz, *Chem. Phys. Lett***. 1979**, *67*, 552-554.

27. N. P. Ernsting, A. Lochschmidt, N. Eilers – Konig, N. Heineking, *J. Phys. Chem. A* **1999**, *103*, 1776–1784.

28. S. N. Batchelor, C. W. Kay, K. A. McLauchlan, M. T. Yeung, *J. Phys. Chem*. **1993**, *97*, 4570–4572.

29. T. Autrey, C. Devasoss, B. Sauerwein, J. A. Franz, G. B. Schuster, *J. Phys. Chem*. **1995**, 99, 869–871.

30. P. T. Litak, J. M. Kauffman, *J. Heterocyclic Chem*. **1994**, *31*, 457-479.

31. P. O. Tawney, R. H. Snyde, R. P. Conger, K. A. Leibbrand, C. H. Stiteler, A. R. Williams, *J. Org. Chemistry* **1961**, *26*, 15–21.

32. D. G. smith, A. Nagamatsu, J. S. Fruton, *J. Org. Chemistry* **1960,** *82*, 4700-4604.

33. M. Hashimoto, E. Majima, S. Goto, Y. Shinohara, H. Terada, *Biochemistry* **1999**, *38*, 1050–1056.

34. G. E. Means, R. E. Feeney, *Chemical Modifications of Proteins* **1971**, Chapter 6, Holden–Day, San Francisco, California.

35. G. D. Reid, D. J. Whittaker, M.A. Day, D. A. Turton, V. Kayser, J.M. Kelly, G. B. Beddard, *J. Am. Chem. Soc.* **2001**, *124*, 5518-5527.

36. G. S. Beddard, M. G. Hyde, *Chemical Physics* **1991**, *151*, 239–248.

37. A. G. Cole, L. M. Yoder, J. J. Shiang, N. A. Anderson, L. A. Walker II, M. M. Banaszak Holl, R. J. Sension, *J. Am. Chem. Soc.* **2002**, 124, 434-442.

38. J. A. Riddick, W. B. Bunger, T. K. Sakan, *Organic Solvents (Physical Properties and Methods of purification)* **4th Edition**, Techniques of Chemistry, Vol. II, Wiley Interscience.

39. Y. Yoshikawa, A. Watanabe, O. Ito, *J. of Photochemistry and Photobiology A: Chemistry* **1995**, *89*, 209-214.

40. P. Suppon, *Chemistry and Light*, The Royal Society of Chemistry **1994**, 44-46.

41. D. H. Williams, I. Fleming, *Specrtoscopic methods in organic chemistry*, **5th** Ed., 20-20.

Chapter 4

4. Preparation of Photocaged Cyclic NPGK

4.1 Introduction

The elucidation of the mechanism by which proteins acquire their unique three-dimensional structure is essential to understand their biological activity. The spontaneous refolding of proteins *in vitro* suggests that the information needed to specify the native structure and biological activity of a protein is contained in its amino acid sequence.[1-4] Experimental investigations and computational studies of folding reactions may provide the information needed to discover the link between the structure and function of proteins. This information, which is referred to as the second half of the genetic code, will complete the genome project and create a revolution in protein physics, chemistry and biology.[5-7]

One problem with experimental approaches is the difficulty in triggering and observing the dynamics of refolding on the submicrosecond timescale, which is the timescale of secondary structure and hydrophobic collapse formation.[8-12] Advances in laser technologies have allowed the development of methods which can trigger and observe reactions on nanosecond timescales. To date, these methods have been applied almost exclusively to study the folding of specific families of proteins that either bear amenable prosthetic groups (hemoproteins) or undergo cold denaturation.[13-16] Introduction of photolabile moieties in proteins would allow a more general application of laser methodologies. One promising approach has been successfully applied to the study of the folding of small peptides, where photolabile linkers may be easily introduced by solid phase synthesis.[17, 18] The introduction of unnatural amino acids into a protein is more challenging, and so photolabile crosslinks have not yet been applied to the study of protein folding.[19] Our general strategy to study folding events on the nanosecond timescale makes use of chemical labelling methods to prepare photocaged proteins so that folding can be initiated photochemically (Figure 4.1).

1 MNKKTIRDVD VRGKRV**WV**RV DFNVPMEQGA ITDDTRIRAA LPTIRYLIEH GAKVILASHL 60
61 GRPKGKVVEE LRLDAVAKRL GELLERPVAK TNEAVGDEVK AAVDRLN**C**GD VLLLENV**C**FY 120
121 PGEEKNDPEL AKAFAELADL YVNDAFGAAH RAHASTEGIA HYLPAVAGFL MEKE

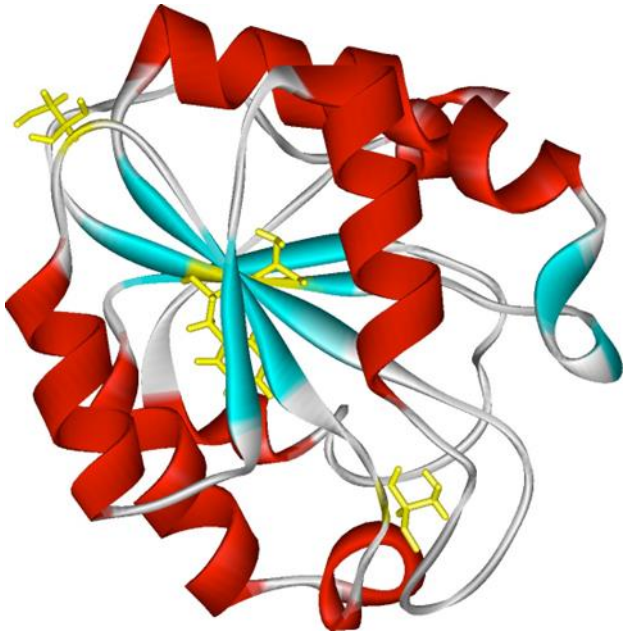


Figure 4.2. Amino acid sequence (1-174) and three dimensional structure of NPGK (derived from the crystal structure of *B. stearotermophilus* PGK). The residues subjected to mutation (C18V F17W E108C R118C) are indicated as sticks in the three dimensional structure and in bold in the amino acid sequence.

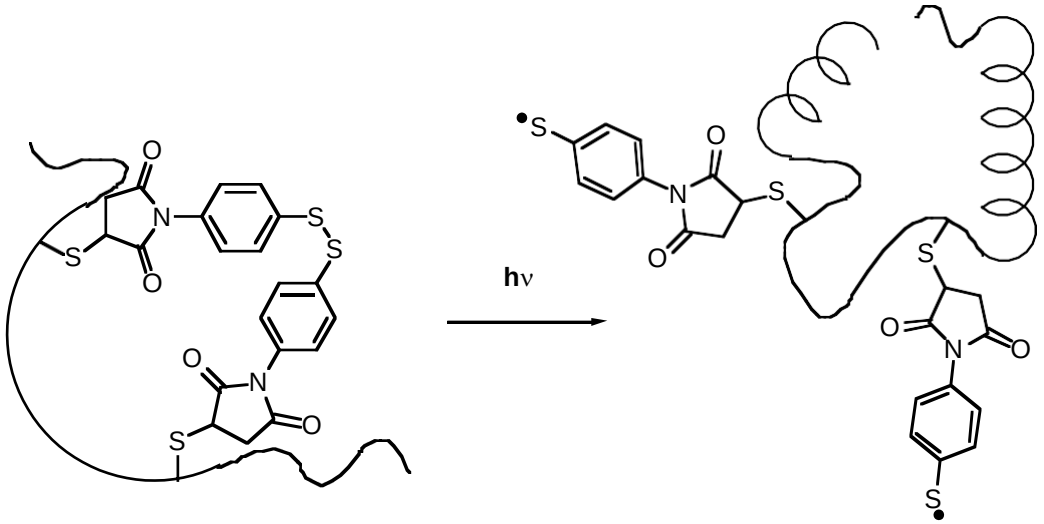


Figure 4.1. Schematic illustration of an optical trigger mechanism based on aryl disulfide crosslinks for the initiation of protein folding.

The synthesis of cyclised proteins has received new interest due to experimental observations and computational approaches which suggest that the topology of the native state is a major determinant of protein folding pathways and rates.[20, 21] Cyclisation has been achieved by two methods: formation of disulfide crosslinks between thiols inserted in the protein structure, and formation of new peptide bonds between specific residues.[22-24] In this chapter we will describe attempts to prepare a cyclic photocaged protein starting from a mutant of the N-terminal domain of phosphoglycerate kinase from *Bacillus stearothermophilus* (NPGK **1**, Figure 2.3)[25] The protein mutant **2** bears the four amino acid mutations shown in Figure 4.2: the native cysteine at position 18 has been mutated to a valine, a phenylalanine residue in position 17 has been mutated to tryptophan and two residues at position 108 and 118 have been mutated to cysteines. Tryptophan was inserted in the mutant, because it may be used as a probe for following the refolding event after flash photolysis. There is no other suitable probe in wild-type NPGK, which contains four tyrosines and no tryptophan.[26]

The design and preparation of the mutant proteins is reported elsewhere.[27] The positioning of the two cysteines was based on the X-ray crystal structure of the entire protein and on the NMR data for the N-terminal domain (NPGK).[25, 28] The residues at

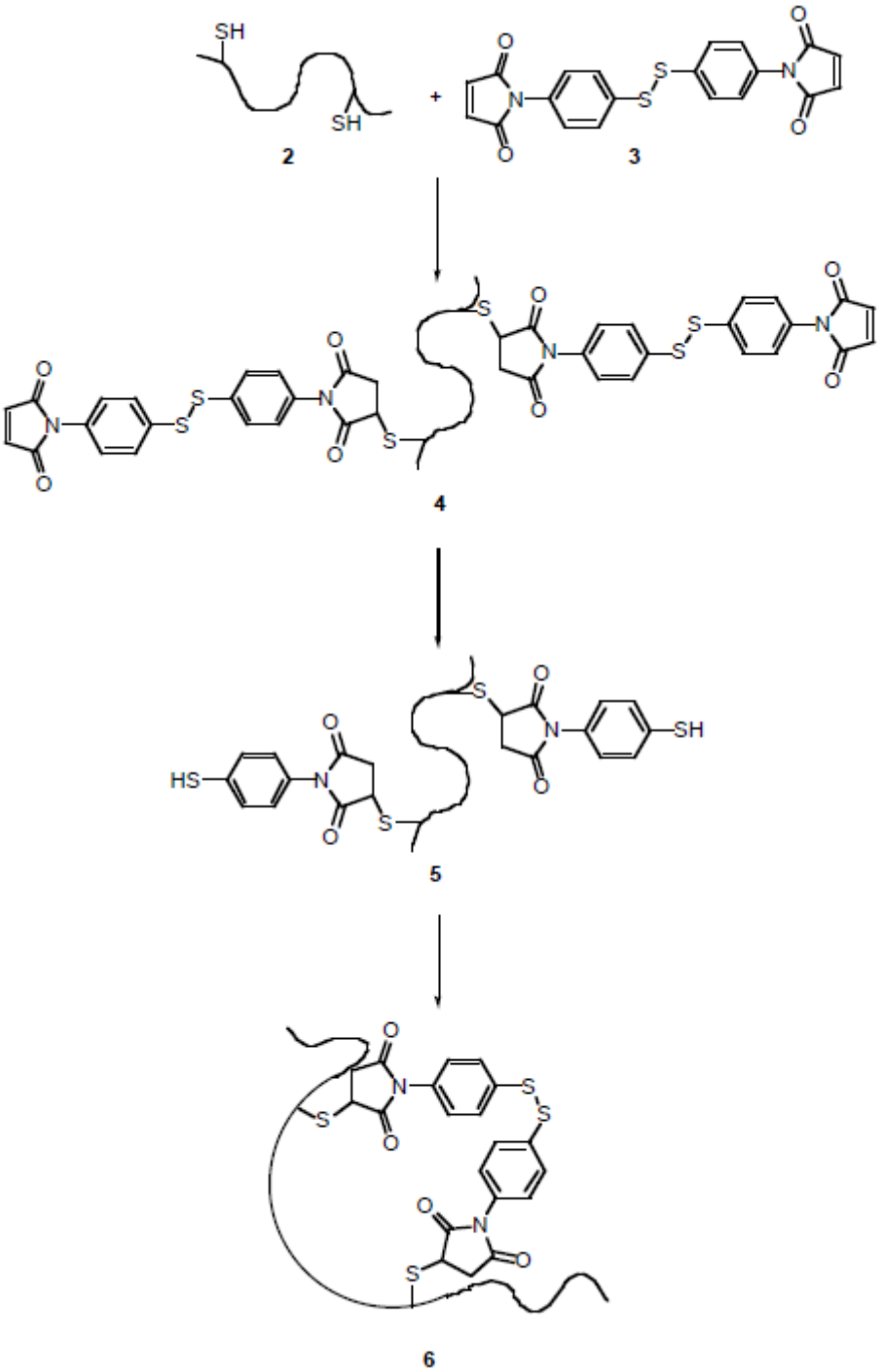


Figure 4.3. Strategy for the preparation of cyclic NPGK.

positions 108 and 118 are located on the surface of the protein, and although they are only 10 amino acids apart, they are distant in the folded protein. Therefore crosslinking should yield a structure incompatible with a native-like fold. The relatively small separation in amino acid sequence is expected to promote

intramolecular crosslinking over intermolecular reactions.[29-31] Since an excessive disruption of the native fold due to the presence of the crosslink may result in a cyclic protein too unstable to be isolated or with a refolding behaviour not comparable with that of native NPGK, residues 108 and 118 were chosen. They are in a flexible area of the native protein structure and are not involved in surface interactions or in secondary structure formation. The steric presence of the linker between these two residues will prevent the native structure being adopted. Therefore the intermolecular crosslink should perturb but not completely disrupt the native fold

4.2 Results and Discussion

4.2.1 Strategy for the preparation of the photocaged protein

The procedure that we decided to set up in order to prepare the crosslinked protein is depicted in Figure 4.3: the first step consists of labelling the cysteine residues of **2** with the bismaleimide derivative **3** in order to prepare the intermediate **4** which can then be reacted with a reducing agent to give **5**. At this stage, the intramolecular crosslink may be introduced by reaction of **5** with an oxidising agent. Unreacted alkylating agent or protein oligomers, formed as result of intermolecular reactions, should be removed by size exclusion chromatography. In order to establish the conditions for the labelling reaction, preliminary experiments were carry out using NPGK instead of the mutant which is more difficult to obtain pure

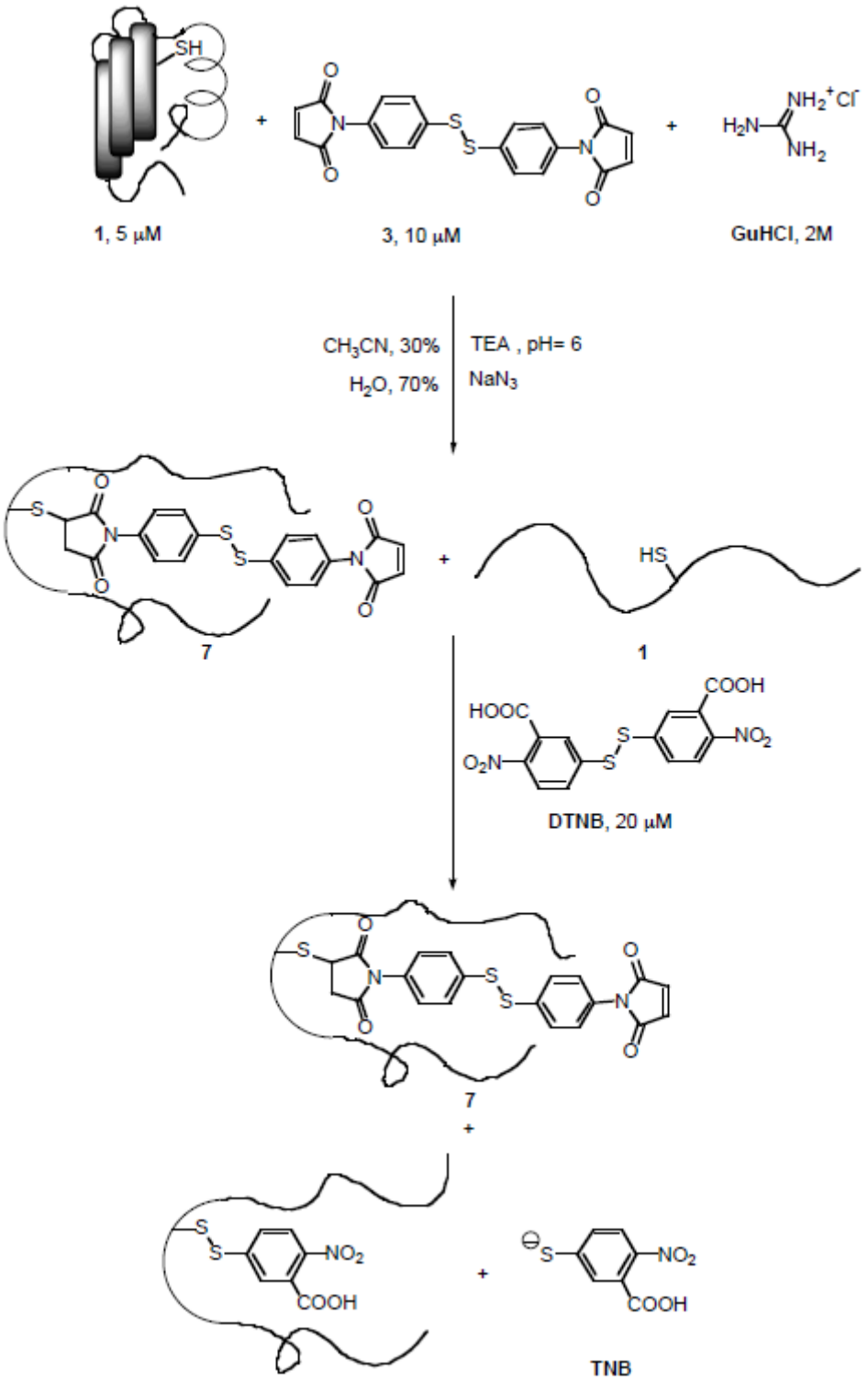
4.2.2 Labelling the single cysteine of NPGK

4.2.2.1 Kinetic experiments

The success of a labelling reaction depends on both the protein and reagent chemistry. In terms of protein chemistry, one requirement is the accessibility of the residue involved in the chemical modification. In the case of NPGK, the cysteine residue is buried in the protein interior and so the labelling reaction must be carried out in the unfolded state. Since previous studies have shown that NPGK is completely unfolded at 2 M concentration of guanidine hydrochloride (GuHCl), this concentration of denaturant was used for the labelling reaction.[28] The exposure of the thiol residue may facilitate the formation of intermolecular disulfides resulting in dimers. These side reactions may be avoided by working at high protein dilution (5-6 μM). Other side products may be formed as a result of a lack of specificity of the labelling reaction, because the maleimide reagents can alkylate histidine and lysine residues. [32, 33] These side reactions may be suppressed by working in slightly acidic conditions (TEA, triethanolamine buffer, pH 6) which protonates the amines.

Another requirement for the labelling to work efficiently is that both the protein and the maleimide reagent are in solution. Due to the insolubility of **3** in water, 30% acetonitrile in water was used. To avoid degradation of the protein, an excess of bacteriocide (sodium azide) was added to the reaction mixture. Although the reaction of maleimide with cysteine is complete in seconds at pH 7 and millimolar concentrations, longer reaction times are usually observed at lower concentrations and more acidic pH.[34-37] Thus the kinetics of the reaction of **1** with **3** were examined to estimate the time required for the reaction. The reaction was monitored by testing the amount of unreacted thiol left at different delay times after the addition of **3** to the protein solution. Unreacted thiol was detected by adding 5,5'-dithiobis (2-nitrobenzoic acid) (DTNB, Ellman reagent) and measuring the optical density at 412 nm of the thiolate (NTB) released (Scheme 4.1).[38]

Initially we carried out experiments in order to find conditions for the reaction of **1** with the Ellman reagent (Figure 4.4). Ideally, the reaction of **1** and Ellman reagent


SH
1, 5 μM
3, 10 μM
GuHCl, 2M
CH3CN, 30%
H2O, 70%
TEA , pH= 6
NaN3
7
HS
1
HOOC
COOH
O2N
NO2
DTNB, 20 μM
7
+
NO2
COOH
TNB

Scheme 4.1

should be instantaneous, with respect to the reaction of the protein with the maleimide reagent, in order to obtain quantitative data for the reaction kinetics. However when a 50-100 excess of DTNB was used at a protein concentration of 5.3 μM, it was difficult to detect any change, because the percentage of reacted DTNB was very small, and the product absorbance was swamped by starting material. Thus the concentration of Ellman reagent was reduced to 4 fold excess which allowed detection of an increase in the optical density at 412 nm (Figure 4.4). This increase takes around 10 minutes to reach a steady state, indicating that the reaction is complete by this time. The change in intensity observed at 412 nm is 70% of the value expected for 5.3 μM thiol using literature values of the extinction coefficient for the DTNB reagent (ε = 12600 M^{-1} cm^{-1} at pH 6) [39], but the absorbance is very low due to the low protein concentration and so the error in this measurement is high.

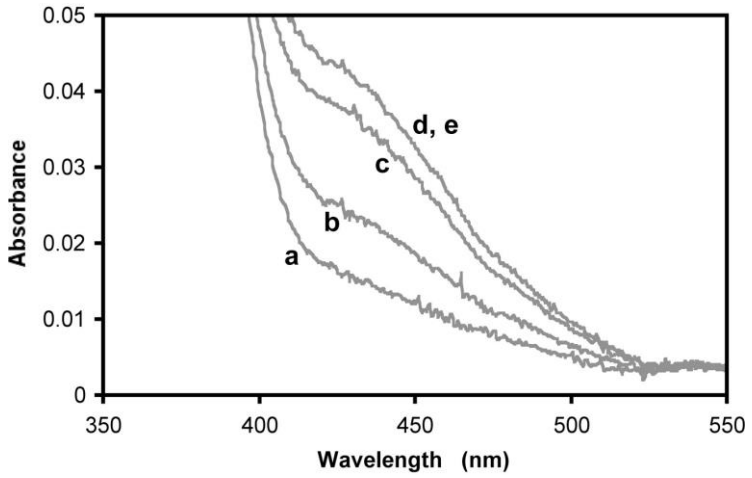


Figure 4.4. UV/Vis absorption spectra of the reaction of **1** with Ellman reagent. Spectra were taken immediately (a) and then at delay times of 2 minutes (b), 7 minutes (c), 12 minutes (d), and 30 minutes (e) after the addition of Ellman reagent to the protein solution. The reaction mixture contained: 5.3 μM protein, 20 μM DTNB, 2 M GuHCl, 50 mM TEA pH 6, 3 mM sodium azide.

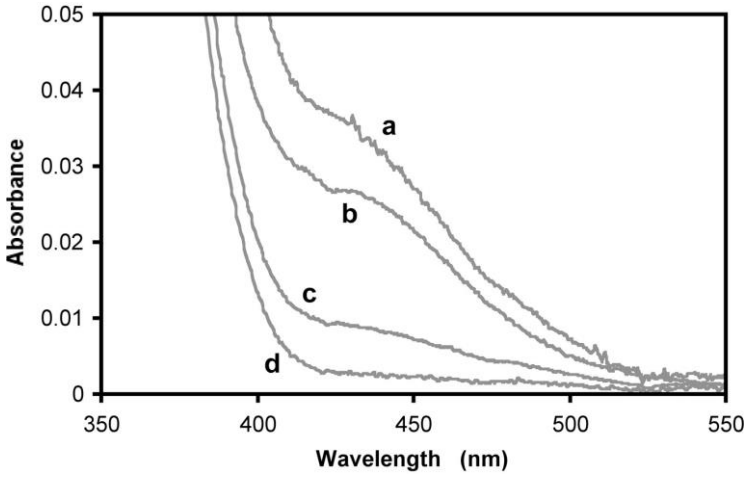


Figure 4.5.UV/Vis absorption spectra of the reaction of Ellman reagent with residual protein thiols after different reaction times of 5.3 μM **1** with 10 μM **3**: **1** and **3** were allowed to react for 30 seconds (b), 8 minutes (c) and 15 minutes (d) and then they were incubated with Ellman reagent for 7 minutes. Trace (a): UV/Vis absorption spectrum recorded after 7 minutes of reaction of **1** with Ellman reagent. The reaction mixture contained: 20 μM Ellman reagent, 2 M GuHCl, 50 mM TEA pH 6, 3 mM sodium azide.

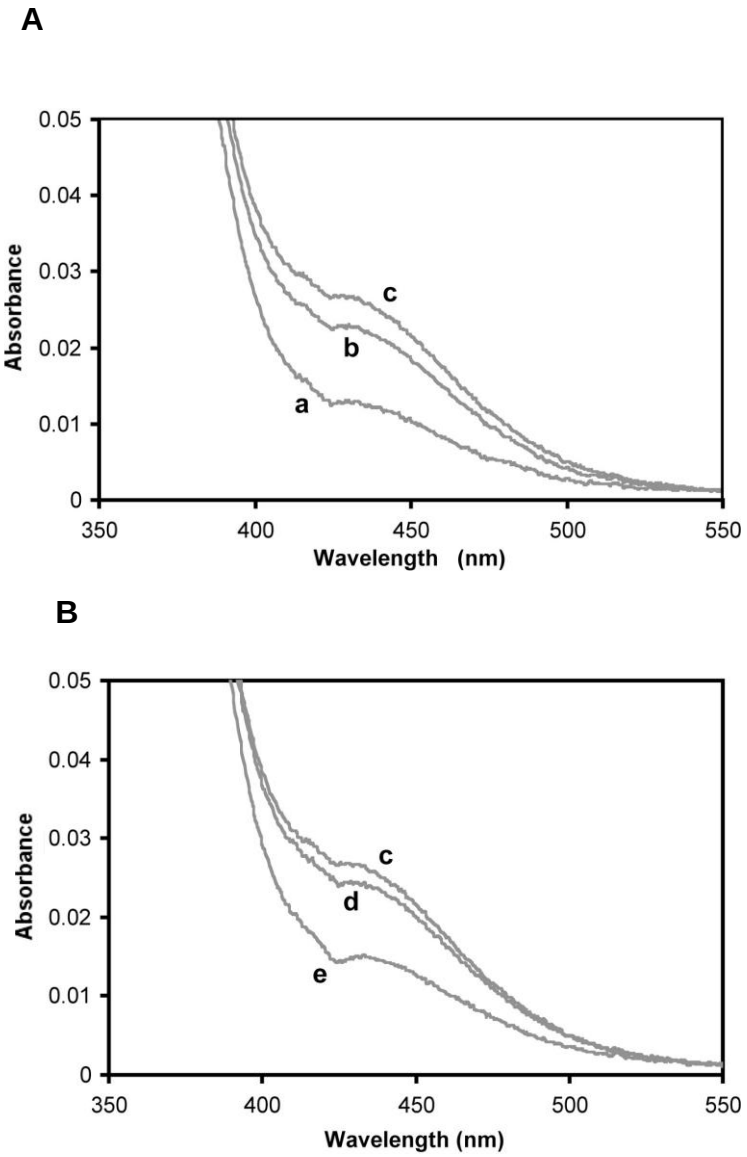
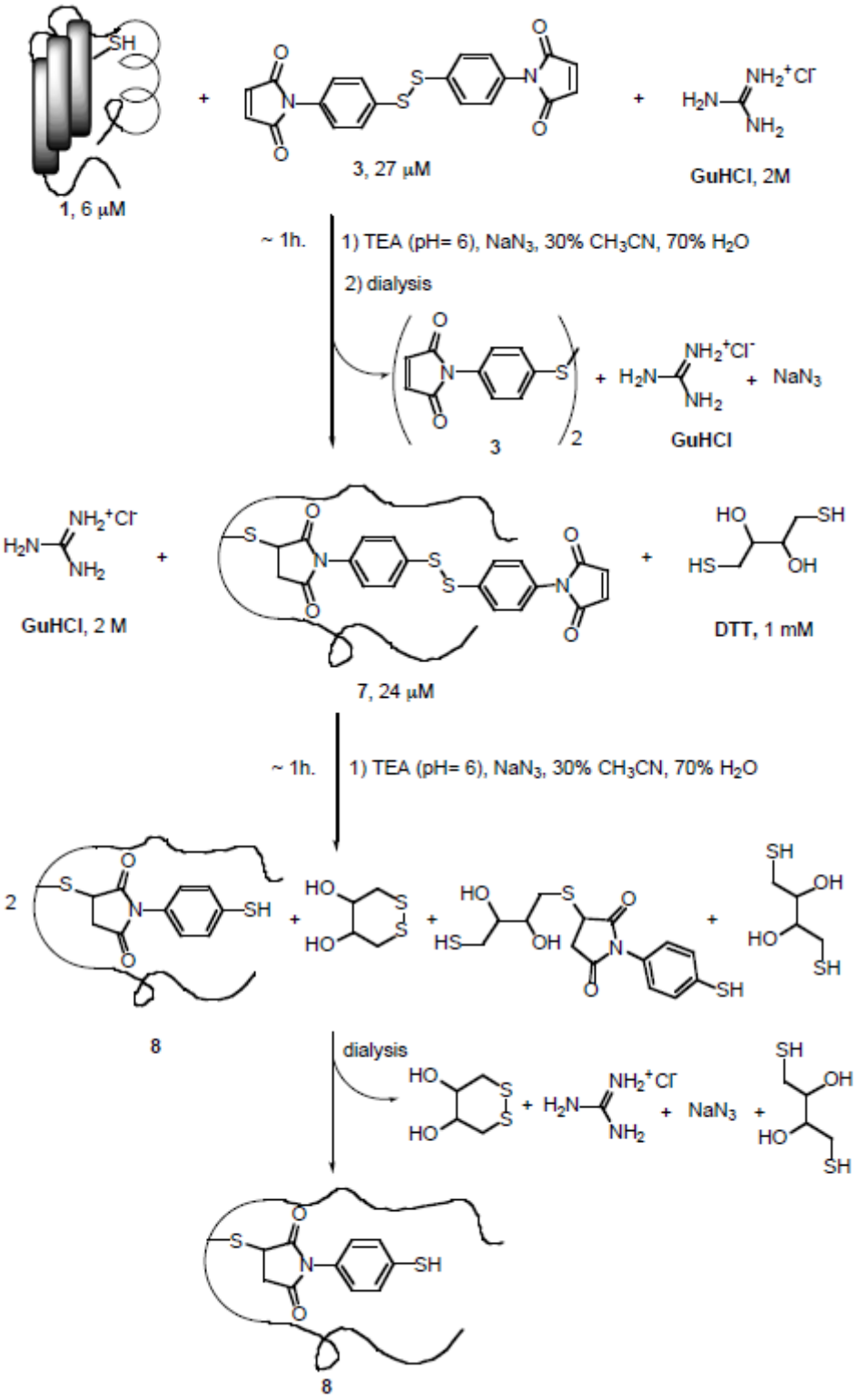


Figure 4.6. UV/Vis absorption spectra of the reaction of Ellman reagent with residual protein thiols after different reaction times of **1** with **3**: (a) **1** and **3** were allowed to react for 30 seconds (a), 3 minutes (b) and 7 minutes (c) B) **1** and **3** were allowed to react for 7 minutes (c), 12 minutes (d) and 30 minutes (e) The reaction mixture contained: 5.3 μM **1**, 10 μM **3,** 20 μM Ellman reagent, 2 M GuHCl, 50 mM TEA pH 6, 3 mM sodium azide.

The time course of the reaction of **1** with **3** was followed by adding a 2 fold excess of **3**, followed by a 4 fold excess of Ellman reagent after different delay times (Scheme 4.1, Figure 4.5). The optical density at 412 nm decreases with increasing reaction time indicating that the protein thiol reacts with the maleimide reagent. The spectra in Figure 4.5 show that within the first 15 minutes, the concentration of unreacted


1, 6 μM
3, 27 μM
GuHCl, 2M
~ 1h.
1) TEA (pH= 6), NaN3, 30% CH3CN, 70% H2O
2) dialysis
3
GuHCl
GuHCl, 2 M
7, 24 μM
DTT, 1 mM
~ 1h.
1) TEA (pH= 6), NaN3, 30% CH3CN, 70% H2O
8
dialysis
8

Scheme 4.2

protein falls below the detection limit of the Ellman assay. Thus the rates of reaction of NPGK with Ellman reagent and **3** are comparable, making it difficult to obtain kinetic parameters. This situation is further complicated by the observation that the absorbance at 412 nm reaches a maximum after 10 minutes, but then decreases. We attributed the decrease to the reaction of the thiol form of DTNB with the maleimide moiety of **3** (Figure 4.6). This explanation is supported by the observation that when ligand **3** is not present, the absorbance of the DTNB anion does not decreases (Figure 4.4).

Despite these side reactions, the kinetic experiments provide a good indication of the time required for the labelling reaction (15 minutes) so we can avoid unnecessarily long incubation times, which may result in nonspecific labelling.

4.2.2.2 Synthesis of labelled NPGK

Since the strategy for the synthesis of the cyclic protein **6** (Figure 4.3) requires a procedure for the labelling of protein thiols with **3** and a procedure for the reduction of the aromatic disulfide attached to the protein structure, we decided to prepare labelled NPGK **7** and then convert it to **8** (Scheme 4.2). In order to characterise the labelled proteins, the synthesis was carried out on a small scale (~ 0.3 mg of protein), and the products were analysed by mass spectroscopy. NPGK and **3** were incubated for one hour and the reaction was terminated by diluting ten fold in water. The solution was dialysed to increase the concentration of **7** and reduce the concentrations of salts and malemide reagent. Addition of a 100 fold excess of dithiothreithol (DTT) at pH 6 gave **8** after dialysis. The mass spectrum of **1**, **7** and **8** are shown in Figure 4.7.

The spectrum of compound **7** showed unreacted NPGK at MH =19258, and the presence of three groups of proteins each containing two species which differ by 18 in their molecular weight. This pattern may be explained as follows: the peak at MH^+= 19666 corresponds to the expected product **7** and the peak at 19684 corresponds to **7** plus water, possibly due to hydrolysis of maleimide.[40-42] The peaks at 18906–18888 and 1834-18216 correspond to truncated forms of NPGK (loss of 9

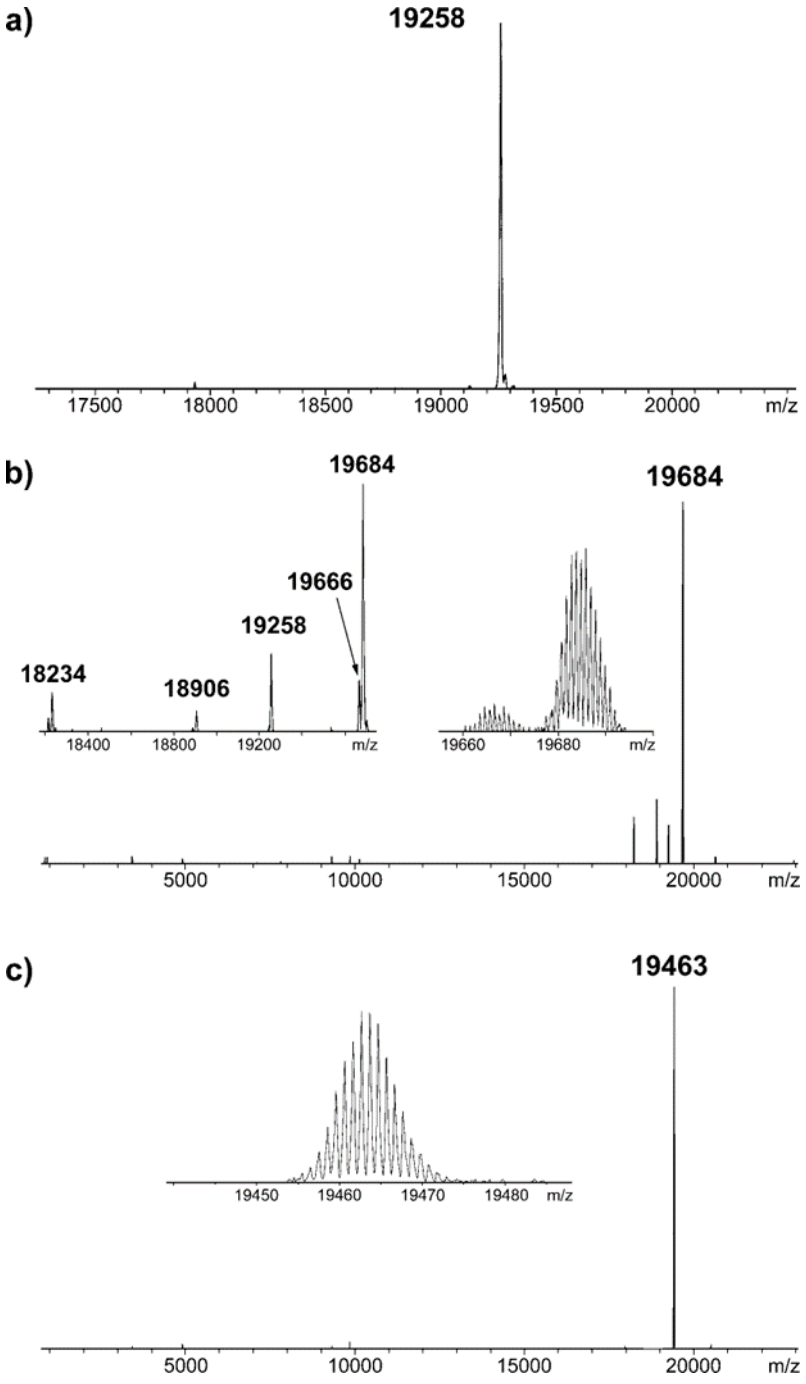


Figure 4.7. ESI-FTICR (electrospray ionisation-Fourier transform ion/cyclotron resonance) mass spectra of labelled and unlabelled NPGK. a) Deconvoluted spectrum of **1** (~ 50 μM) dissolved in a 1:1 solution of water and acetonitrile with 2% formic acid. b) Deconvoluted spectrum of **7** (~20 μM in 2% formic acid**)** and expansions. The isotopic distribution of the signal corresponding to **7** (MH^+ =19666 and [$MH^+ + H_2O$] = 19684) is also shown. c) Deconvoluted and isotopically resolved mass spectra of **8. 8** (~ 50 μM) was dissolved in a 1:1 solution of water and acetonitrile with 2% formic acid.

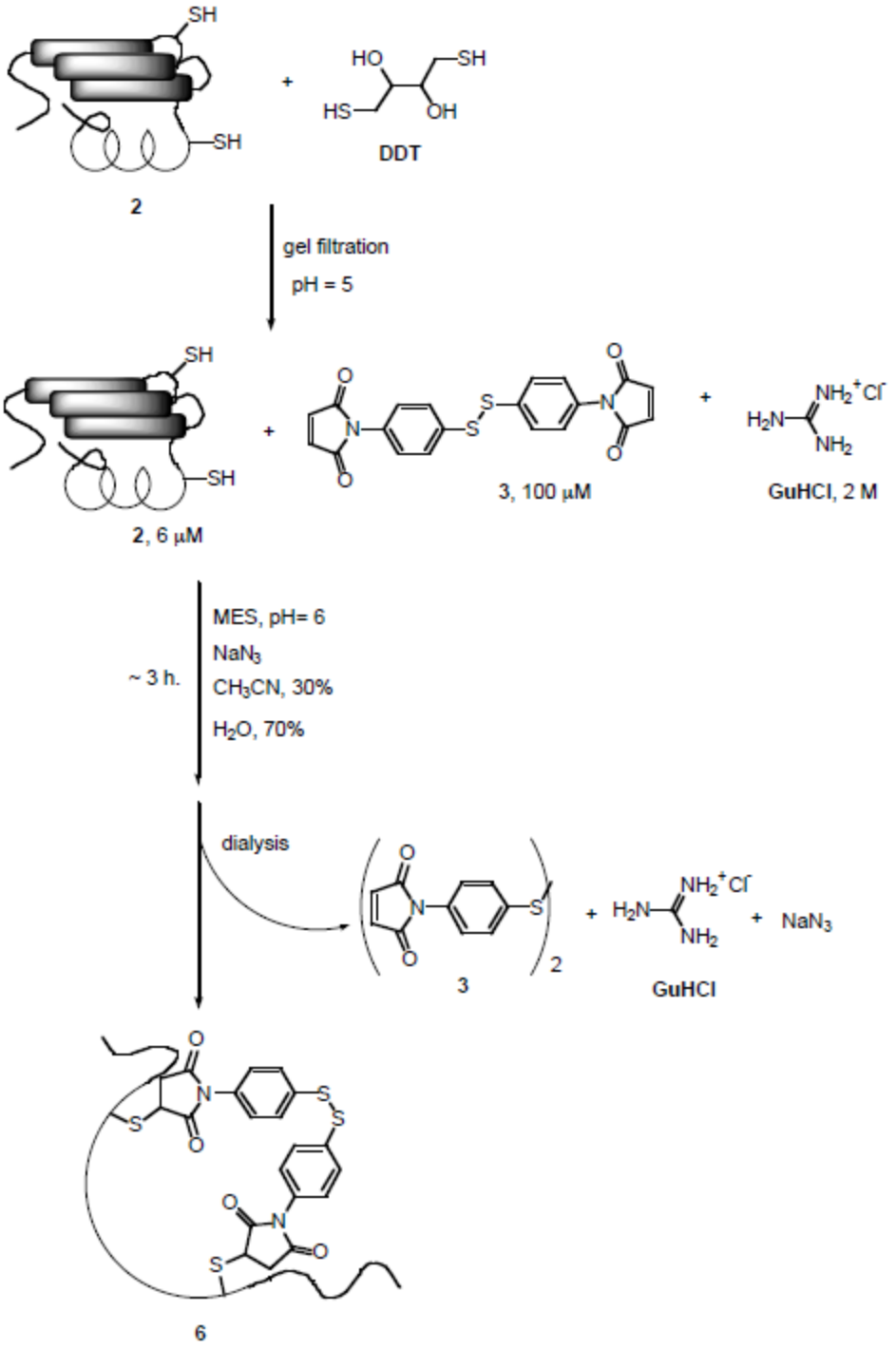
and 13 amino acids from the carboxy-terminus) labelled with the maleimide reagent **3**, together with the corresponding water adducts. The observation that no hydrolytic fragments are present in the solution of pure **1** suggests that the modified protein may be more sensitive to proteases or that the sample of **7** has been accidentally contaminated by proteases. We suggest that the latter is the case, because no truncated protein species were observed in the mass spectrum of compound **8**. The only specie present was the expected product **8** at MH^+ = 19463. The +18 water adduct is not present in this sample either.

The observation that unreacted protein was present in the spectrum corresponding to **7** but not in the spectrum corresponding to **8** is due to a difference in the reaction times in two different runs (for **7**, **1** was treated with **3** for 60 minutes, whereas for **8** the incubation time was 90 minutes). Thus we have found experimental conditions that allow specific and quantitative labelling of NPGK with a bismaleimide derivative.

4.2.3 Preparation of crosslinked NPGK

4.2.3.1 Procedure for the preparation of an analytical sample of crosslinked NPGK

Initial experiments were aimed at preparing an analytical sample of crosslinked NPGK following a procedure similar to that described above. However working with the mutant required some adjustments of the conditions. The cysteine residues of the mutant are on the protein surface, and this makes them more susceptible to oxidation to disulfides which may lead to the formation of protein oligomers. To avoid this problem the mutant was stored in excess DTT. However, DTT can react with the maleimide reagent, so it was necessary to find a procedure for removing DTT from the mutant stock solution prior to the labelling reaction.[43, 44] According to the literature, the removal of DTT should be performed at acidic pH to suppress thiol-disulfide exchange reactions.[45, 46] Thus DTT was removed by gel filtration (10 minutes) under acidic conditions (pH 5) and maintaining the protein concentration at around 20 μM. Analyses of the eluate by mass spectroscopy (Figure 4.8) showed a peak corresponding to the reduced protein at m/z = 19214 and a very small peak at


SH
SH
2
+
HO
SH
HS
OH
DDT
gel filtration
pH = 5
SH
SH
2, 6 μM
+
O
N
S
S
N
O
O
3, 100 μM
+
NH2+Cl-
H2N
NH2
GuHCl, 2 M
~ 3 h.
MES, pH= 6
NaN3
CH3CN, 30%
H2O, 70%
dialysis
O
N
S
O
2
3
+
NH2+Cl-
H2N
NH2
GuHCl
+
NaN3
O
N
S
S
S
O
O
N
O
S
6

Scheme 4.3

m/z = 19085 which corresponds to loss of methionine from the N-terminus of the protein. A double charged molecular ion was observed at m/z = 9607 together with a small amount of dimer at m/z = 38424. The dimer may correspond to a disulfide or to a non-covalent complex, but it is a very small peak.

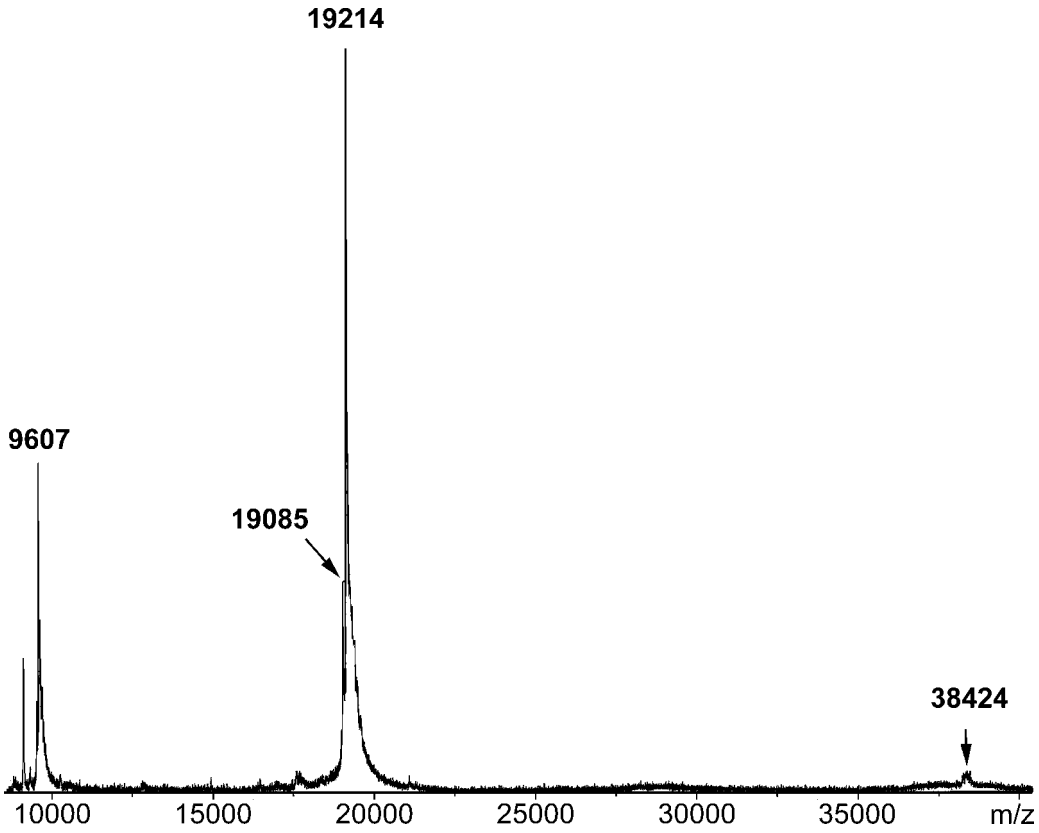


Figure 4.8. MALDI (matrix- assisted laser desorption/ ionisation) mass spectrum of **2** (~15 μM) in a 1:1 solution 0.1%of acetic acid and sinapinic acid (3, 5-dimethoxy-4-hydroxycinnamic acid, 10 mg/mL in acetone).

This shows that the NPGK mutant is a homogenous sample, and the removal of DTT does not result in formation of oligomers. From the mass spectrum, it is not possible to rule out the formation of intramolecular disulfide crosslinks, so a sample of the protein was reacted with the Ellman reagent. The assay suggested that only 60 % of the protein thiols were in the reduced state, but the experiment, and the reactivity described below suggests 100 % of the thiols are available. **2** was reacted with **3** under the following conditions: MES [2-(N-Morpholino) ethane sulphonic acid] was used to buffer the solution at pH 6, a 16.6 excess of **3** was added and the reaction mixture was incubated for 3 hours (Scheme 4.3). The buffer was changed from the

NPGK procedure to obtain the desired pH 6 (mixing a triethanolamine solution of **3** with the acidic solution of **2** gave a final pH of 5.5). The reaction time was increased based on the previous experiments with NPGK, which showed that one hour is not long enough to complete the reaction. The reaction mixture was purified by dialysis and analysed by gel electrophoresis and mass spectrometry. As shown in Figure 4.9, no oligomeric species were detected: the labelled proteins migrate similarly to the unlabelled protein.

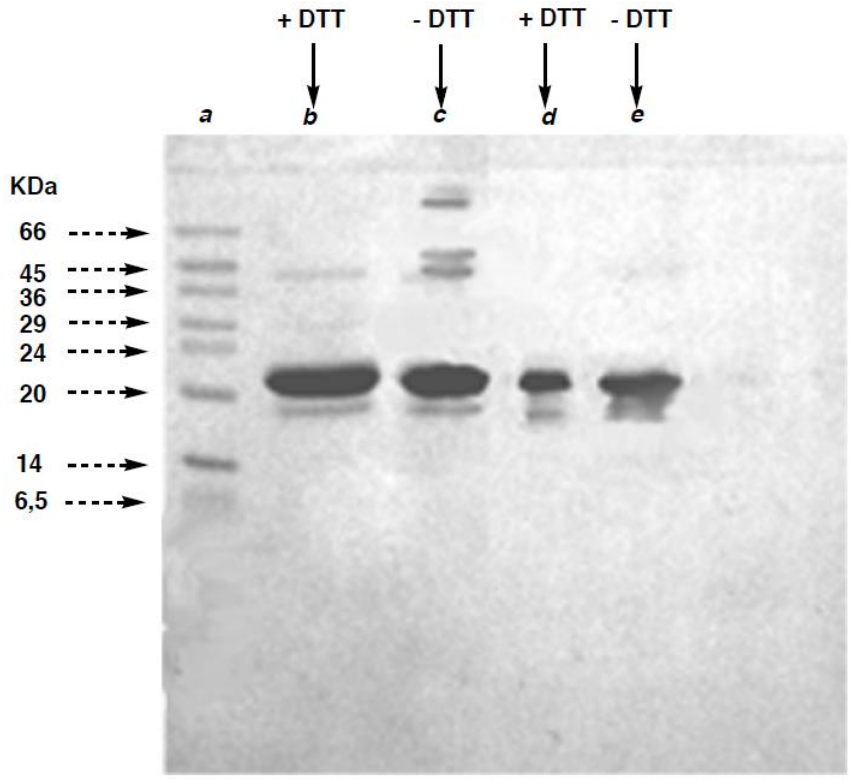


Figure 4.9. SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) of the products of the reaction of **2** and **3.** Lane *a*: Molecular weights markers. Lane *b* and *c*: **2** loaded with SDS buffer plus and minus DTT. Lane *d* and *e*: reaction mixture loaded with SDS buffer plus and minus DTT. Samples (20 μL, ~ 20 μM) were prepared as a 1:1 mixture of SDS loading buffer : protein solution, and they were heated at 100^0 C for 5 minutes prior to loading. Protein bands were stained with a solution of Coomassie blue R.

The mass spectrum of the products is shown in Figure 4.10. Four different protein species are present in the reaction mixture, none of which corresponds to the expected product **4** (MW = 20030, Figure 4.3). No unreacted protein is present which means that intramolecular disulfides were not present in the starting material

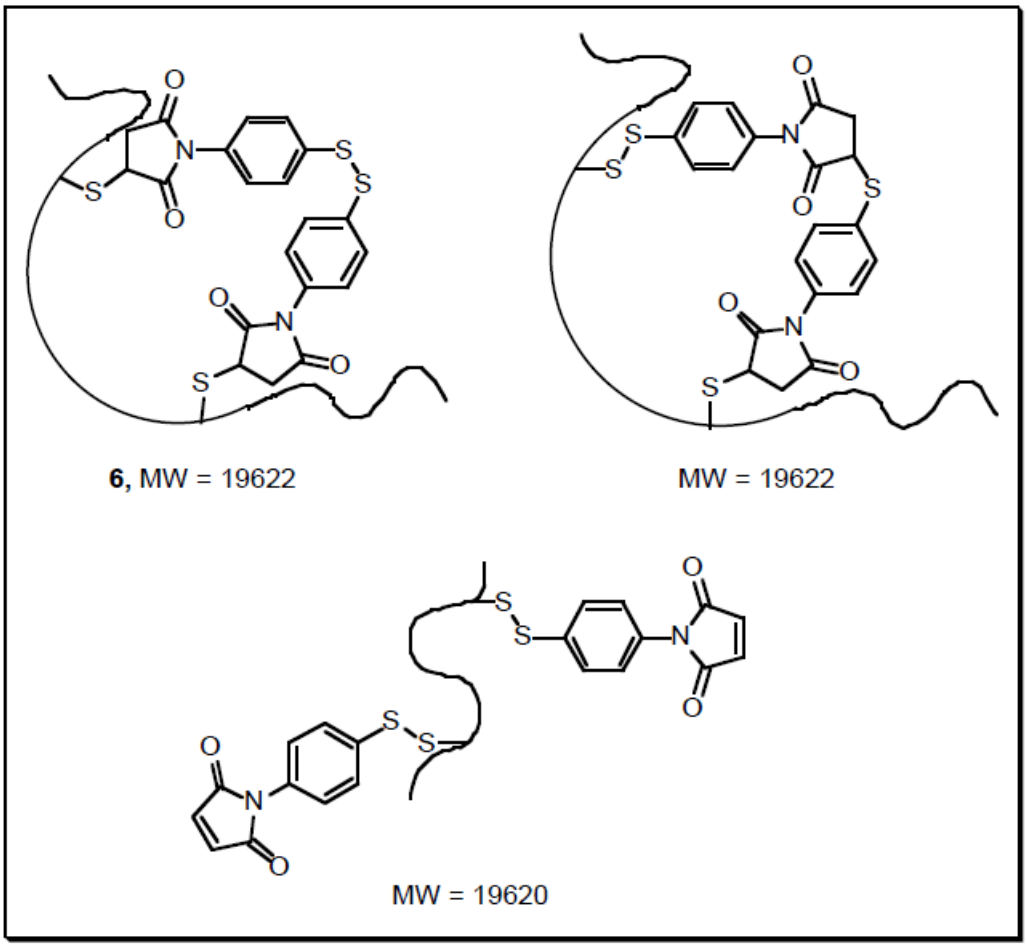


Figure 4.11. Possible structures consistent with the mass spectrum of labelled **2**

2. The major product appears at m/z = 19622, and this corresponds to the molecular weight of the final target, the cyclised protein **6**. The peak at m/z = 19491 corresponds to the cyclic crosslinked form of the truncated protein which has lost a methionine from the N-terminus sequence (present in the mutant stock solution, Figure 4.8). The peaks at m/z = 19418 and 19826 differ from 19622 by ± 204, the mass of half the malemide reagent **3**. This suggests that a thiol-disulfide exchange reaction has taken place. If this happens, then the signal at 19622 may be due to several different species, some of which are depicted in Figure 4.11.

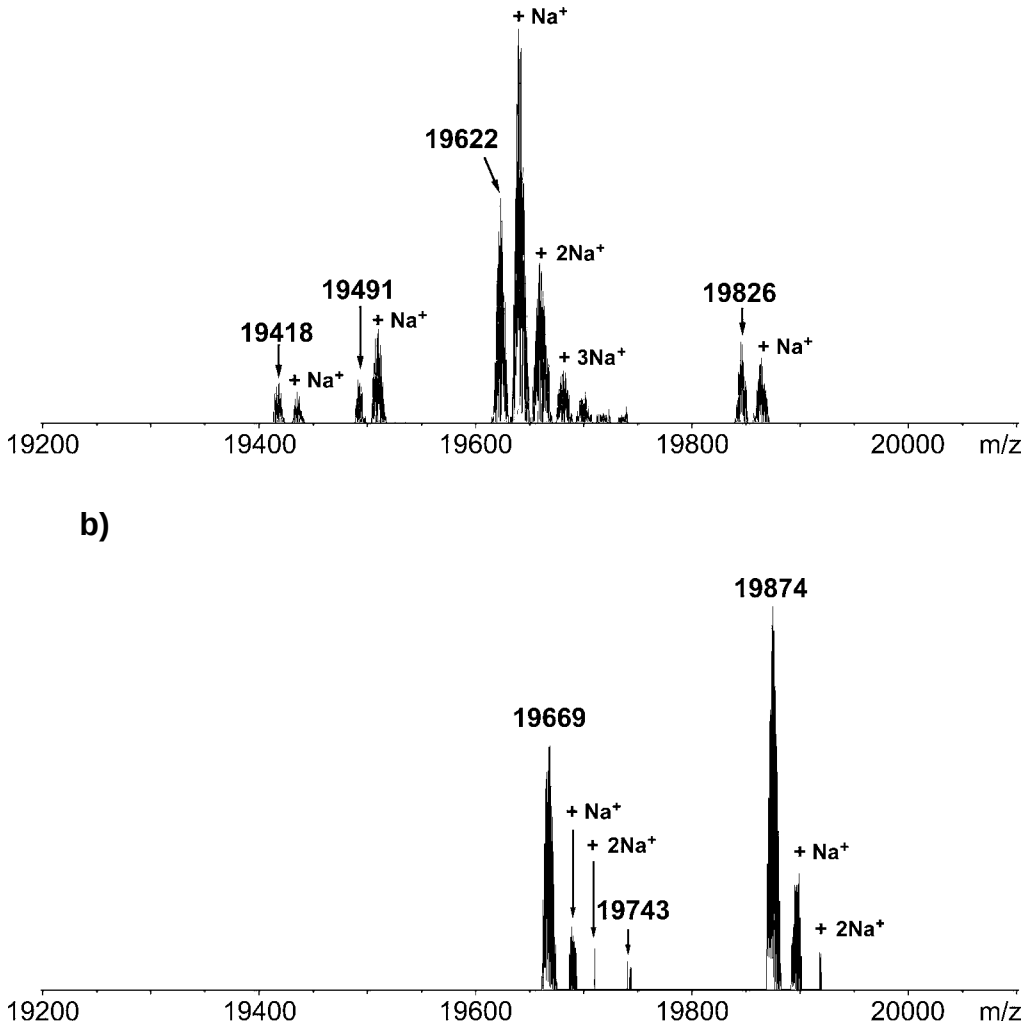
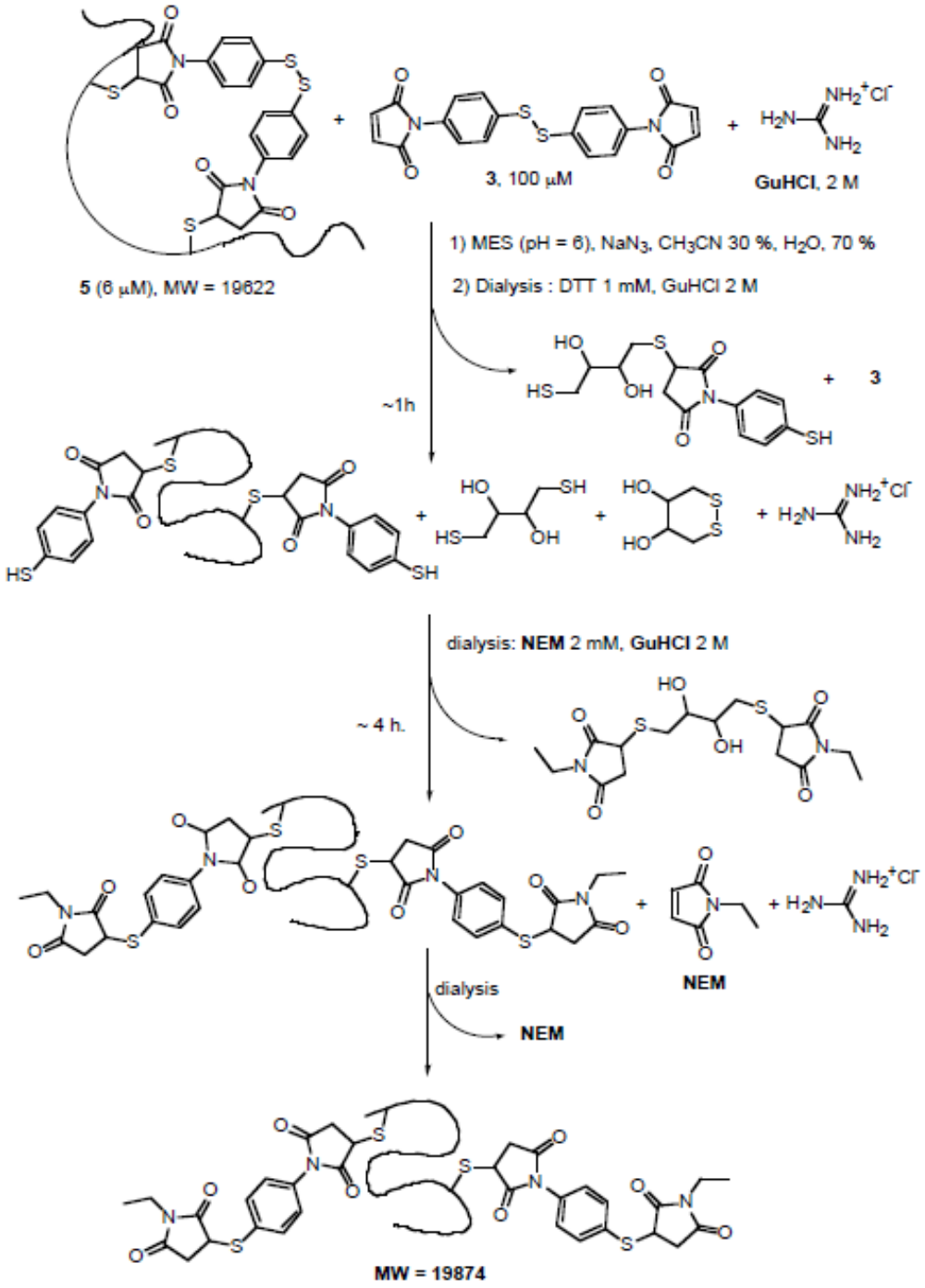


Figure 4.10. Deconvoluted ESI-FTICR mass spectra of the products obtained after reaction of **2** with **3** (a), followed by labelling with NEM (b). Proteins (~40 μM) were dissolved in a 1:1 solution of water and acetonitrile with 2% formic acid.


5 (6 μM), MW = 19622
3, 100 μM
GuHCl, 2 M
1) MES (pH = 6), NaN3, CH3CN 30 %, H2O, 70 %
2) Dialysis : DTT 1 mM, GuHCl 2 M
~1h
3
dialysis: NEM 2 mM, GuHCl 2 M
~ 4 h.
NEM
dialysis
NEM
MW = 19874

Scheme 4.4

Similarly the signals at 19418 and 19826 may be attributed to more than one structure. Another experiment was therefore carried out in an attempt to narrow down the number of possible protein species present in the reaction mixture (Scheme 4.4). **2** and **3** were reacted using the procedure previously described (Scheme 4.3). DTT was then added to the reaction mixture via dialysis to cleave all the disulfides. Then a second dialysis with a 200 fold excess of N-ethylmaleimide (NEM) functionalised all the thiols groups so formed. The dialyses were carried out under unfolding conditions by addition of guanidine hydrochloride to the DTT and N-ethylmaleimide solutions, in order to maintain the protein in the unfolded state.[47] The mass spectrum of the products showed three signals at higher mass than the original reaction mixture (Figure 4.10). The relative intensities of the signals was very different from the pattern observed for the **2** + **3** reaction mixture. It looks like the mass ion at 19826 has combined its intensity with the mass at 19418 in the course of the NEM reaction sequence. The mass values of the proteins before and after the NEM labelling differ by 2 NEM fragments (2 x 125), and an interpretation is summarised in Figure 4.12.

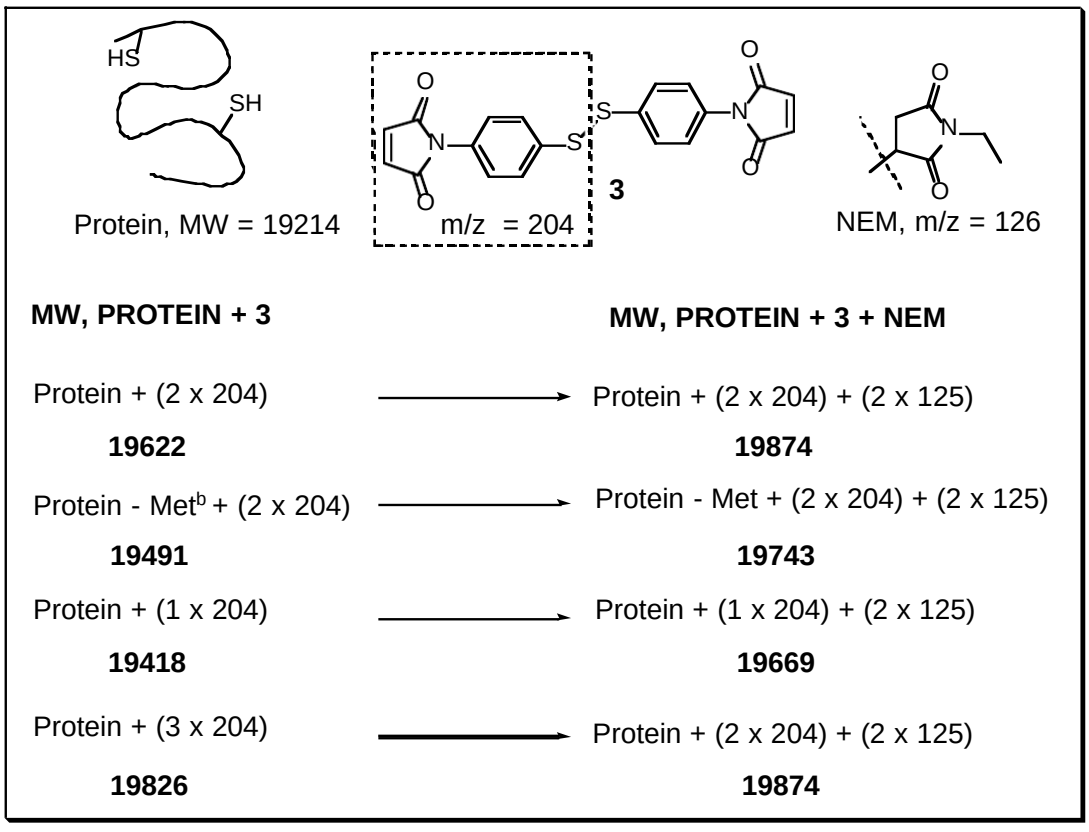


Figure 4.12. Summary of the mass values found for (**2** + **3**) reaction mixture before and after NEM labelling sequence. [b] Met = Methionine (MW =131).

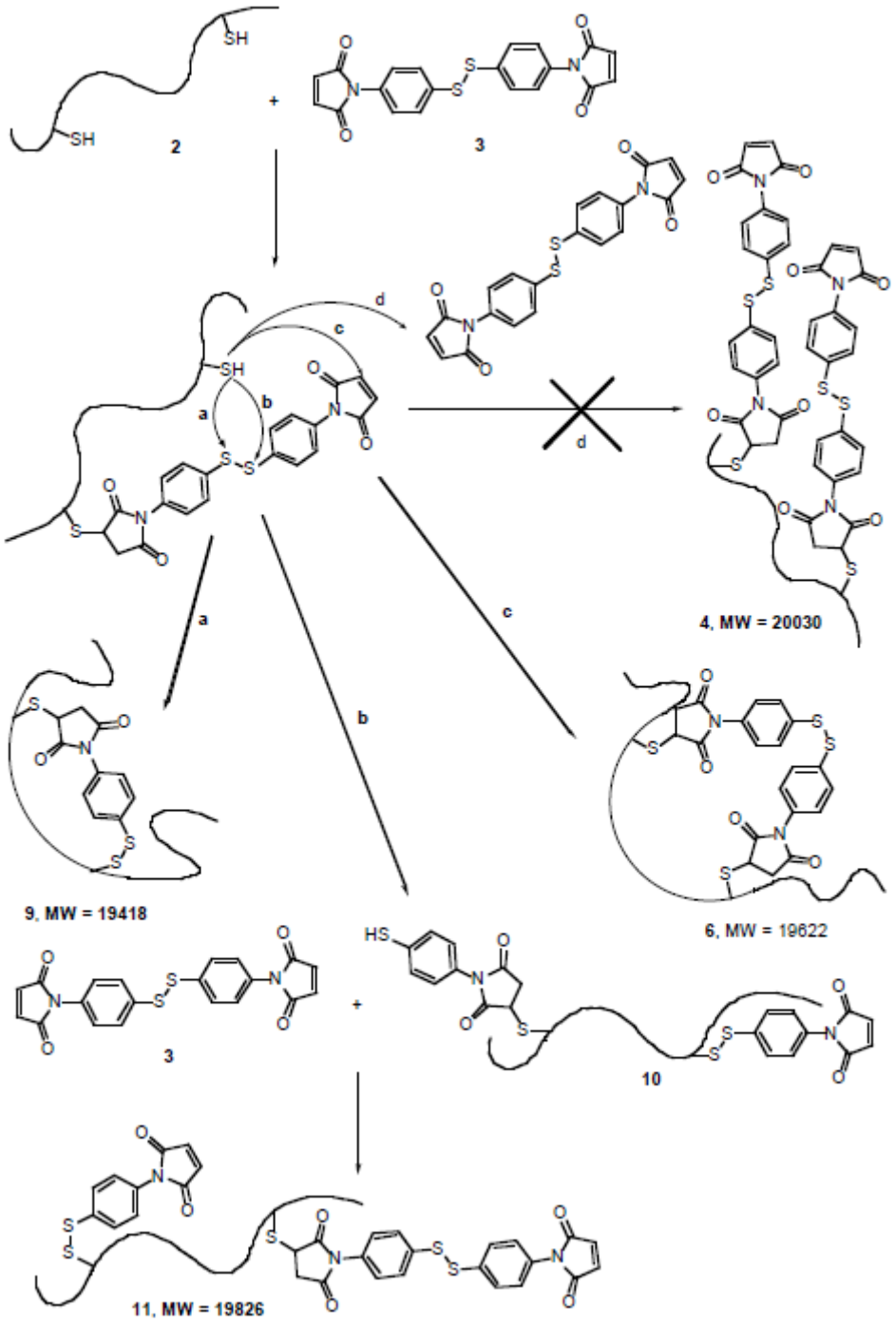


Figure 4.13. Proposed mechanism for the reaction of **2** with **3**.

The NEM experiment indicates that the proteins at m/z = 19622, 19491, 19418 each contain one disulfide, and the protein at m/z =19826 must have lost one or two fragments corresponding to half the maleimide **3** [-(1 x 204) or – (2 x 204)], added two NEM fragments (2 x 125) and thus give the signals at m/z = 19874 or 19669. The intensity pattern suggests that this protein lost 2 half maleimides as indicated. Thus this protein contains two disulfides. The only possible structure which can account for the signal at m/z = 19418 is **9** (Figure 4.13). Thus it seems that one protein thiol reacts with **3** and then the second thiol reacts intramolecularly with the attached maleimide. For this reaction to compete with the excess **3** in solution, the effective molarity of the protein thiols must be high, so that the intramolecular reaction is faster than the intermolecular one. According to literature data, the effective molarity of thiol pairs separated by 9-15 amino acid residues in an unfolded polypeptide is in the millimolar range whilst the concentration of **3** in solution was 10^{-4} M.[30, 31, 48]

Previous studies (chapter 3) have shown that thiol-disulfide exchange is observed when cysteine and **3** react in stoichiometric amounts but not when excess **3** is present in solution. The intramolecular reaction is analogous to stoichiometric addition of cysteine to **3** and thus it is not surprising that we see some thiol-disulfide exchange. The solution reactivity of **3** (chapter 3) suggests that the intramolecular addition of the maleimide moiety should be faster than attack on the disulfide bond (Figure 4.13). Thus the major product of the intramolecular process should be the target cyclic protein **6**, which explain why the signal at m/z = 19622 is the most intense in the mass spectrum.

A third intramolecular reaction is possible for the key monofunctionalised protein. Intramolecular thiol-disulfide exchange can take place in a different way giving **10**. However **10** still has a free thiol group. We can explain the fourth product if this aromatic thiol reacts on the disulfide faster than on the maleimide group. If the intermolecular reaction on the disulfide of **3** is faster than the intramolecular reaction with the maleimide bound to the protein, product **11** results. This would account for the formation of a product corresponding to the addition of one and a half ligand **3** (408 + 204) (m/z =19826). On reduction, two half ligand fragments [-(2 x 204)] are

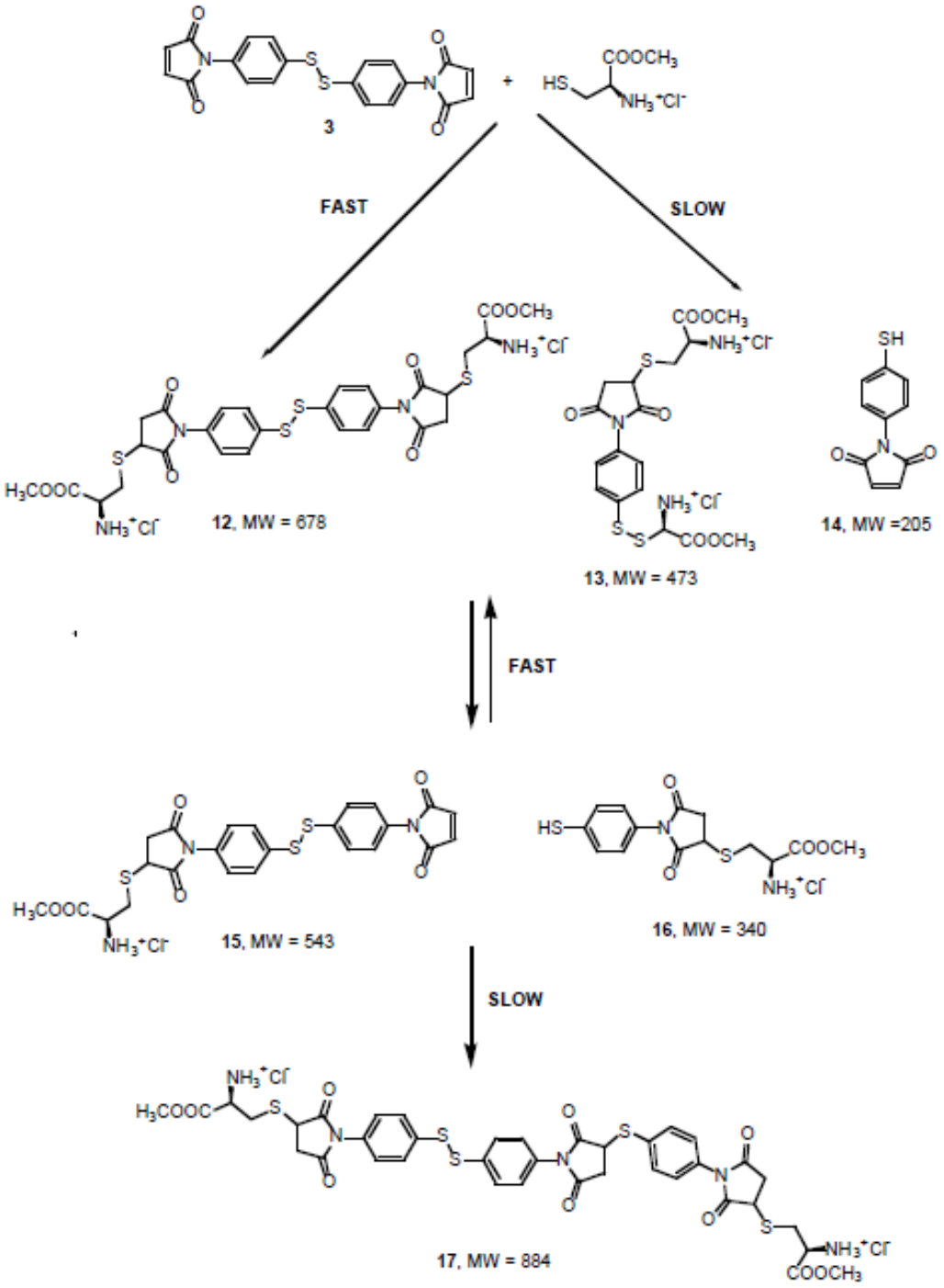


Figure 4.14. Proposed mechanism for the reaction of cysteine methyl ester with **3**.

lost giving two free thiols. Addition of two NEM fragments gives the signal at m/z = 19666 observed in the mass spectrum of the NEM labelled proteins (Figure 4.10). An alternative approach to the one used in here would be to carry out the labelling reaction under folding conditions. If the effective molarity of the thiols in the folded state is lower than in the unfolded state, then the formation of side products due to intramolecular crosslinking could be suppressed thus leading to a cleaner sample of functionalised protein.

The analysis above suggests that aromatic thiols undergo disulfide exchange faster than addition to maleimide, but the reactivity of aliphatic thiols is the other way around. This requires further investigations, but it also explains the product distribution observed when **3** was reacted with cysteine methyl ester (Figure 4.14). We have shown that the reaction of cysteine with the maleimide is faster than the thiol disulfide exchange reaction, therefore **12** is the major product and **13** and **14** are minor products. The aromatic thiol **14** should preferentially undergo thiol-disulfide exchange with **12** giving **15** and **16**. Once a thiol-disulfide exchange equilibrium is established, slow addition of **16** to the maleimide of **15** leads to the formation of a small percentage of **17**. Compound **17** is analogous to the heavier molecular weight product formed in the reaction of **3** with **2**.

Although the mass spectra are consistent with the formation of the target cyclic protein **6** in high yield, more experiments are needed to test whether or not this modified protein is suitable for initiating protein folding by flash photolysis. One requirement for the triggering of a refolding event is that the intramolecular crosslink should result in a structure which is incompatible with a native-like folded structure. Thus the synthesis of **6** was carried out on a preparative scale (3 mg) to investigate the structural properties of the cyclic protein.

4.2.3.2 Procedure for the preparation of crosslinked NPGK on a preparative scale

2 and **3** were reacted under the conditions described above on a mg scale. However, in the final dialysis step to remove excess **3** and the denaturing agent, a precipitate formed. The formation of precipitate is due to the removal of denaturant, because

this did not occur when dialysis was carried out in 2 M GuHCl. No precipitate was observed for the analytical preparation of sample **6**, but this is probably due to the lower concentration of protein and shorter dialysis times.[49] To test if the rapid removal of guanidine could overcome the problem, isolation of **6** was attempted by gel filtration. The reaction mixture was concentrated by dialysis in 2 M GuHCl and then applied to a gel filtration column. A precipitate was observed in the solution of the labelled protein eluted from the column. After removal of the first precipitate, a second precipitate was found in the solution stored overnight at 4 ^{o}C. The precipitate was redissolved in 2 M GuHCl and analysed along with the supernatant fraction. The homogeneity of **6** was tested by gel electrophoresis of the soluble protein fraction, because GuHCl interferes with the migration of proteins on the gel. The results are shown in Figure 4.15 and indicate that no oligomers are present in the sample of **6**, which was as homogeneous as the sample of the starting material **2**.

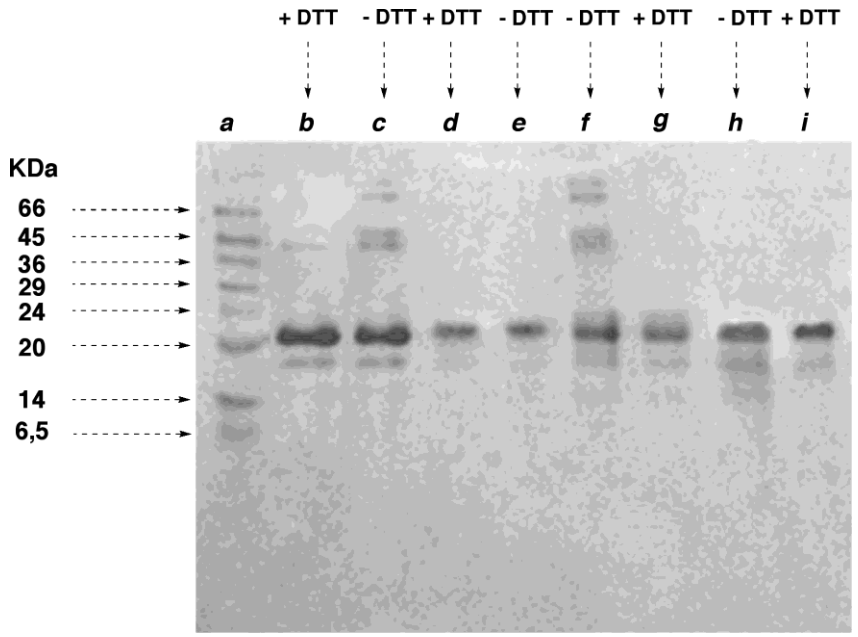


Figure 4.15. SDS-PAGE of **2** and **6** (soluble fraction). Lane *a*: Molecular weights markers. Lane *b* and *c*: **2** (20 μM) loaded with SDS buffer plus and minus DTT. Lane *d* and *e*: **6** (~ 10 μM) loaded with SDS buffer plus and minus DTT. Lane *f* and *g*: **2** (10 μM) loaded with SDS buffer plus and minus DTT. Lane *h* and *i*: **18** (20 μM) loaded with SDS buffer plus and minus DTT. Samples (10 μL) were prepared as 1:1 mixture of SDS loading buffer: protein solution and they were heated at 100^{0} C for 5 minutes prior to loading. Protein bands were stained with a solution of Coomassie blue R.

The results of the gel and the observation that the soluble and precipitate fractions of **6** show a similar behaviour when analysed by ^{1}H NMR spectroscopy and unfolding equilibrium measurements (see next section), suggest that it is the low solubility of the cyclic protein that causes formation of a precipitate in the purification step. This behaviour is expected if the crosslink alters the conformational stability and thus physical properties of the protein.

One way to determine the yield of the crosslinking reaction is from the UV/Vis absorption spectrum. The data of **6** (soluble fraction), **2** and compound **12** are shown in Figure 4.16. The optical density of **6** at 250 nm is increased relative to that of **2**, which correlates well with the fact that the maximum absorption of **12** is centred at 250 nm, so it is not possible to determine the protein concentration by optical density measurements. Thus the concentration of **6** (precipitate and supernatant) was determined with a standard Bradford assay which gave an approximate yield for the labelling reaction of 70 % of which 50 % corresponded to the protein in solution (~ 20 μM).[46, 50]

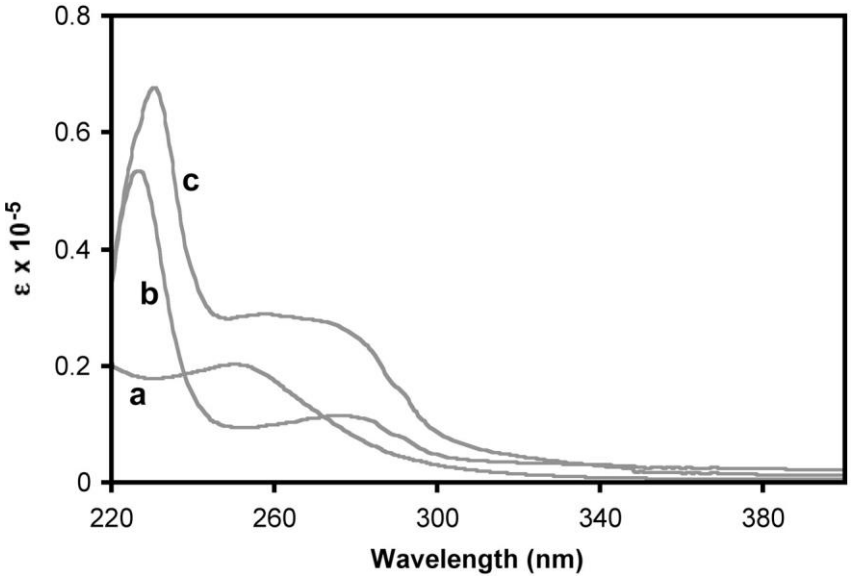
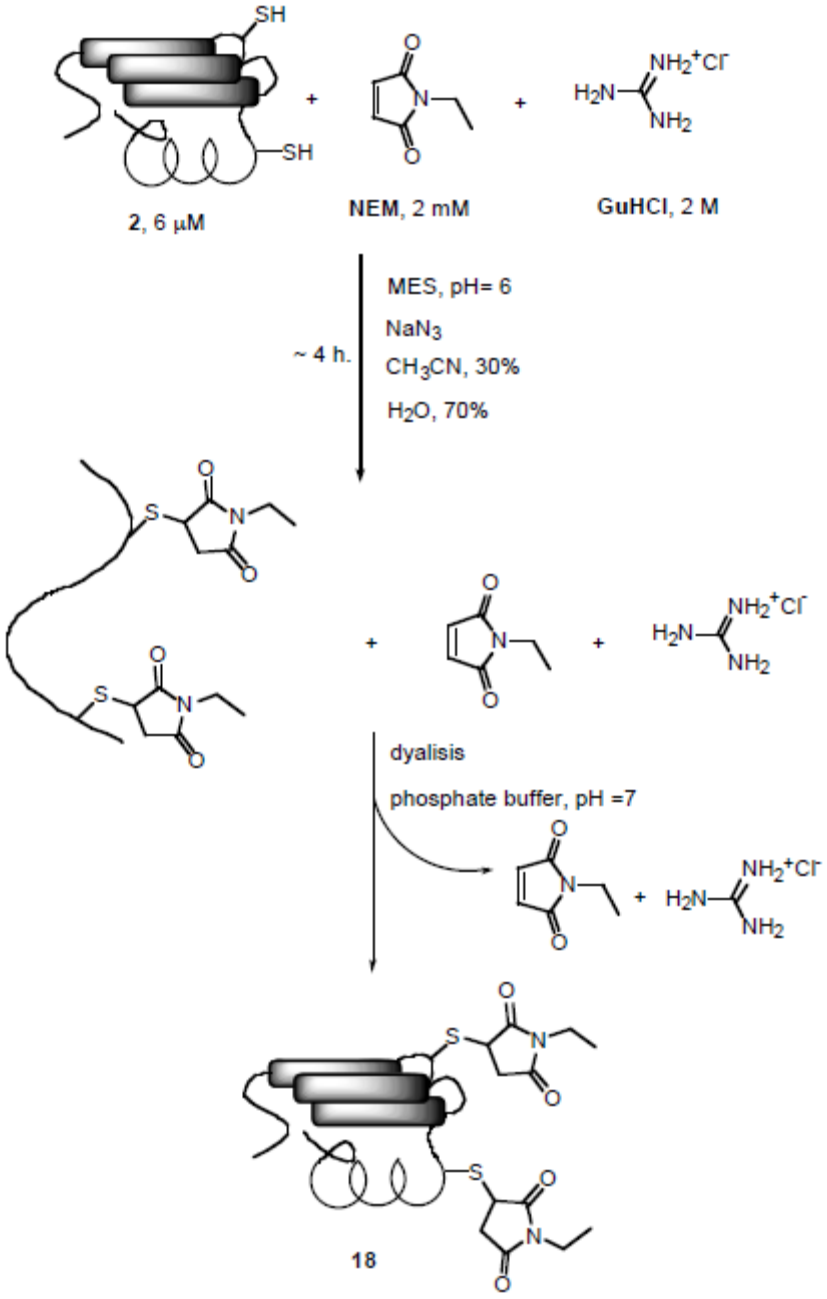


Figure 4.16. Normalised UV/Vis absorption spectra of **12** (a), **2** (b), **6** (c). For **6**, the concentration was calculated with a standard Bradford assay.


SH
SH
2, 6 μM
NEM, 2 mM
GuHCl, 2 M
~ 4 h.
MES, pH= 6
NaN3
CH3CN, 30%
H2O, 70%
dyalisis
phosphate buffer, pH =7
18

Scheme 4.5

4.2.4 Equilibrium unfolding measurements of crosslinked NPGK

The aim of the equilibrium unfolding measurements was to verify that the structure of the cyclic proteins **6** is perturbed compared to the unlabelled mutant and to test if the soluble and precipitate fractions of cyclic protein are homogeneous. To obtain information on the protein structure, we used CD (circular dichroism, far-UV) and fluorescence measurements, which are sensitive to the average amount of α-helix, and β−sheet content, and to the solvent exposure of the aromatic side chains respectively. As a control, a new derivative was synthesised by covalent modification of the protein thiols of **2** with NEM (Scheme 4.5). Compound **18**, like **2**, should fold like the native wild type protein, but has the advantage of not undergoing polymerisation or cyclisation due to the presence of free thiol groups.

The synthesis of **18** was carried out on a preparative scale following the procedure described for the labelling of the compound **6** with NEM (Scheme 4.4). However in the case of **18**, the protein could be refolded to a final concentration of 40 μM without precipitation. Compound **18** showed similar behaviour to both **2** and **6** on a SDS gel (Figure 4.15). The unfolding curves of **6** (soluble and precipitate fractions), compound **18** and **1** are shown in Figure 4.17. For all the unfolding equilibrium measurements, the experimental data were fitted assuming a two-state model, where only the unfolded and folded states are populated at equilibrium (Figure 1.2). This treatment is based on the assumption that the stability of the unfolded state increases linearly with the concentration of denaturant.[51, 52] According to this hypothesis the dependence of the thermodynamic parameters on denaturant is described by:

$$\frac{\Delta G_{(F-U)}}{RT} = \frac{\Delta G_{(F-U)w}}{RT} + mD \qquad \text{Equ. 4.1}$$

where:

$\Delta G_{(F\text{-}U)}$ and $\Delta G_{(F\text{-}U)w}$ are the changes in free energy between the F (folded) and U (unfolded) states at a given denaturant concentration (D) and in water (w) respectively,

R is the molar gas constant,

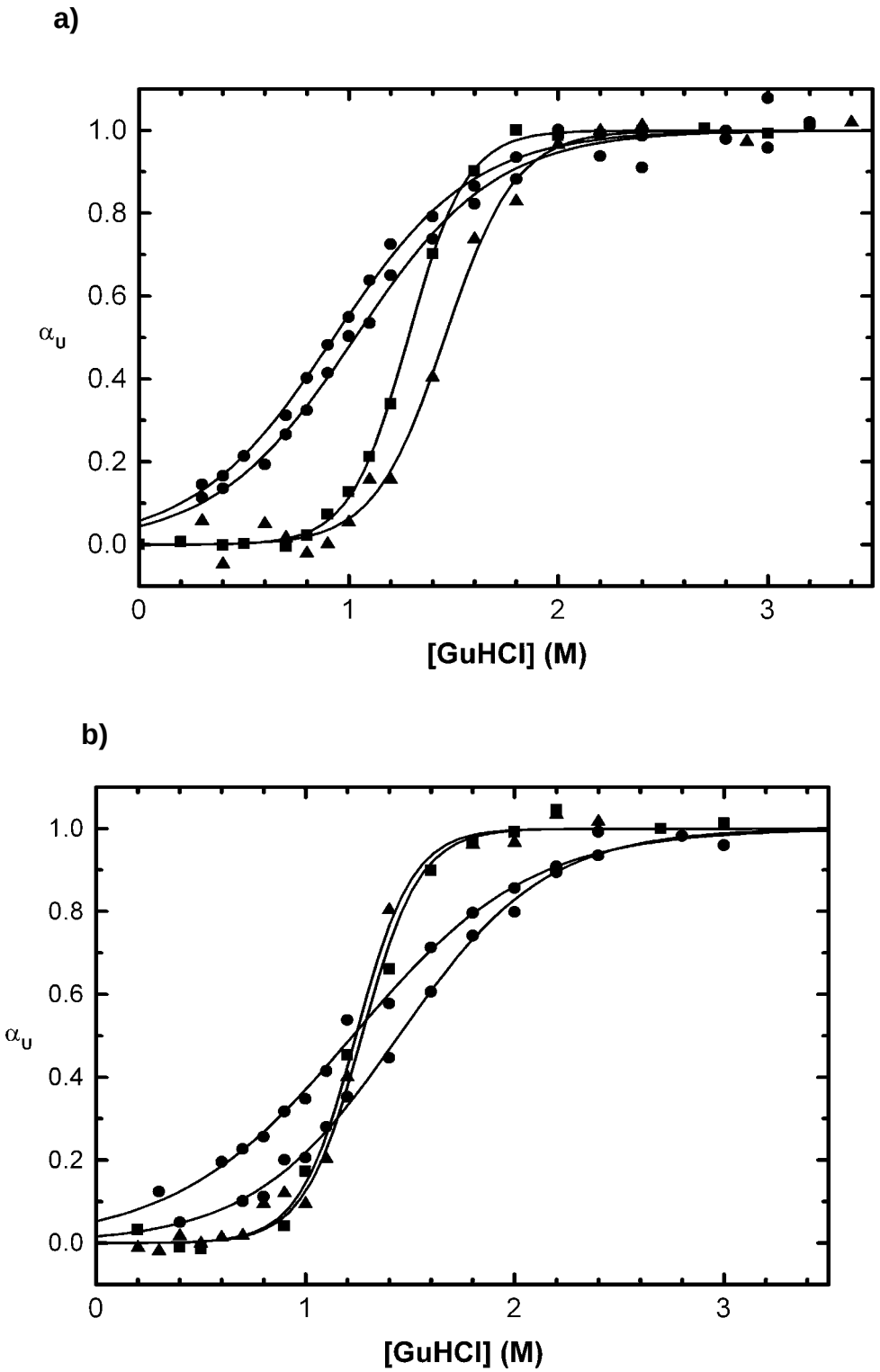


Figure 4.17. Equilibrium unfolding profiles of **1** (square), **18** (triangles), and **6** (circles). Fluorescence data. b) CD data. The mole fraction of the unfolded state (α_U) is plotted as a function of denaturant concentration. The curves were obtained by fitting the data to equation 4.2

T is the absolute temperature,
m describes the sensitivity of the unfolding reaction to the denaturant concentration relative to the folded state ($m_F = 0$). This empirical parameter is related to the changes in the number of denaturant binding sites and thus it is proportional to the increase of protein surface accessible to the solvent between the F and U states.[53, 54] Qualitatively, a steep unfolding transition in a plot of spectroscopic signal *vs*. denaturant concentration is indicative of a high *m* value, and conversely, a shallow unfolding transition is indicative of a low *m* value.[55, 56] For a two-state model, equilibrium unfolding curves are obtained fitting the experimental data to equation 4.2:

$$I = \alpha_F \cdot I_F + \alpha_U \cdot I_U \qquad \text{Equ. 4.2}$$

Where α_F and α_U are the fractional proportion of molecules in the folded and unfolded states respectively and I, I_U, I_F are the intensity of the measured, folded and unfolded signals respectively. The fitting procedure allows calculation of the equilibrium constant for the unfolding reaction in water ($K_{F/U(w)}$) and the *m* value for the unfolded state (m_U), according to the following relationships:

$$\alpha_F = \frac{K_{F/U}}{1 + K_{F/U}} \qquad \text{Equ. 4.3}$$

$$\alpha_U = 1 - \alpha_F \qquad \text{Equ. 4.4}$$

$$K_{F/U} = K_{F/U(W)} \cdot \exp(m_U \cdot x) \qquad \text{Equ. 4.5}$$

where $K_{F/U}$ is the equilibrium constant for the unfolding reaction in a given solvent, and equation 4.5 is derived by combing the Arrenhius equation with equation 4.1.[51]

The unfolding profiles in Figure 4.17 are expressed as intensities relative to the unfolded state and are normalised for comparative purposes. To obtain the unfolding profiles, the absorbance intensity at 222 nm of the CD measurements and the fluorescence emission at 305 nm or at 320-340 nm were used (see Experimental Section). In the case of **1** and **18** a sloping baseline was found on the folded and

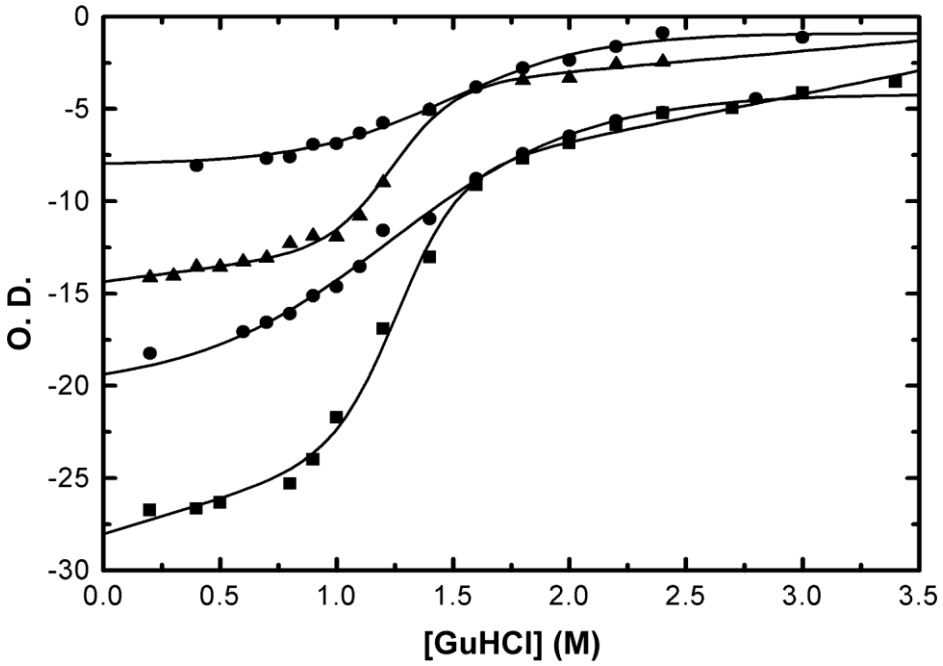


Figure 4.18. Equilibrium unfolding profiles of **1** (square), **18** (triangles), and **6** (circles). obtained by CD measurements. The optical density at 222 nm is plotted as a function of denaturant concentration. The curves were obtained by fitting the data to equation 4.2 and are not normalised

unfolded side of the transition as it is shown in Figure 4.18. The unfolding profiles shown in Figure 4.17 were obtained by subtracting the observed slopes from the data which were then fitted to equation 4.2.[57] The unfolding profile of **1** was compared with **18** to test for possible perturbations of the fold as a result of the maleimide labelling. They are very similar, so the labelled protein conserves a native-like structure. The unfolding profiles of the two cyclic protein fractions are very similar but they are very different from that of **1** and **18**. The unfolding curves are shallower which indicate a less cooperative transition and a non-native structure.[58, 59]

In principle it is possible that the cyclic protein conserves a native/like structure but unfolds populating intermediates (Figure 4.19).[56, 57] A detailed comparison of the structural properties of **6** and **18** is necessary to rule this out.

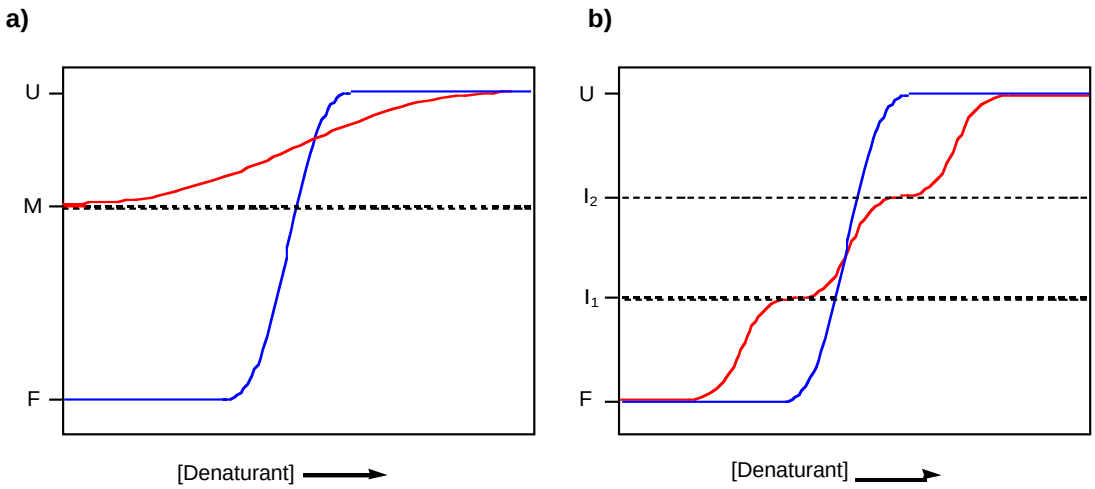


Figure 4.19. Illustration of a two-state folding transition compared with a non-cooperative transition. a) The starting point of the transition is an unstructured protein b) The transition involves intermediate states.

The average amount of secondary structure measured by CD for both proteins is almost the same (Figure 4.20). The coincidence of the CD spectra may be due to interference of the aryl disulfide chromophore or to uncertainty in the protein concentration.[17, 60] However, the maximum in the fluorescence spectrum of **6** is clearly red-shifted relative to that of **18**. This is characteristic of the exposure of tryphtophan to a more polar environment and thus suggests that **6** is partially unfolded.[63] One possibility is that the structure of **6** resembles that of a molten

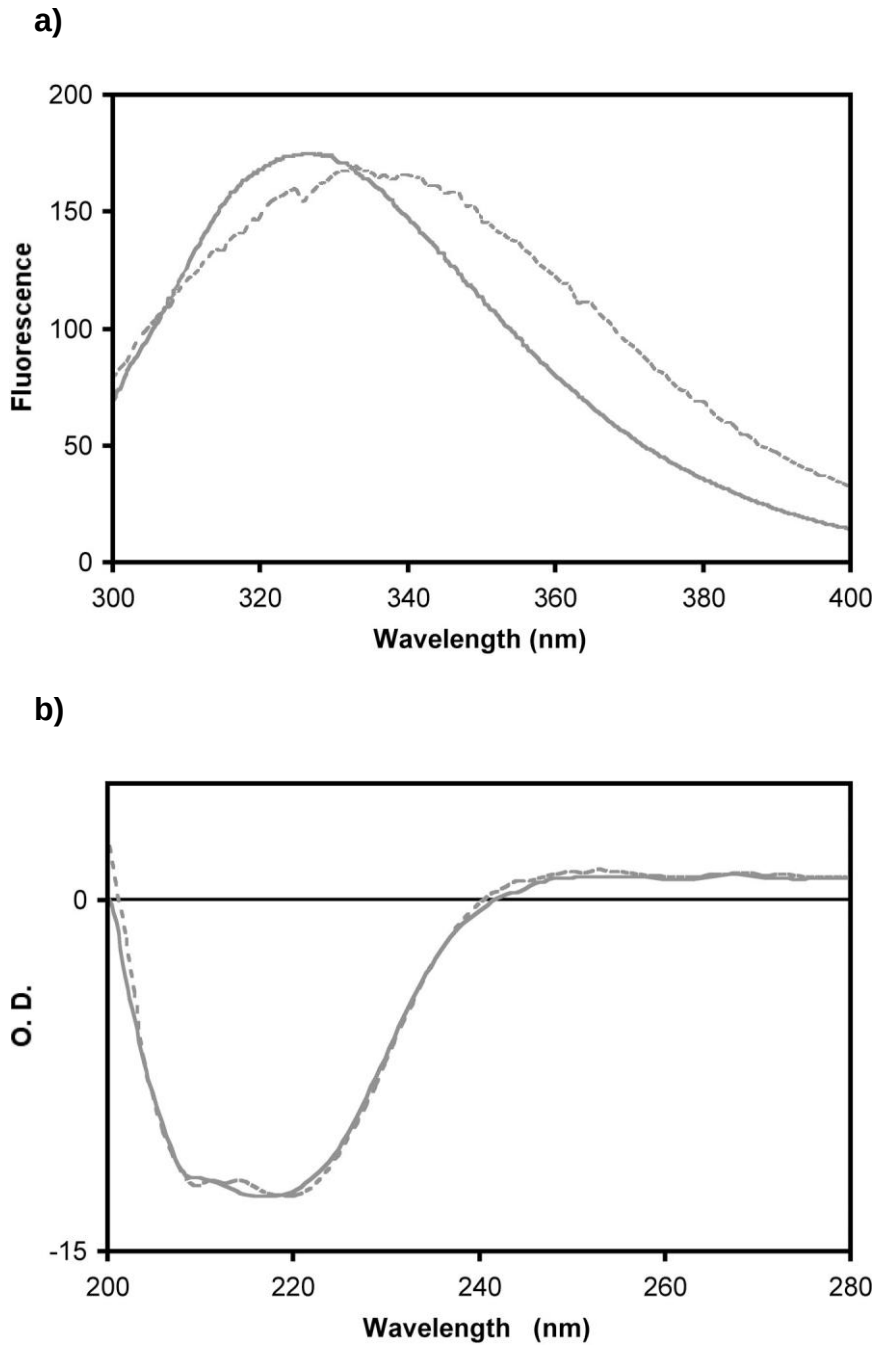


Figure 4.20. Fluorescence (a) and CD (b) spectra of ~~**18**~~ ( ) and **6** (---)**.** 10 μM protein for CD and 5 μM protein for fluorescence, 50 mM sodium phosphate buffer, pH 7

globule, a state rich in secondary structure but more hydrated than the native folded state.[57, 61]

6, **18** and the **2** were also analysed by ^{1}H NMR spectroscopy (Figure 4.21). The signal at 1.23 ppm is due to the threonine methyl groups, and it has collapsed in the spectrum of **6**, which suggests that the crosslink induces a perturbation of the folded state.[56, 62] Thus the spectroscopic data support the idea that the shallow unfolding profile observed in the case of **6** can be attributed to a transition in which the starting point is a protein less structured than the native one.

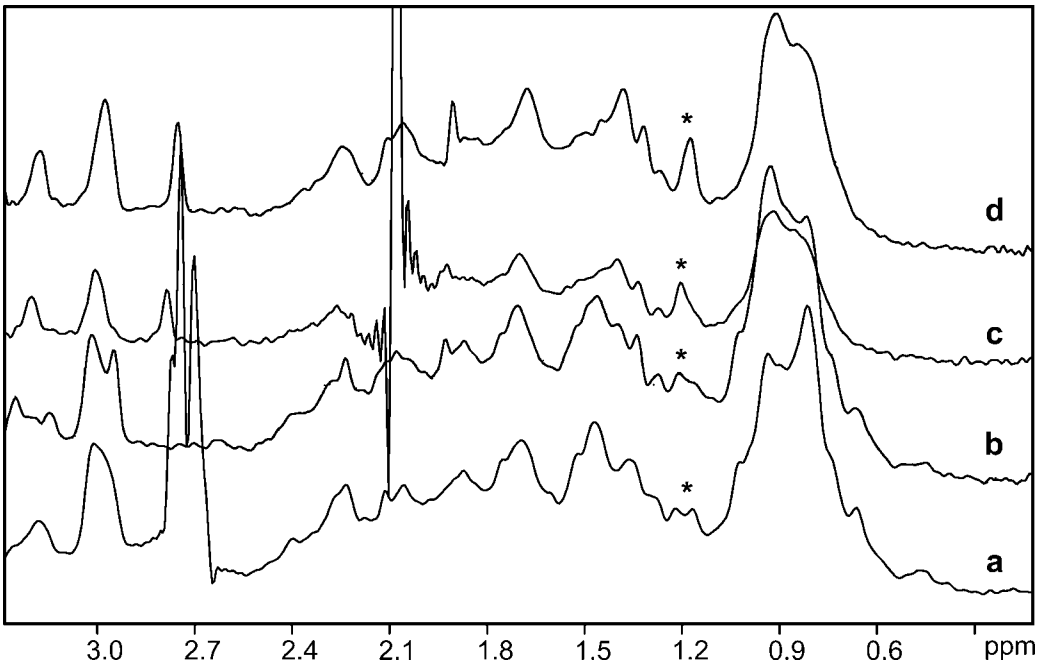


Figure 4.21. Part of the 500 MHz ^{1}H NMR spectra of **2** (a), **18** (b), precipitate fraction of **6** (c), soluble fraction of **6** (d). (20-40) μM **6**, **18** and 300 μM **2**, 50 mM sodium phosphate, 0.6 M GuHCl, 10% D_2O, 800 μM TSP (3- trimethylsilyl-propionate), pH 7. Solution of **2** also contained 8 mM DTT.

4.3 Conclusions

The results obtained so far show that we have set up a procedure which allows crosslinking of a protein with a photoactive moiety. Preliminary data also suggest that this procedure may be used to force a folded protein into an unfolded state. Although the solubility of the modified protein is only 20 μM, this concentration is enough to follow a refolding reaction by time-resolved fluorescence spectroscopy. We are currently waiting for the laser facilities located at Leeds University to be set up in order to carry out the photolysis experiments.

4.4 Experimental Section

4.4.1 Materials

All chemicals were purchased from Sigma Aldrich, except for triethanolamine, sodium dihydrogen orthophosphate and disodium hydrogen phosphate, which were purchased from BDH, and were used without further purification. Acrylamide solutions and TEMED (*N*, *N*, *N'*, *N'*,-tetramethylenediamine) for SDS-PAGE were purchased from Biorad. Ultrafiltration membrane (regenerated cellulose, YM 10) was purchased from Amicon/Millipore. Vivaspin Centrifugal filter devices (MW cut off = 5000, Polyethersulphone membrane, 20 mL capacity) were purchased from Sartorius/Vivascience. Prepacked Sephadex G-25 PD 10 columns (exclusion limit, MW = 5000) were purchased from Amersham Pharmacia.

4.4.2 Instruments

The pH of buffer solutions was measured using a Jenway 3020 pH meter. Protein solutions (< 2 mL) were dried using a speedvac Plus SC110A Savant. UV/Visible spectra were recorded on a Varian Cary 3 Bio spectrophotometer. Fluorescence measurements were carried out with a Varian Cary Eclipse spectrometer. Circular Dichroism measurements were carried out with a Jasco J-810 Spectropolarimeter using a 1mm path length cuvette. ^{1}H NMR spectra were recorded in a 500 MHz Bruker AMX spectrometer equipped with a cryoprobe. MALDI/MS were obtained with a Micromass TOF (time of flight) Spec2E mass spectrometer. ESI-FTICR /MS (electrospray ionisation - fourier transform ion-cyclotron resonance mass spectra) were obtained using a 9.4 T-Bruker Bio-apex II spectrometer and an external electrospray ionisation source (Analitica, Bradford, CT, USA). This instrument has been described in detail elsewhere.[63] All instrument operations and data analysis were performed by Liam McDonnel using the x. mass v 5.0 (Bruker Daltonics Gnbh Bremen Germany) and they were part of the EPSRC FTICR services located at Warwick University. The error in the measurement of the protein mass is between 0.5 - 2 ppm.

4.4.3 General Methods

Protein solutions.
A number of buffer solutions were used to carry out the different experiments on proteins. The composition of these buffer solutions is detailed in the following list:

- TEA (pH 6): 50 mM triethanolamine, 3 mM sodium azide.
- MES (pH 6): 50 mM [2-(N-morpholino) ethane sulphonic acid], 3 mM sodium azide.
- CH_3COOK (pH 5): 20 mM potassium acetate, 3 mM sodium azide.
- TRIS (pH 8): 50 mM trishydroxymethylamine, 3 mM sodium azide.
- PBS 50 mM (pH 7): 0.021 M sodium dihydrogen ortho phosphate, 0.029 M sodium di-sodium hydrogen phosphate, 3 mM sodium azide.
- GuHCl (pH 6): 50 mM TEA, 6M guanidine hydrochloride.
- GuHCl (pH 6): 50 mM MES, 6 M guanidine hydrochloride.
- GuHCl (pH 7): 50 mM PBS, 6 M guanidine hydrochloride.

The final pH was adjusted by addition of aqueous solutions of 1N NaOH or HCl, except for CH_3COOK buffer, which was adjusted with glacial acetic acid. All buffers were filtered through a Y10 membrane prior to use.

The exchange of buffer, or the removal of low molecular weight compounds (MW < 5000) from protein solutions, was carried out by dialysis or gel permeation chromatography, using respectively Vivaspin centrifugal devices and G75 PD 10 columns. Unless otherwise stated, dialysis was carried out by transferring the sample into the centrifugal device, (0.5-3mL), diluting with the desired buffer to 20 mL and reducing the volume by centrifugation to 500 μL. Then the process of dilution and concentration was repeated twice more, to finally obtain 500 μL of sample with the new buffer. Gel permeation chromatography was used in those cases were precipitation of the sample could take place during the dialysis procedure. Typically, a sample of 300-1000 μL, was diluted to 2.5 mL with the desired buffer. Then it was loaded into the column, previously equilibrated with the same buffer, and the first 3

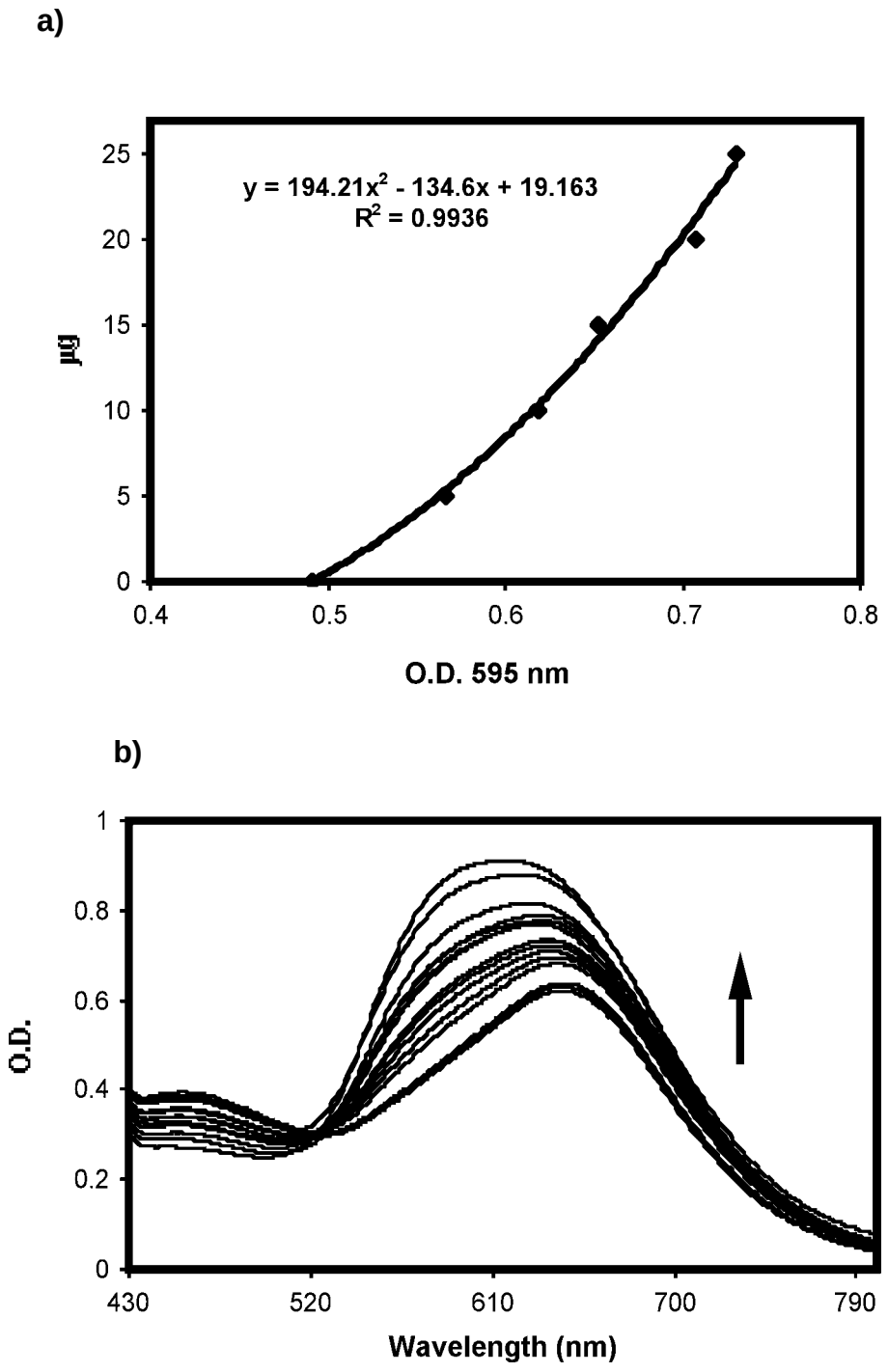


Figure 4.22. a) calibration curve used for the determination of labelled NPGK concentration. b) UV/Vis. spectra of the diluted solution of Coomassie brilliant blue G- 250 upon addition of protein. The OD at 595 increases with an increase in the amount of protein added (0-30 μg in 100 μl of protein solution)).

mL of eluate were collected. These methodologies also allowed the protein concentration to be changed when required.

Estimation of protein concentration was carried out using different methods, depending on the particular protein used. The concentration of NPGK **1** was estimated by the UV absorption of tyrosine residues at 280 nm. An extinction coefficient of 1100 $M^{-1}cm^{-1}$ was taken for each tyrosine residue (ε_{280nm} = 4400 M^{-1} cm^{-1} for **1**) and the molecular weight of the protein was determined by the database entry of the protein amnio acid sequence (19258 Da).[64, 65] The concentrations of protein **2** and NPGK **6** were estimated by standard Bradford essay, using the following protocol:
The dye binding solutions were prepared following the Bradford method (Concentrated dye solution : 98 mg of Comassie brilliant blue G-250 were dissolved in a solution of 25% v/v of ethanol 95% and 50% v/v of concentrated phosphoric acid. This solution was diluted 5 fold in water for the measurements).[50] The calibration curve was obtained from protein **1** solutions, using the following samples: 7.3 μL (5 μg), 14.6 μl (10 μg), 22 μl (15 μg), 29μL (20 μg), 36 μL (25 μg), 44 μl (30 μg) of 56 μM of **1** in PBS (pH 7), that were diluted to 100 μL with PBS (pH 7). 5 mL of diluted dye solution was added to each sample and the optical density at 595 nm was recorded. Protein samples were run in duplicate, and a control sample without the addition of protein was used as reference. The diluted and concentrated dye solutions were stored protected from light at 5 ^{0}C. The optical density at 595 nm of the diluted dye solution added to 100 μL of PBS (pH 7) was constant for at least 6 months. The UV spectra of the dye solution upon addition of increasing amount of protein, and the calibration curve used for the determination of protein concentration are shown in Figure 4.22.

Solutions of protein and reagents used in labelling reactions were deoxygenated using the following method: the sample container, provided with a magnetic stirrer was transferred into a larger flask, subjected to a partial vacuum that was broken with nitrogen. This process was repeated al least 5 times.

SDS-PAGE

The following buffers were prepared for SDS-PAGE:

- Upper buffer (stacking gel): 0.5 M trishydroxymethylamine, 0.4 % (w/v) SDS, pH 6.8.
- Lower buffer (resolving gel): 1.5 M trishydroxymethylamine, 0.4 % (w/v) SDS, pH 6.8.
- Running buffer: 25 mM trishydroxymethylamine, 0.19 M glycine, 0.1 % (w/v) SDS, pH 8.3.
- Loading buffer 1: 50 mM trishydroxymethylamine, 100 mM DTT, 2 % (w/v) SDS , 0.1 % (w/v) bromophenol blue, 10 % (v/v) glycerol.
- Loading buffer 2: 50 mM trishydroxymethylamine, 2 % (w/v) SDS , 0.1 % (w/v) bromophenol blue, 10 % (v/v) glycerol.
- Stain: 0.4 % (w/v) Coomassie blue R, 8 % acetic acid (v/v), 46 % methanol (v/v).
- Destain: 10 % acetic acid (v/v), 30 % (v/v) methanol.
- Resolving gel: 2.5 mL of lower buffer, 4 mL of 40 % (w/v) acrylamide/bisacrylamide (29/1), 3.5 mL distilled water, 100 μL of 10% (w/v) ammonium persulfate, 10 μL of TEMED.
- Stacking gel: 2.5 mL of upper buffer, 1.15 mL of 40 % (w/v) acrylamide/bisacrylamide (29/1), 6.4 mL distilled water, 110 μL of 10% (w/v) ammonium persulfate.

Gels were prepared by pouring in the assembled glass plates of a Bio-Rad mini-protean II apparatus, the 16% polyacrylamide resolving gel solution followed by the 4% polyacrylamide stacking gel. Samples (20-10 μL) were prepared as a 1:1 mixture of SDS loading buffer: protein solution (20-40 μM). Samples were heated at 100^{O} C for 5 minutes prior to loading. Gels were clamped in the electrophoretic chamber for 1.5-2 hours at a voltage of 80 V during stacking and 140 V during resolution. The resulting gels were stained overnight and then destained with shaking at room temperature. The wet gels were placed on a light box and photographed.

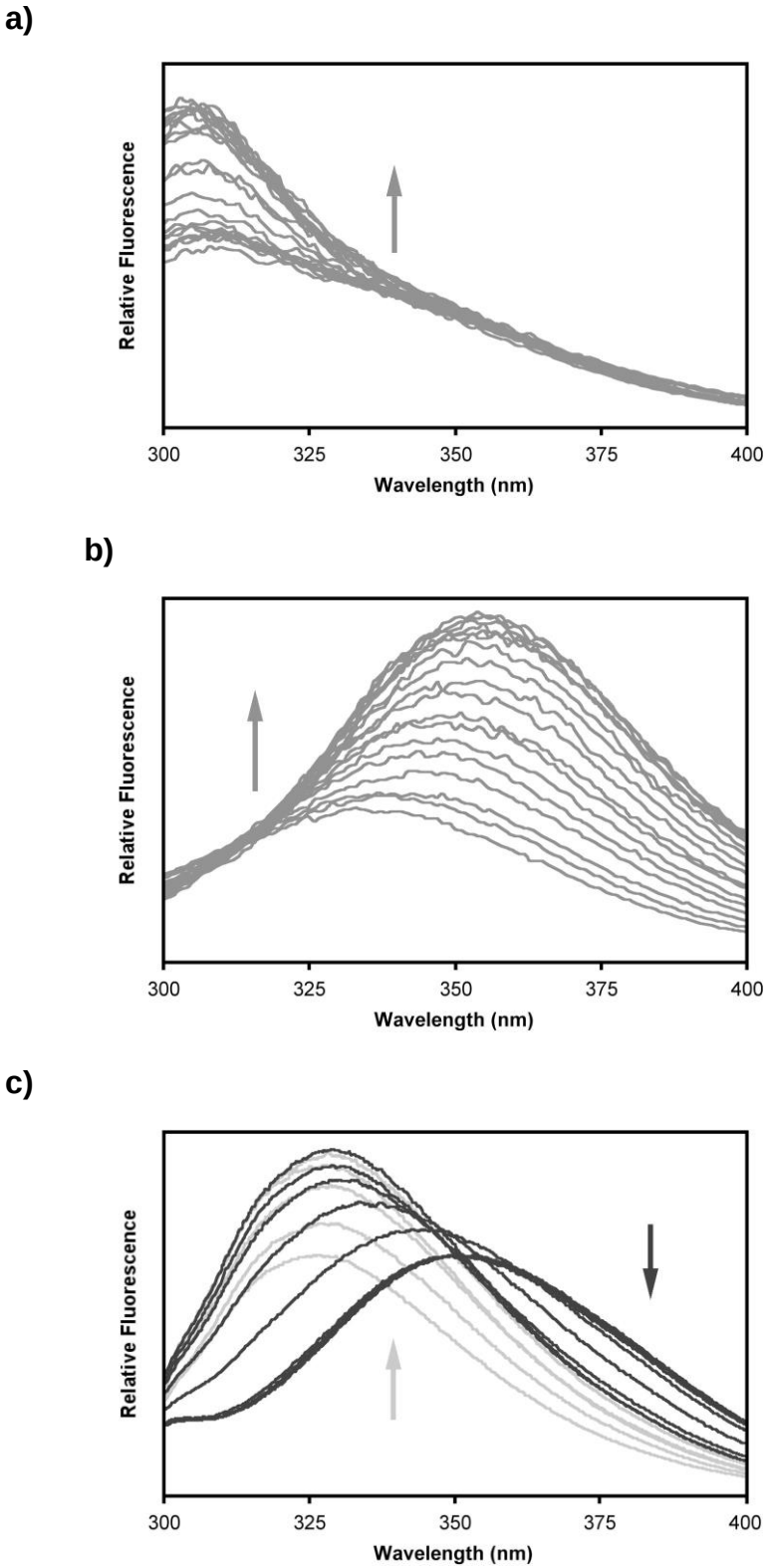


Figure 4.24. Fluorescence emission spectra of **1** (a), **18** (b) and **6** (c). at different denaturant (GuHCl) concentration. The arrows indicate increasing denaturant concentration. For **1** and **6** the fluorescence intensity increases upon unfolding, for **18** the fluorescence intensity increases first and then decreases upon unfolding

Unfolding equilibrium measurements

Unfolding profiles were obtained by recording the Fluorescence or CD spectra of protein solutions with increasing amount of denaturant. A number of protein solutions (between 10 and 20) with the same concentration of protein and a concentration of GuHCl ranging from 0 to 3 M were prepared from stock solutions of protein 20-40 μM and GuHCl 6M, both in PBS (pH 7) buffer. The desired concentration was achieved by diluting the stock solutions in PBS (pH 7) buffer. The solutions were then allowed to re-equlibrate for a minimum of 20 minutes at 25 ^{0}C. For CD experiments, the final concentration of protein used was between 10 and 20 μM, and the spectra were collected over a 300 – 200 nm range using a 2 nm slit width (Figure 4.23). Unfolding profiles were obtained as a plot of absorbance at 222 nm *versus* denaturant concentration. For fluorescence, the concentration of protein was between 4 and 5 μM, and the spectra were obtained by excitation of the tyrosines at 272 nm (NGPK **1**) or tryptophan at 290 nm (proteins **6** and **18**). Emission was collected over a range of 280 – 400 nm using a slit width of 5 nm (Figure 4.24).

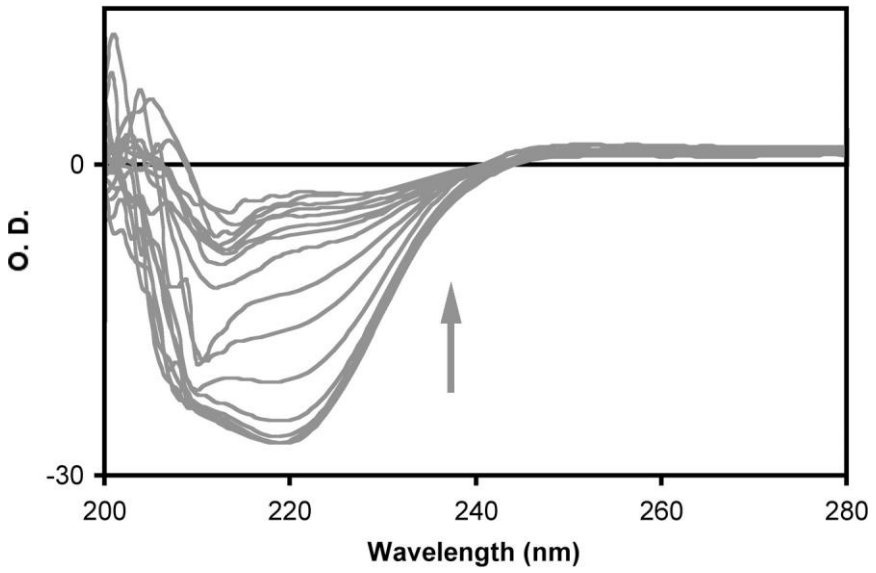


Figure 4.22. Circular dchroism spectra of **1**. Similar spectra were obtained for 5 and 18. The arrow indicates increasing concentration of denaturant. The OD at 222 nm decreases with increasing GuHCl concentration.

Unfolding profiles were obtained as follow:

- for **1** the intensity at 305 nm was plotted *versus* denaturant concentration.
- for **6** and **18** the difference in the fluoresce emission at 340 and 320 nm was plotted *versus* denaturant concentration.

Experimental data were fitted to equation 4.2 using Microcal Origin (version 5.0, Microcal Software).

.

^{1}H NMR spectroscopy experiments

For ^{1}H NMR spectroscopy experiments, protein solutions were prepared as follow :

- 400 μL of 20 μM **6** (soluble fraction) in PBS were added to 50 μl GuHCl 6 M in PBS, 50 μl of D_2O and 4 μl of 100 mM TSP (3- trimethylsilyl-propionate in D_2O), pH 7.
- 250 μL of 40 μM **6** (insoluble fraction) stored in 1M GuHCl, PBS were diluted to 500 μl with 8.3 μL of GuHCl 1M in PBS, 50 μl of D_2O, 4 μl of 100 mM TSP and 188 μL of PBS, pH 7.
- 400 μL of 400 μM **2** stored in DTT 8 mM, PBS were diluted to 500 μLwith 50 μl GuHCl 6 M in PBS, 50 μl of D_2O and 4 μl of 100 mM TSP, pH 7.

Spectra were recorded with 64 scans for **2** and 2048 scans for **6** and **18**. Water suppression was achieved using a presaturation pulse during the relaxation delay between scans. Spectra were processed using FELIX 2000 (MSI, San Diego) and referenced against TSP present in the sample.

4.4.4 Kinetic experiments

Qualitative kinetic experiments were carried out by UV-Vis spectroscopy. Samples were prepared in deoxygenated solutions, using mixtures of TEA (pH 6) buffer and acetonitrile as solvent. The protocol for the different experiments is here detailed:

*Reaction of **1** with DTNB*

In a 3 ml quartz cuvette equipped with a magnetic stirrer were added in the following order:

800 μL of 6M GuHCl in TEA (pH 6)

750 μL of distilled water

10 μl of 1M solution of sodium azide

800 μl of acetonitrile

The cuvette was transferred into a 50 mL flask and the solution was deoxygenated. To the solution were added, in this order:

10 μL of 6 mM solution of DTNB in acetonitrile (freshly prepared and deoxygenated)

40 μL of **1** 320 μM in TEA, (pH 6).

The solution was gently shaken and UV spectra at 250-550nm were recorded immediately and after intervals of 2, 7, 12 and 30 minutes between each scan. A control sample without the addition of protein and DTNB was used as reference (Figure 4.4).

*Reaction of **1** with **3***

Working solutions were prepared as described for the reaction of **1** with DTNB, with the following difference: 10 μL of 2.5 mM solution of **3** in acetonitrile were added to each sample instead of DTNB. After intervals of 30 seconds, 8 minutes and 15 minutes, 10 μl of the DTNB stock solution (6 mM, pH 6) were added to each protein solution and the UV spectra were recorded immediately and after intervals of 2, 7 and 12 minutes between each scan(Figure 4.5).

4.4.5 Preparation of an analytical sample of 7 (~ 0.3 mg)

Protein **7** was prepared from protein **1** and disulfide **3**. The reaction mixture was prepared following essentially the same procedure and quantities described above for the kinetic experiment of **1** with **3**, using a 10 mL vial instead of a UV cuvette. The vial was transferred in a 100 ml flask and deoxygenated. The reaction mixture was left under gentle stirring for 60 minutes, protected from light, and under nitrogen. Then it was transferred into a centrifugal device and dialysed. 5 μL of 97% acid formic solution were added over the final 500 μL of solution and the sample was submitted to mass spectroscopy.

As a control, a sample of pure **1** was prepared for mass spectroscopy analyses. 10 μL of 0.32 mM **1** TEA (pH 6) were diluted to 20 mL with distilled water in a centrifugal

filter device, and the sample was dialysed. The resulting sample (500 μL) was lyophilised. The dry sample was redissolved in 200 μL of a 1:1 mixture of acetonitrile: 4% formic acid and analysed by mass spectroscopy (ESI –FTICR).

Expected for **1** (1 - 174 amino acids, $C_{853}H_{1380}N_{246}O_{253}S_4$): 19258. Found (calculated as averaged of the 15^+ and 16^+ charge state): 19258.
Expected for **7** (1- 174 amino acids, $C_{873}H_{1393}N_{248}O_{257}S_6$): 19666. Found (calculated as averaged of the 15^+): m/z = 19666 (MH^+), 19684 ($MH^+ + H_2O$), 19258 (**1**), 18888 (1-168 of MH^+), 18906 [1 – 168 of ($MH^+ + H_2O$)], 18216 (1-161 of MH^+), 18234 [1 – 161 of ($MH^+ + H_2O$)].

4.4.6 Preparation of an analytical sample of 8 (~ 0.6 mg)

The reaction mixture of **1** and **3** (2.4 mL) was prepared as described for the preparation of **7**. After 90 minutes the reaction mixture was transferred to a centrifugal filter device and diluted to 20 ml with a fresh solution of 1mM DDT in GuHCl 2M, 50 mM TEA, 3 mM sodium azide (pH 6). The solution volume was reduced by centrifugation to 800 μL, transferred to a 2 mL eppendorf vial equipped with a magnetic stirrer and incubated for 1 hour. The reaction mixture was then transferred into a centrifugal filter device and dialysed. The resulting sample (500 μL) was lyophilised. The dry sample was redissolved in 200 μL of a 1:1 mixture of acetonitrile- 4% formic acid v/v and analysed by mass spectroscopy (ESI –FTICR).

Expected for **8** ($C_{863}H_{1387}N_{247}O_{255}S_5$): 19643. Found (calculated as averaged of the 16^+): 19643.

4.4.7 Procedure for the removal of DTT from stock solution of 2

Protein **2** (10-12 mg/ml) was prepared by Clare Jelinska and was stored as an ammonium sulphate suspension in 10 mM DTT, 50 mM PBS, 3mM sodium azide (pH 7). Since DTT may undergo oxidation with time, the DTT buffer was exchanged for a fresh stock solution (0.02 M DTT, 50 mM TRIS, 3mM sodium azide, pH 8) to minimise the possibility of polymerisation of protein prior to labelling reactions and

mass spectroscopy experiments. The change of buffer was carried out by dialysis and the concentration of the new solutions was estimated by Bradford assay. Also, most of the experiments are incompatible with the presence of DTT in solution. In these cases the DTT was removed by exclusion size chromatography. The protocol of exchange of buffer and removal of DTT is detailed as follows:

For the exchange of buffer, 100 μL of a stock solution of **2** were transferred in a 1 mL eppendorf vial and centrifuged at 13000 rpm for 20 minutes. The supernatant was discarded. The solid was redisolved in 1 mL of the fresh DTT stock, transferred into a centrifugal filter device and dialysed in the fresh DTT stock solution. The final 500 μL solution were transferred to a 1 mL eppendorf vial and incubated overnight. After that time, the concentration of protein in the final solution was determined by a Bradford assay, using two 10 μL aliquots of solution diluted ten fold with PBS buffer (pH 7). The estimated concentration was 600 μM (OD_{595} = 0.72) (Figure 4.21).

The elimination of DTT was carried out using size exclusion chromatography. Typically, 500 μL of fresh stock solution of **2** were diluted to 2.5 ml with CH_3COOK buffer and then applied to an PD10 column previously re-equilibrated in the same buffer. The first 3 mL were collected. Then the concentration was measured by measurement of the OD at 280 nm. The number of free thiol groups was estimated by the Ellman method, using equal volumes of protein solution and a fresh solution containing 100 μM DTNB, 4 M GuHCl, 3 mM sodium azide, 50 mM PBS (pH 7) previously deoxygenated.

4.4.8 Procedure for the preparation of an analytical sample of 2

An analytical sample of **2** was prepared by removing DTT from 200 μL of **2** stock solution in fresh DTT. In this case, the eluent used was 0.1 % acetic acid. Seven fractions, of 500 μL each, were collected. The sixth fraction (30 μM) was submitted to mass spectroscopy analyses (MALDI-TOF).

Expected for **2** (1-174 amino acids, $C_{852}H_{1376}N_{244}O_{251}S_5$): 19214. Found: m/z = 19214 (MH^+), 19085 (MH^+ - N-terminus methionine), 9607 (MH^+ / 2), 38424 (2 MH^+).

4.4.9 Procedure for the preparation of an analytical sample of crosslinked NPGK 6(~0.3 mg)

In a 2.4 ml vial equipped with a magnetic stirrer were added in the following order:

800 μL of 6 M GuHCl in MES (pH 6)

188 μL of distilled water

2 μL of sodium azide 1 M

760 μL of acetonitrile

96 μL of 2.5 mM solution of **3** in acetonitrile

The vial was transferred to a 50 mL flask and the solution was deoxygenated. 600 μL of a solution 24 μM of **2** in CH_3COOK buffer (pH 5) were then added under nitrogen. The pH of the mixture was 6, as estimated by measuring the pH on a control sample without protein. The reaction mixture was stirred gently during 3 hours under nitrogen and protected from light. Then the solution was transferred into a centrifugal filter device and dialysed, using distilled water, to yield 500 μL sample. 10 μL of the sample were used for SDS-PAGE (Figure 4.9). The remaining 490 μL were lyophilised and then redissolved in 200 μL of a 1:1 mixture of acetonitrile/4% formic acid v/v for mass spectroscopy analysis (ESI –FTICR).

Calculated for **6** ($C_{872}H_{1390}N_{246}O_{255}S_7$): 19622. Found: m/z = 19622 (MH^+), 19645 (MNa^+), 19668 ($M2Na^+$), 19691 ($M3Na^+$), 19714 ($M4Na^+$), 19491 (MH^+ - Methionine), 19514 (MNa^+ - Methionine), 19418 (MH^+ - 204), 19441 (MNa^+ - 204), 19826 (MH^+ + 204), 19849 (MNa^+ + 204).

4.4.10 Procedure for the reaction of an analytical sample of crosslinked NPGK 6 with NEM

The reaction mixture of **2** and **3** (2.4 mL) was prepared as described for the preparation of an analytical sample of **6**. After 3 hours the reaction mixture was transferred in a centrifugal filter device and diluted to 20 mL with a fresh solution of 1mM DTT, 2 M GuHCl 50 mM MES and 3 mM sodium azide (pH 6). The solution was centrifuged to a final volume of 700 μl. This solution was then transferred to a 1

mL eppendorf vial equipped with a magnetic stirrer and incubated for 1 hour protected from light. The solution was then transferred into a centrifugal filter device and diluted to 20 mL with a fresh solution of 2 mM NEM, 2 M GuHCl, 50 mM MES and 3 mM sodium azide (pH 6). The volume was reduced by centrifugation to 800 μL and the resulting sample transferred into a 2 mL eppendorf vial equipped with a magnetic stirrer. After 4 hours of stirring, protected from light, the reaction mixture was transferred in a centrifugal filter device and dialysed in distilled water. The resulting sample (500 μL) was lyophilised and then redissolved in 200 μL of a 1:1 mixture of acetonitrile/4% formic acid v/v for the mass spectroscopy analysis (ESI – FTICR).
Expected for labelling of **6** ($C_{884}H_{1406}N_{248}O_{259}S_7$): 19874. Found: 19874 ($MH^+$), 19897 ($MNa^+$), 19990 ($M2Na^+$), 19743 ($MH^+$ - Methionine), 19669 (MH^+ - 204).

4.4.11 Procedure for the preparation of crosslinked NPGK 6 on a preparative scale (~ 3 mg)

For the synthesis of 3 mg of **6**, the procedure used for the preparation of an analytical sample was scaled up nearly 10 times:
in a 50 ml flask equipped with a magnetic stirrer were added in the following order:
9 mL of 6 M GuHCl in 50 mM MES (pH 6)
2.2 mL of distilled water
60 μL of 1M sodium azide
8 mL of acetonitrile
1.1 mL of 2.5 mM solution of **3** in acetonitrile
The solution was deoxygenated as described in methods. To the resulting solution, and under nitrogen, 6.8 mL of a solution 22 μM of **2** in CH_3COOK buffer, pH 5, were added. This solution was obtained by removal of DTT from stock solution of **2**, using the procedure described above. In practice, three aliquots of 350 μL of stock solution **2** (180 μM in 20 mM DTT, 50 mM TRIS, 3 mM sodium azide, (pH 8)) were loaded on a PD10 column. The first 3 mL of eluate were taken in each case, and the concentration calculated by measuring their optical density. The solutions were added to the reaction mixture as soon as they were obtained from the column, 3 mL from the first column, 3 mL from the second and 0.8 mL from the third, to a total of

6.8 mL. The reaction mixture was stirred gently under nitrogen and protected from light for 3 hours. The mixture was then split in 5 portions of 5.4 mL, each transferred into a centrifugal filter device and diluted to 20 ml with a solution 2 M GuHCl, 50 mM PBS, 3 mM sodium azide (pH 7). The solutions were centrifuged to a final volume of 1 mL each and then all of them were transferred into a single centrifugal filter device. The resulting solution (5 mL) was centrifuged to a final volume of 2.5 mL and loaded on a PD10 column previously reequilibrated with PBS buffer (pH 7)*. In the first 3 mL of eluate appeared a precipitate. The solid was separated by centrifugation and the 2.9 mL of supernatant were stored at 4 ^{0}C overnight. After this time a second precipitate appeared and was also separated. All precipitate fractions were pooled and dissolved in 800 μL of 2 M GuHCl in PBS 50 mM, 3 mM sodium azide (pH 7).

To estimate the yield of the reaction, the resulting solutions (supernatant and precipitated fractions) were subjected to a Bradford assay. 20 μL of the supernatant solution and 20 μL of the precipitated solution, the latter diluted ten fold with PBS buffer (pH 7), were used to carry out the assay. For the supernatant solution, the optical density observed at 595 nm was 0.6 and, for the precipitated fraction, was 0.54. Comparison with the calibration curve (Figure 4.21) allowed us to estimate the concentration of both solutions, being 20 μM for supernatant fraction (2.6 mL) and 80 μM for the precipitated fraction (800 μL). A 70 % yield was then estimated considering a molecular weight of 19622 for the product.

The homogeneity of the sample was tested by SDS- PAGE, performed over 5 μL of supernatant solution (Figure 4.15). These solutions were then used for unfolding profile measurements and ^{1}H-NMR experiments as described in general methods.

* The removal of the low molecular weight reactants could not be made by the normal dialysis procedure due to problems of solubility (20-30 μM maximum solubility, as found after the experiment) of the proteins in absence of GuHCl. This problem could be solved by not reducing so drastically the volume during dialysis, but then the cycles required to obtain the desired samples would make it unpractical. For this reason the methodology was changed in favour of size exclusion chromatography.

4.4.12 Procedure for the preparation of crosslinked NPGK 18 on a preparative scale (~ 3 mg)

Compound **18** was prepared following the same procedure described for the preparative preparation of **6**, adding 100 μL of 48 mM NEM in 50 mM MES, 3 mM sodium azide (pH 6) to the reaction mixture (24 mL) instead of disulfide **3.** The reaction mixture was stirred gently under nitrogen for 4 hours. After gel filtration on a PD10 column, 3.2 ml of compound **18** in 50 mM PBS (pH 7) were collected. The concentration of this sample was 38.8 μM (97% yield) as calculated by OD at 280 nm (0.40) using the extinction coefficient of **2** (ε_{280} = 10400).

The homogeneity of the sample was tested by SDS-PAGE (Figure 4.15). This solution was then used for unfolding equilibrium measurements.

4.5 References

1. C. B. Anfisen, *Science* **1973**, *181*, 223-230.

2. K. A. Dill, *Biochemistry* **1990**, *29*, 7134-7154.

3. W. A. Eaton, V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, J. Hofrichter, *Annu. Rev. Biophys. Biomol. Struct*. **2000**, *29*, 327-359.

4. D. Baker, *Nature* **2000**, *405*, 39-42.

5. G. L. G. Miklos, G. M. Rubin, *Cell* **1996**, *86*, 521-529.

6. D. K. Pettit, W. R. Gombotz, *Trends in Biotechnology* **1998**, *16*, 343-349.

7. L. P. Encell, D. M. Landis, L. A. Loeb, Nature Biotechnology **1999**, *17*, 143-147.

8. K. Kuwajima, H. Yamaya, S. Miwa, S. Sugai, T. Nagamura, *Febs Lett*. **1987**, *221*, 115-118.

9. H. Roder, G. A. Elove, S. W. Englander, *Nature* **1988**, *335*, 700-704.

10. A. F. Chaffotte, C. Cadieux, Y. Guillou, M. E. Goldberg, *Biochemistry* **1992**, *31*, 4303-4308.

11. C. R. Matthews, *Ann. Rev. Biochem.* **1993**, *62*, 653-683.

12. A. Matouschek, L. Serrano, E. M. Meiering, M. Bycroft, A. R. Fersht, *J. Mol. Biol.* **1998**, *277*, 973-983.

13. C. M. Jones, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11860-11864.

14. T. Pascher, J. P. Chesick, J. R. Winkler, H. B. Gray, *Science* **1996**, *271*, 1558-1560.

15. W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5897-5899.

16. J. Sabelko, J. Ervin, M. Gruebele, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6031-6036.

17. H. S. M. Lu, M. Volk, Y. Kholodenko, E. A. Gooding, R. M. Hochstrasser, W. F. DeGrado, *J. Am. Chem. Soc.* **1997**, *119*, 7173-7180.

18. K. C. Hansen, R. S. Rock, R. W. Larsen, S I. Chan, *J. Am. Chem. Soc.* **2000**, *122*, 11567-11568.

19. D. Mendel, *Annu. Rev. Biophys. Biomol. Struct*. **1995**, *24*, 435-462.

20. A. E. Baker, *Curr. Opin. Struct. Biol*. **1999**, *9*, 189-196.

21 V. Munoz, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11311-11316.

22. J. A. Camarero, J. Pavel, T. W. Muir, *Angew. Chem. Int. Ed.* **1998**, *37*, 347 –349.

23. C. P. Scott, E. A. santos, M. Wall, D. C. Wahnon, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638-13643.

24. V. P. Grantcharova, D. Baker, *J. Mol. Biol.* **2001**, *306*, 555-563.

25. G. J. Davies, S. J. Gamblin, J. A. Littlechild, H. C. Watson, *Proteins: Struct. Funct. Genet*. **1993**, *15*, 283-289.

26. L. L. P. Hosszu , M. J. Parker, J. Spencer, M. Kelly, C. J. Craven, J. P. Waltho, A. R. Clarke, *Biochemistry* **1997**, *36*, 333-340.

27. C. Jelinska, *Equilibrium and kinetic studies of folding and unfolding of the N-terminal domain of phosphoglycerate kinase* **2002**, PhD thesis.

28. L. L. P. Hosszu, C.J. Craven, J. Spencer, M. J. Parker, M. Lorch, A. R. Clarke, *Nature Structural Biology* **1997**, *4*, 801-804.

29. G. Illuminati, L. Mandolini, *Accounts of Chemical Reasearch* **1981**, *14*, 102-109.

30. T. E. Creighton, N. J. Darby, *J. Mol. Biol.* **1993**, *232*, 873-896.

31. T. E. Creighton, *Proteins: structure and molecular properties*, 2nd Edition **1993**, W. H. Freeman and Company, 179-180.

32. C. F. Brewer, J. P. Riehm, Analytical Biochemistry 1967, 18, 248-255.

33. G. E. Means, R. E. Feeney, *Chemical Modifications of Proteins* **1971**, 110-112, Holden – Day, San Francisco, California.

34. G. Gorin, P.A. Martic, G. Doughty, *Arch. of Bioch. and Biophys* **1966**, *115*, 593-597.

35. E. J. Bures, J. O. Hui, Y. Young, D. T. Chow, V. Katta, M. F. Rohde, L. Zeni, R. D. Rosenfeld, K. L. Stark, M. Haniu, *Biochemistry* **1998**, *37*, 12172-12177.

36. M. Hashimoto, E. Majima, S. Goto, Y. Shinoara, H. Terada, *Biochemistry* **1999**, *38*, 1050-1056.

37. H. J. Gruber, G. Kada, B. Pragl, C. Riener, C. D. Hahn, G. S. Harms, W. Ahrer, T. G. Dax, K. Hohenthanner, H. G. Knaus, *Bioconjugate Chem*. **2000**, *11*, 161-166.

38. G. E. Means, R. E. Feeney, *Chemical Modifications of Proteins* **1971**, 156-157, Holden – Day, San Francisco, California.

39. T. Ozawa, A. Hanaki, *Chem. Pharm. Bull*. **1981**, *29*, 1101-1105.

40.J. D. Gregory, *J.O.C.***1955**, *77*, 3922-3923.

41. D. G. Smith, H. Tuppy, *Bioch. et Bioph. Acta* **1968**, *168*, 173-180.

42. E. Majima, S. Goto, H. Hori, Y. Shinoara, Y. Hong, H. Terada, *Bioch. et Bioph. Acta* **1995**, *1243*, 336-342.

43. V. De Pinto, J. A. Al Jamal, R. Benz, G. Genchi, F. Palmieri, *Eur. J. Biochem*. **1991**, *202*, 903-911.

44. M. A. Harrison, J. Murray, B. Powell, Y. Kim, M. E. Finbow, J. B. C. Findlay, *The Journal of Biological chemistry* **1999**, *274*, 25461-25470.

45. G. E. Means, R. E. Feeney, *Chemical Modifications of Proteins* **1971**, 150-152, Holden – Day, San Francisco, California.

46. A. Persechini, R. H. Kretsinger, *The Journal of Biological chemistry* **1988**, *263*, 12175-12178.

47. T. Y. Yen, R. K. Joshi, H. Yan, N. O. L. Seto, M. M. Palcic, B. A. Macher, *Journal of Mass spectrometry* **2000**, *35*, 990-1002.

48.E. Tatsumi, D. Yoshimatsu, M. Hirose, *Biochemistry* **1998**, *37*, 12351-12359.

49 R. Jhenicke, R. Seckler, *Advances in Protein Chemistry* **1997**, *50*, Chapter 1.

50. M. M. Bradford, *Anal. Biochem*. **1976**, *72*, 248-254.

51. M. J. Parker, J. Spencer, A. R. Clarke, *J. Mol. Biol*. **1995**, *253*, 771-786.

52. T. R. Sosnick, L. Mayne, S. W. Englander, *Proteins: Struct. Funct. Genet*. **1996**, *24*, 413-426.

53. C. N. Pace, *Rev. Bioch*. **1975**, *3*, 1-43.

54. M. Oliveberg, *Acc. Chem. Res.* **1998**, *31*, 765-772.

55. R. A. Staniforth, S. G. Burston, C. J. Smith, G. S. Jackson, I. G. Badcoe, T. Atkinson, J. J. Holbrook, A. R. Clarke, *Biochemistry* **1993**, *32*, 3842-3851.

56. M. J. Pandya, P. B. Williams, C. E. Dempsey, P. R. Shewry, A. R. Clarke, *The Journal of Biological chemistry* **1999**, *274*, 26828-26837.

57. R. A. Staniforth, M. G. Bigotti, F. Cutruzzola', C.T. Allocatelli, M. Brynori, *J. Mol. Biol*. **1998**, *275*, 133-148.

58. T. E. Creighton, *Proteins: structure and molecular properties*, 2nd Edition **1993**, W. H. Freeman and Company, 289-291.

59. T. Okuno, S. Hirota, O. Yamauchi, *Biochemistry* **2000**, *39*, 7538-7545.

60. A. Chakrabarty, T. Kortemme, S. Padmanabham, R. L. Baldwin, *Biochemistry* **1993**, *32*, 3842-3851.

61. B. Nolting, R. Golbik, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10668-10672.

62. K. Wuthrich, *NMR of proteins and nucleic acids* **1986**, Wiley, 16-19.

63. M. Palmblad, K. Hakansson, P. Hakansson, X. Feng, H. J. Cooper, A. E. Giuannakopulus, P. S. Green, P. J. Derrick, *Eur. J. Mass. Spectr*. **2000**, *6*, 267-275.

64. L. L. P. Hosszu , M. J. Parker, G. S. Jackson, S. G. Burston, J. Spencer, C. J. Craven, J. P. Waltho, A. R. Clarke, *Biochemistry* **1996**, *35*, 15740-15752.

65. http: // *ca. expasy. org./cgi.-bin/protparam* (entry code for bsPGK: P18912).

Chapter 5

5. Preparation of Photocaged NPGK

5.1 Introduction

It has been shown that the chemical modification of a single amino acid residue may induce structural perturbation of a native protein. An example is creatine kinase, which can be partially unfolded by labelling a single thiol with 5,5'-dithiobis (2-dinitrobenzoic) acid (DTNB) but refolds upon removal of the chemical modification by addition of dithiothreitol.[1, 2] Single amino acid modification has been used to study the refolding of cytrochrome c by labelling of a single methionine residue with a photocleavable *o*-nitrobenzyl group. Photochemical removal of the modification initiated the refolding reaction on a time scale of microseconds.[3]

To further extend this methodology to the triggering of folding reaction on the nanosecond time scale, we have functionalised the N-terminal domain of phosphoglycerate kinase (**1**, NPGK) using a disulfide that can be photocleaved in picoseconds. Since NPGK contains a single cysteine that is buried in the hydrophobic core of the domain, this system appeared ideal for the development of an approach based on single amino acid modification to introduce an optical trigger for the study of folding (Figure 5.1).[4]

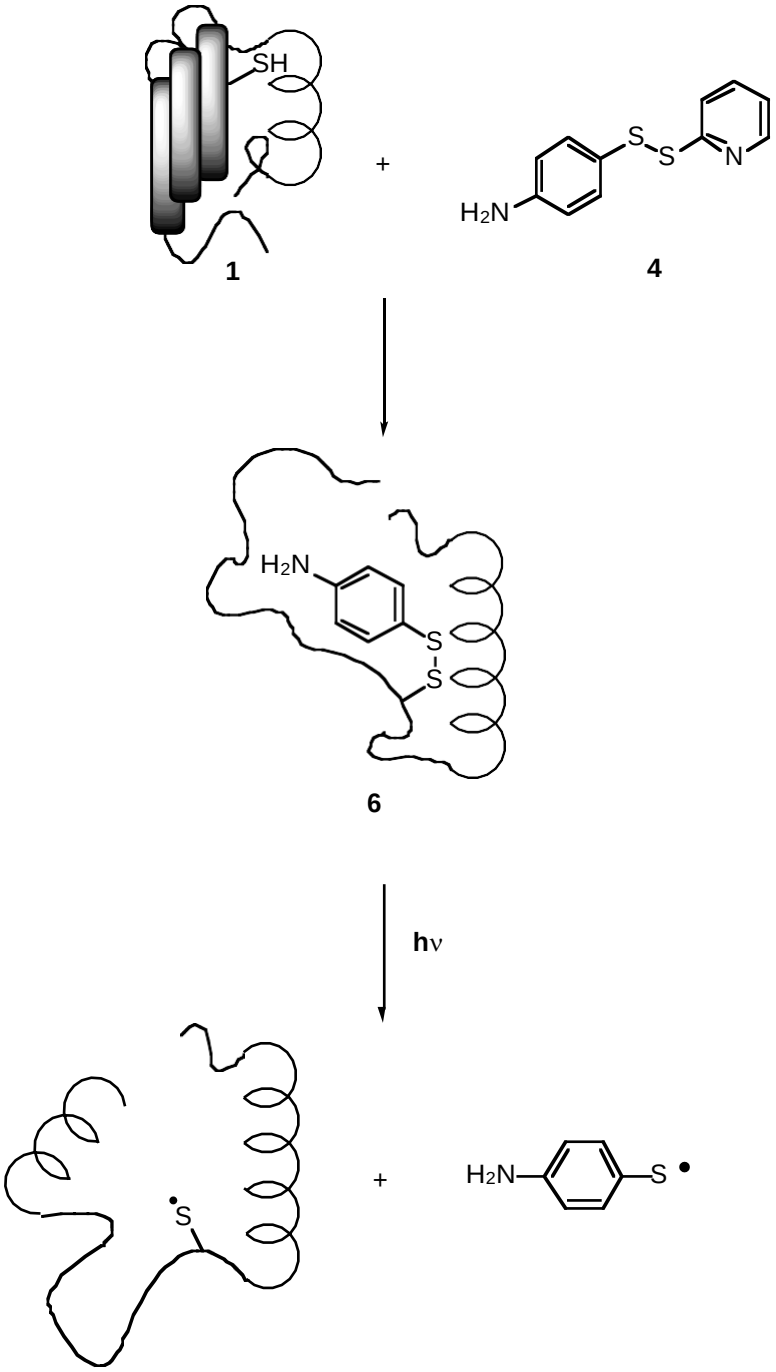


Figure 5.1

In this chapter we describe the synthesis of **4**, **5** and photocaged NPGK **6** as well as time resolved laser photolysis experiments of **5** which was used as a model compound to test the efficiency of the optical trigger mechanism (Figure 5.2).

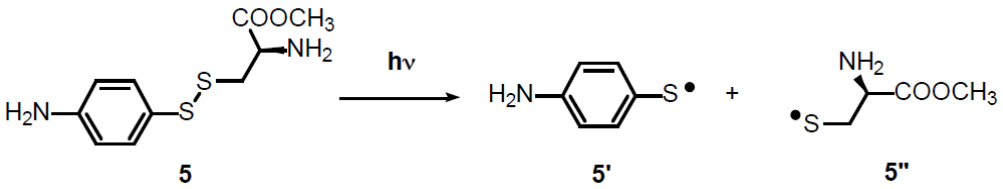


Figure 5.2

5.2 Results and Discussion

5.2.1 Synthesis and photolysis of a new photolabile unsymmetric disulfide

Disulfide **4** was synthesised in two steps (Scheme 5.1). A thiol-disulfide exchange reaction between *p*-aminothiophenol **2** and bis (2-pyridyl) disulfide **3** yielded **4**. The yield was lowered due to the reaction of **4** with unreacted **2**, that lead to the formation of bis (*p*-aminophenyl) disulfide, which was separated from **4** by column chromatography. The structure of **4** was confirmed by X–ray crystallography (Figure 5.3). **4** was reacted with cysteine methyl ester hydrochloride to obtain **5**.

Photolysis of **5** was carried out under the same experimental conditions described for the photolysis of the maleimide-based ligands. The transient decay of the absorption signal for **5'** (Figure 5.2) observed at 550 nm in acetonitrile and in an aqueous solution of acetate buffer (pH 5) are shown in Figure 5.4. The decay of the radical absorption was fit to equation 3.5 and the resulting parameters are listed in Table 5.1.[5, 6]

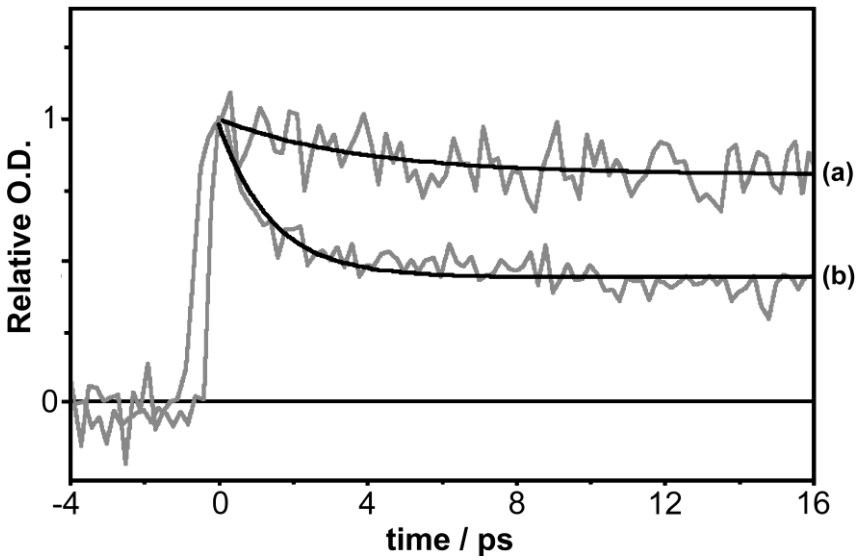
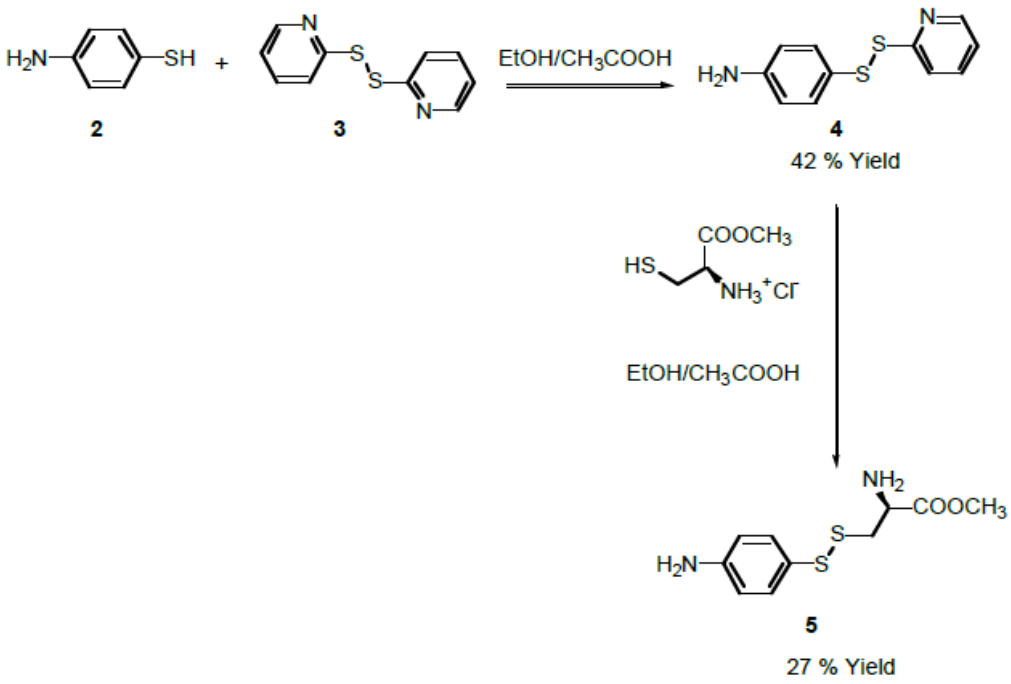


Figure 5.4. Evolution of the radical absorption signal at 550 nm after photolysis of **5** and the fitting to a monoexponential function in acetonitrile (a) and 10 mM acetate buffer solution (b). The fits to a monoexponential decay are also shown. The vertical axis is the optical density (OD) of the thiyl radical at 550 nm relative to the OD of the ground state disulfide at the same wavelength.



Scheme 5.1

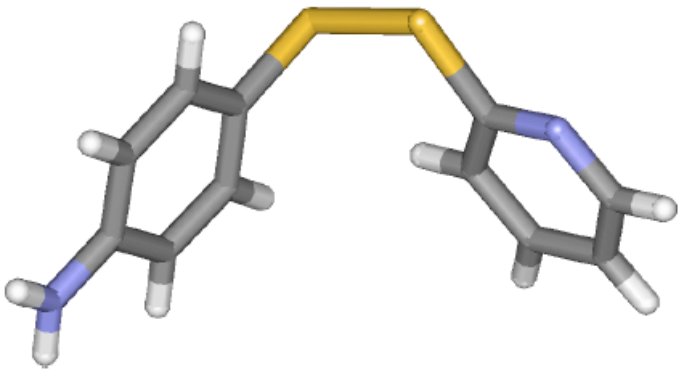


Figure 5.3. Crystal structure of **4**.

$$I(t) = Ae^{(-t/\tau)} + B \qquad \text{Equ. 3.5}$$

Since photolysis of **5** produces two different populations of thiyl radicals (Figure 5.2), it is not possible to compare the kinetic behaviour with that of thiyl radicals from symmetric disulfides. However the decay of **5** and bis (*p*-aminophenyl) thiyl radicals is qualitatively very similar. In water, a 30% greater decay of **5'** was observed compared with acetonitrile. Although there is some radical recombination over the first few picoseconds, approximately 50% of the thiyl radicals **5'** escape cage recombination in water. Therefore this system is ideal for photolabile covalent modification of proteins.

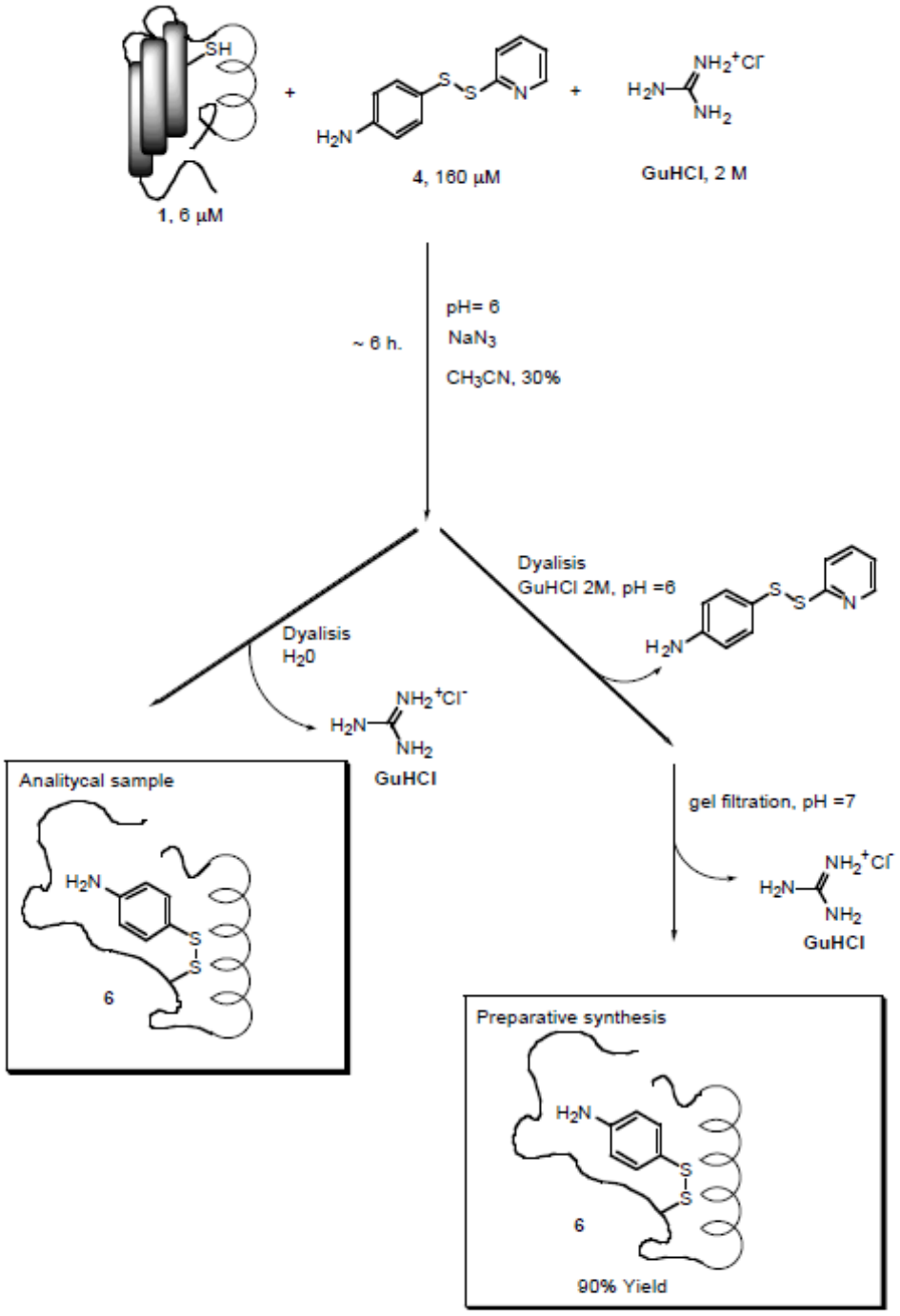
Table 5.1

Solvent (viscosity)[a]	τ (ps)	k_e (ps^{-1})	k_r (ps^{-1})	$Ax10^{-5}$	$Bx10^{-5}$	ϕ [b]
CH_3CN (0.53)	3.7	0.20	0.08	3.23	13.5	0.82 ± 0.50
H_2O (1.00)	1.4	0.30	0.31	7.73	6.18	0.44 ± 0.06

[a] Viscosity values are given in centipoise and are taken from reference 7. [b] The error in the parameters from the fitting of the curves is taken as twice the standard deviation of the experimental data points calculated for the A value.

5.2.2 Synthesis and conformational analysis of photocaged NPGK

Compound **6** was synthesised by following a procedure similar to the one used for the labelling of NPGK with the maleimide-based reagent **3** (Scheme 5.2). Thus to make the protein thiol accessible to the labelling reagent **4**, **1** was unfolded by addition of guanidine hydrochloride.[8] The reaction was carried out at pH 6, because we have shown that at slightly acidic pH NPGK reacts rapidly with disulfides bearing good leaving groups such as 5,5'-dithiobis(2-nitrobenzoic) acid (DTNB). In addition, in the case of **4**, acid conditions catalyse the labelling reaction by increasing the leaving group ability of the 2-thiopyridine due to the protonation of the nitrogen.[9,10] The concentration of **1** was maintained at 6 μM in order to suppress intramolecular thiol-disulfide exchange reactions and thus avoid formation of protein dimers. NPGK and **3** were incubated for 6 hours and the reaction was terminated by diluting ten folds in water. The solution was then dialysed in water for the


SH
1, 6 µM
+
S
S
N
H2N
4, 160 µM
+
NH2+Cl-
H2N
NH2
GuHCl, 2 M
~ 6 h.
pH= 6
NaN3
CH3CN, 30%
Dyalisis
H20
NH2+Cl-
H2N
NH2
GuHCl
Dyalisis
GuHCl 2M, pH =6
H2N
gel filtration, pH =7
NH2+Cl-
H2N
NH2
GuHCl
Analitycal sample
H2N
S
S
6
Preparative synthesis
H2N
S
S
6
90% Yield

Scheme 5.2

preparation of an analytical sample of labelled protein **6** which was analysed by mass spectroscopy. The mass spectrum data showed that **6** was formed in 100 % yield under these conditions (Figure 5.5).

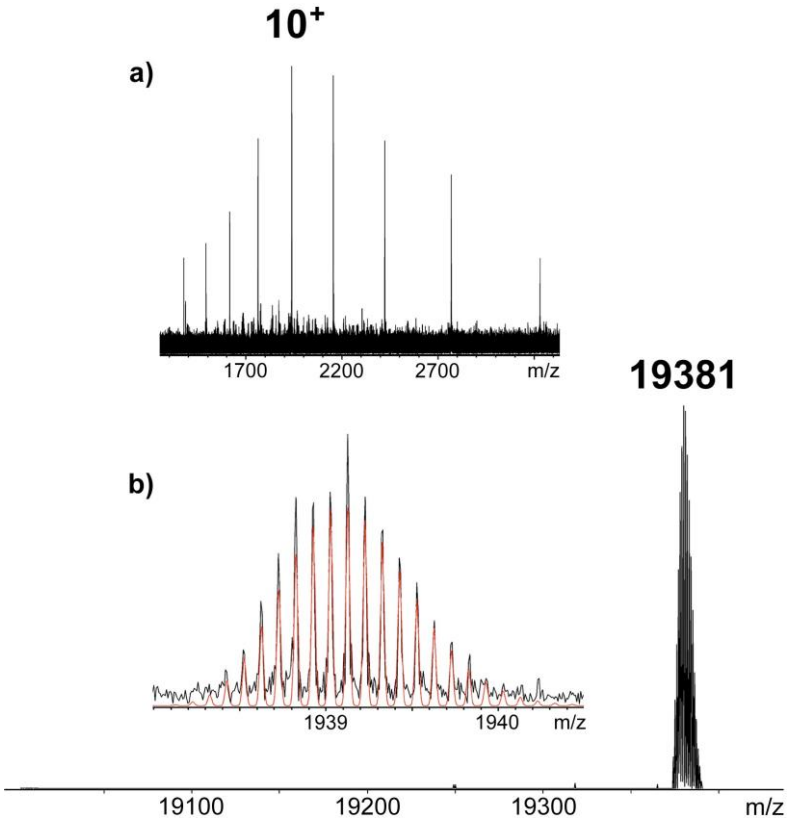


Figure 5.5. ESI-FTICR (Fourier transform ion/cyclotron resonance-electrospray ionisation) mass spectrum of **6**: a) spectrum of intact **6** showing peaks corresponding to protein molecules with varying net charges *z*. b) Overlap of the theoretical and measured isotopic distribution of the 10^+ charge state of **6**, and deconvoluted mass spectrum of **6** calculated from the 10^+ charge state. Protein (~ 40 μM) was dissolved in a 1:1 solution of water and acetonitrile with 2% formic acid.

In order to test if the thiol modification induced changes in the conformation of the protein, the synthesis of **6** was repeated on a preparative scale (Scheme 5.2). In this case, the refolding of the labelled protein was carried out by gel filtration instead of dialysis.

A comparison of the UV/Vis. absorption spectra of **6**, **1** and **5**, showed that upon chemical modification the optical density of the protein was increased at 250 nm and above 300 nm (Figure 5.6).

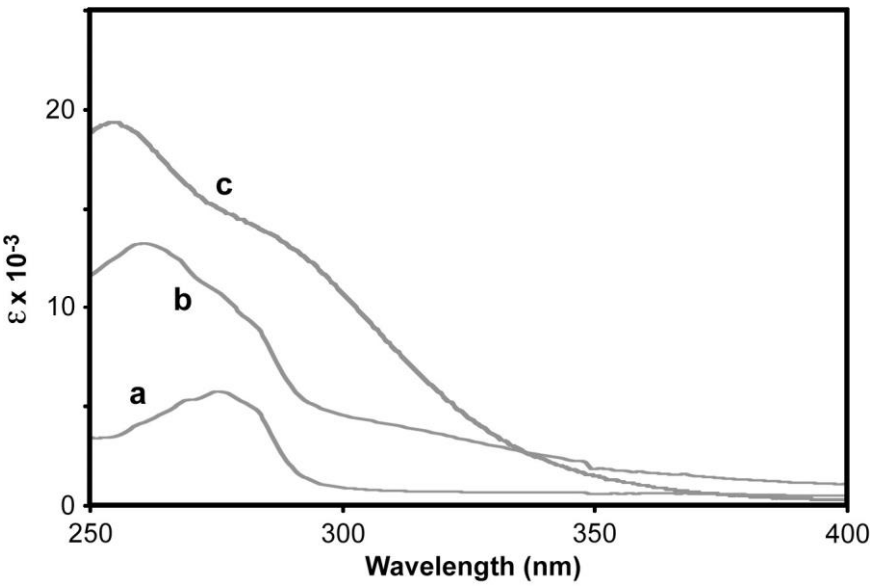


Figure 5.6. Normalised UV/Vis absorption spectra of **1** (a), **6** (b), and **4** (c). The concentration of **6** was determined with a Bradford assay**.**

Since the unlabelled protein does not absorb above 300 nm, an estimate of the reaction yield was initially obtained by calculating the concentration of thioaniline bound to the protein based on the extinction coefficient of **5** at 310 nm (8530 M^{-1} cm^{-1}). However this underestimated the protein concentration by nearly 50% as judged by the Bradford assay and the intensity of the ^{1}H NMR spectrum.[11]

According to the Bradford assay, the concentration of **6** in solution was 50 μM which corresponded to a yield for the labelling of nearly 90 %. Thus it seems that the extinction coefficient of the thioaniline bound to the protein is different from that of **5** probably because it is inside a protein that is at least partially folded and therefore experiences a non–aqueous environment.

In order to test if the labelling reaction resulted in formation of protein dimers, **6** and **1** were subjected to gel filtration analysis. The retention time of both proteins was the same indicating that **6** is a monomer (Figure 5.7).

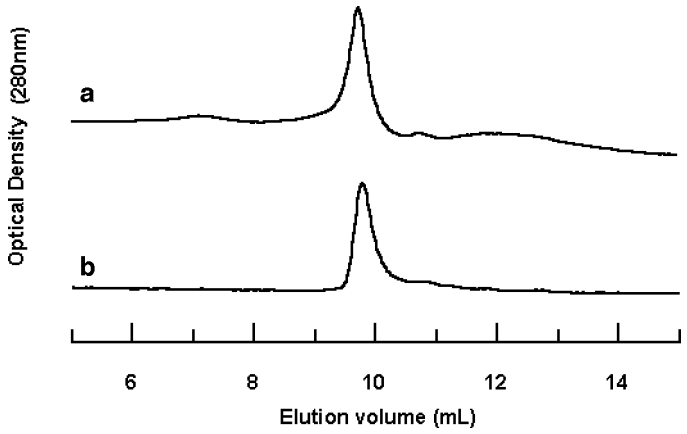


Figure 5.7. Elution profiles for **1** (a) and **6** (b) obtained by analytical gel filtration. Protein concentration was 20 μM in 50 mM sodium phosphate pH 7, 3 mM sodium azide.

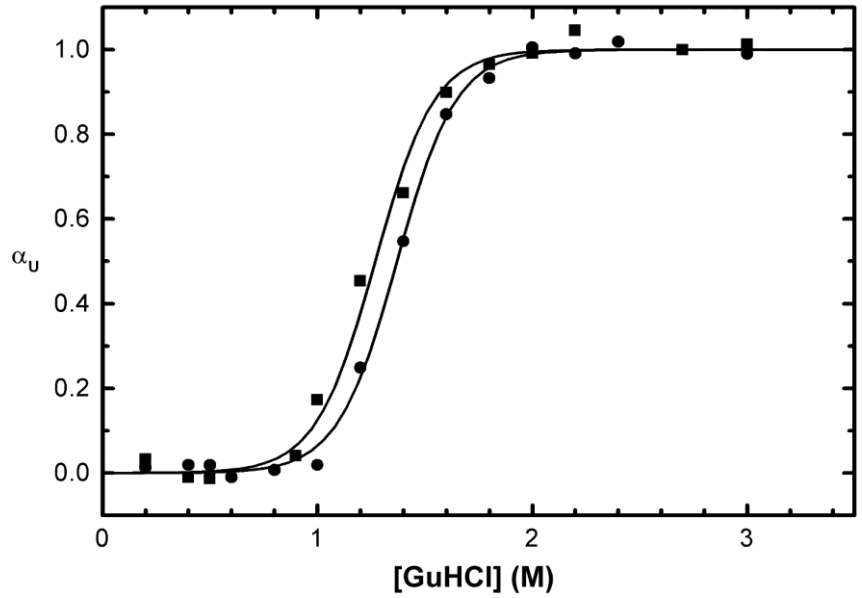


Figure 5.8. Equilibrium unfolding profiles of **1** (squares) and **6** (circles) obtained by circular dichroism measurements (Far-UV CD). The mole fraction of the unfolded state (α_U) is plotted as a function of denaturant concentration. The curves were obtained by fitting the data to equation 4.2.

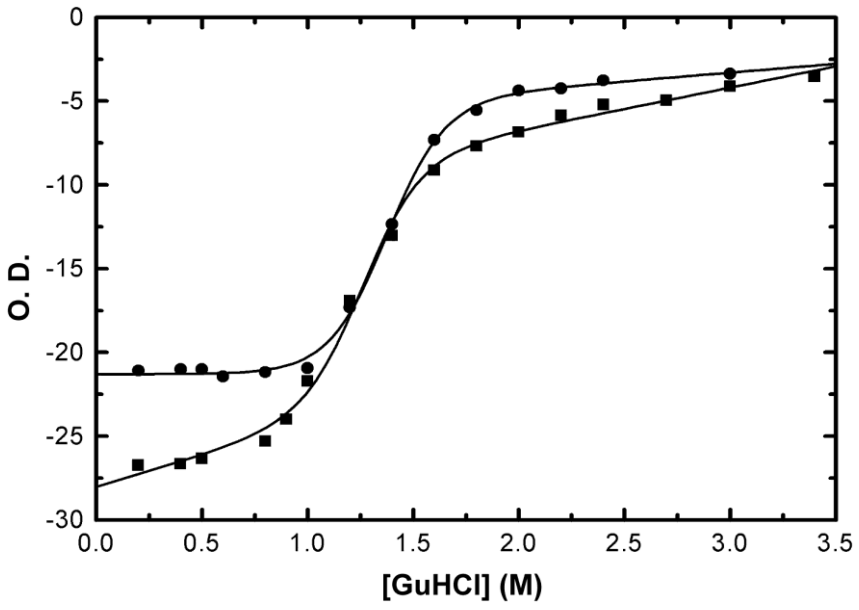


Figure 5.9. Equilibrium unfolding profiles of **1** (squares) and **6** (circles) obtained by circular dichroism measurements (Far-UV CD). The absorbance at 222 nm is plotted as a function of denaturant concentration. The curves were obtained by fitting the data to equation 4.2 and are not normalised.

5.2.3 Unfolding equilibrium measurements of photocaged NPGK

Circular dichroism and fluorescence measurements were carried out on both **1** and **6** to test the effect of the labelling reaction on the protein conformation. For both proteins, measurements were carried out without adding dithiothreitol (DTT) in order to avoid cleavage of the disulfide. Previous studies on the unfolding transition of NPGK have been carried out without addition of reducing agents because at the low concentrations of protein used for the measurements (10 μM) formation of dimers species is unlikely.[8]

The experimental data were fitted to equation 4.2 assuming a two-state model for the unfolding reaction: [12]

$$I = \alpha_F \cdot I_F + \alpha_U \cdot I_U \qquad \text{Equ. 4.2}$$

where α_F and α_U are the fractional proportion of molecules in the folded and unfolded states respectively and I, I_U, I_F are the intensities of the measured, folded and unfolded signals respectively.[9] The circular dichroism measurements indicated that the average secondary structure content was very similar for **6** and **1** throughout the unfolding transition (Figure 5.8 and 5.9).

In the fluorescence spectrum, the intensity of the tyrosine fluorescence in the labelled protein **6** was significantly lower than for **1**, possibly due to energy transfer to the aniline (Figure 5.10). An energy transfer process is likely because the tyrosine residues fluoresce at 305 nm which corresponds to the absorption of compound **5** (Figure 5.6). [13] When **6** was denatured with guanidine hydrochloride, the increase in fluorescence was too small to be useful for quantifying the unfolding transition.

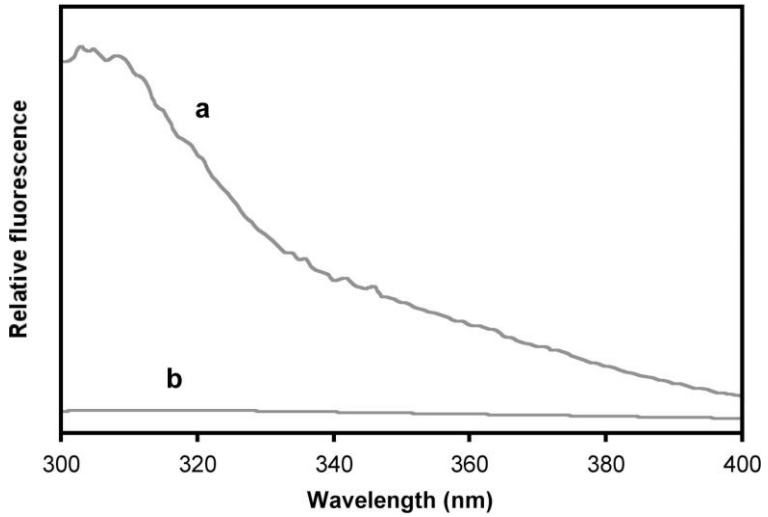


Figure 5.10. Fluorescence spectra of **1** (a) and **6** (b) at 5 μM in 50 μM phosphate buffer, 3 mM sodium azide (pH 7).

In order to evaluate if the labelling reaction induces a significant perturbation in the environment of the aromatic residues, **1** and **6** were analysed by ^{1}H NMR spectroscopy (Figure 5.11).

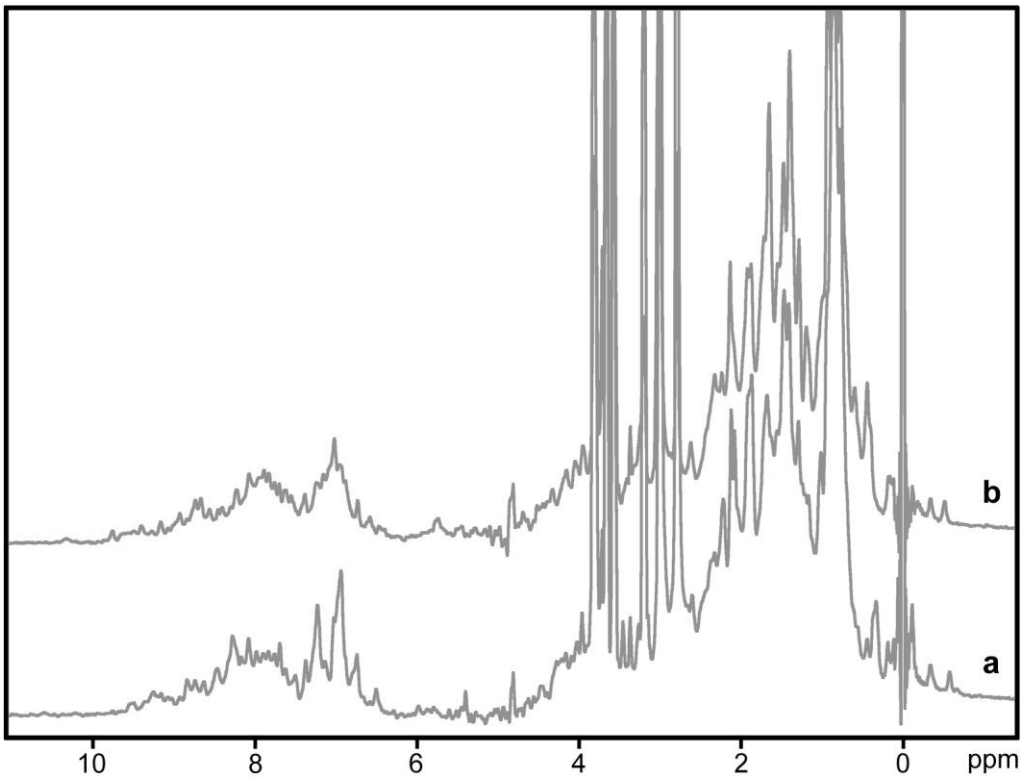


Figure 5.11. 500 MHz ^{1}H NMR spectra of **6** (a) and **1** (b). Protein concentration was 40 μM in 50 mM sodium phosphate pH 7, 3 mM sodium azide, 10 % D_2O, 800 μM TSP (3-trimethylsilyl-propionate).

Chemical shift dispersion is similar in the two spectra indicating a comparable degree of structure. In particular, the methyl singlets at < 0 ppm show that the environment in the hydrophobic core is essentially identical in the two proteins.

5.3 Conclusions

We have established a procedure which gives good yields of a photocaged protein. We have shown that in the case of wild type NPGK, the single amino acid modification is not sufficient to induce a significant perturbation of the native-like protein fold, probably because there is enough space to tolerate the insertion of a small molecule in the hydrophobic core. However, the labelling procedure could be extended to any protein that positions thiols in a less tolerant region of the structure.

5.4 Experimental Section

5.4.1 Materials

Chemicals were purchased from Sigma Aldrich and Sigma, except triethanolamine and sodium dihydrogen ortho phosphate, which were purchased from BDH, and were used without further purification. HPLC (Fluka) grade solvents were used for UV analyses and photolysis experiments. TLC was carried out using Merck Silica Gel 60 (F_{254}) precoated aluminium sheets. Visualisation was under UV light or exposure to iodine vapours. Flash chromatography was performed using Merck silica gel 60 (40-63 μm). Ultrafiltration membranes (regenerated cellulose, YM 10) were purchased from Amicon/Millipore. Vivaspin Centrifugal filter devices (MW cut off = 5000, Polyethersulphone membrane, 20 mL capacity) were purchased from Sartorius/Vivascience. Prepacked Sephadex G-25 PD 10 columns (exclusion limit, MW = 5000) were purchased from Amersham Pharmacia.

5.4.2 Instruments

Aromatic Disulfide experiments

^{1}H and ^{13}C NMR spectra were recorded on a Bruker AC250 spectrometer. All chemical shifts are quoted in ppm and are referenced to the solvent signal. The coupling constants are expressed in Hz. The following abbreviations are used in the assignment of ^{1}H spectra: s singlet, d doublet, dd doublet of doublets, t triplet, q quintet, m multiplet, br broad signal. Positive fast atom bombardment (FAB^{+}) and positive chemical ionisation (CI^{+}) mass spectra were recorded on a Fison VG ProSpec 8 using a matrix of *m*-nitrobenzyl alcohol (NBA) and a magnetic sector detection system. UV/Visible absorption spectra were recorded on a Varian Cary 1 spectrophotometer. For photolysis experiments the femtosecond laser and transient absorption spectrophotometer were built and set up by Dr. Gavin Reid and Prof. Godfrey Beddard at Leeds University. A detailed description of this laser system is reported in reference 14. For the present experiment the laser set up was modified as follows: the pump beam was 520 nm and it was frequency doubled using a 100 micro-metre thick BBO crystal to 260 nm. Samples were excited with 100 fs of the 260 nm light (the energy was a maximum of 100 nJ). The zero time for the probe wavelength was determined by excitation of a solution of *trans* stilbene in acetonitrile. For each experiment the difference in absorbance between the excited and ground state species measured at different pump/probe delay times was less than 10^{-4}. The signal to noise ratio was improved by averaging 10-20 scans. The data were analysed with the same program used for data collection (ExptPPC1, version 3.0) which was kindly given to us by Dr. Gavin Reid.

Protein experiments

The pH of buffer solutions was measured using a Jenway 3020 pH meter. Protein solutions (< 2 mL) were dried using a speedvac Plus SC110A Savant. UV/Visible spectra were recorded on a Varian Cary 3 Bio spectrophotometer. Fluorescence measurements were carried out with a Varian Cary Eclipse Fluorescence spectrometer. Circular Dichroism measurements were carried out with a Jasco J-810 Spectropolarimeter using a 1mm path length cuvette. ^{1}H NMR spectra were recorded in a 500 MHz Bruker AMX spectrometer equipped with a cryoprobe. Analytical gel filtration chromatography was carried out with a Waters 717 HPLC system equipped with an autosampler, photodiode array detector (Waters TM 996) and a Biosep SEC 3000 column. The HPLC system uses Millenium software (Waters) for the data processing. ESI-FTICR/MS (electrospray ionisation-Fourier transform ion-cyclotron resonance mass spectra) were obtained using a 9.4 T- Bruker Bio-apex II spectrometer and an external elctrospray ionisation ESI source (Analitica, Brandford, CT, USA). This instrument has been described in detail elsewhere.[15] All instrument operations and data analysis were performed by Liam McDonnel using the x. mass v 5.0 (Bruker Daltonics) and they were part of the EPSRC FTICR services located at Warwick University. The absolute error in the calculation of the mass values is typically between 0.5 – 2 ppm.

5.4.3 General Methods

Solutions for photolysis experiments

Concentration of **5** was 1 mM in 10 mM sodium acetate buffer (pH 5) or acetonitrile as solvents. The cuvette path length was 1 mm for all the measurements.

Protein solutions

A number of buffer solutions were used to carry out the different experiments on proteins. The composition of these buffer solutions is detailed in the following list:

- TEA (pH 6): 50 mM triethanolamine, 3 mM sodium azide.
- PBS 50 mM (pH 7): 0.021 M sodium dihydrogen ortho phosphate, 0.029 M sodium disodium hydrogen phosphate, 3 mM sodium azide.
- GuHCl (pH 6): 50 mM triethanolamine, 6 M guanidine hydrochloride.
- GuHCl (pH 7): 50 mM PBS, 6 M guanidine hydrochloride.

The final pH was adjusted by addition of aqueous solutions of 1N NaOH or HCl, and the solution was filtered through a Y 10 membrane prior to use.

The exchange of buffer and the removal of low molecular weight compounds (MW < 5000) from protein solutions was carried out by dialysis or gel permeation chromatography using, respectively, Vivaspin centrifugal devices and G75 PD10 columns. Unless otherwise stated, dialysis was carried out by transferring the sample into the centrifugal device, (0.5 - 3 mL), diluting with the desired buffer to 20 mL and reducing the volume by centrifugation to 500 μL. Then the process of dilution and concentration was repeated twice more, to finally obtain 500 μL of sample with the new buffer. For gel permeation chromatography, 2 mL of protein sample were diluted to 2.5 mL in the appropriate buffer and then loaded into the column, previously equilibrated with the same buffer. The first 3.2 ml of eluate was collected. These methodologies also allowed the of protein concentration to be changed when required.

Estimation of protein concentration was carried out using different methods, depending on the particular protein. The concentration of NPGK **1** was estimated by the UV absorption of tyrosine residues at 280 nm. An extinction coefficient of 1100 $M^{-1}cm^{-1}$ was taken for each tyrosine residue (ε_{280nm}= 4400 $M^{-1}cm^{-1}$ for **1**) and the molecular weight of the protein was determined by the database entry of its amino acid sequence (19258 Da).[8, 16] The concentration of protein **6** was estimated by a standard Bradford assay. The protocol used is detailed as follows:
the dye binding solutions were prepared following the Bradford method (Concentrated dye solution: 98 mg of Comassie brilliant blue G-250 were dissolved in a solution of 25 % v/v of ethanol 95% and 50% v/v of concentrated phosphoric acid. This solution was diluted 5 fold in water for the measurements).[11] The calibration curve was obtained from solutions of **1**, using the following samples: 7.3 μL (5 μg), 14.6 μl (10 μg), 22 μl (15 μg), 29 μL (20 μg), 36 μL (25 μg), 44 μl (30 μg) of 56 μM stock solution of **1** in PBS (pH 7), were diluted to 100 μL with PBS (pH 7). 5 mL of diluted dye solution were added to each sample and the optical density at 595 nm was recorded. Protein samples were run in duplicate and a control sample without the addition of protein was used as reference. The diluted and concentrated dye solutions were stored protected form the light at 5 ^{O}C. The optical density at 595 nm of the diluted dye solution added to 100 μL of PBS (pH 7) was constant for at least 6 months. The UV spectra of the dye solution upon addition of increasing amount of protein and the calibration curve used for the determination of protein concentration are shown in Figure 5.12.

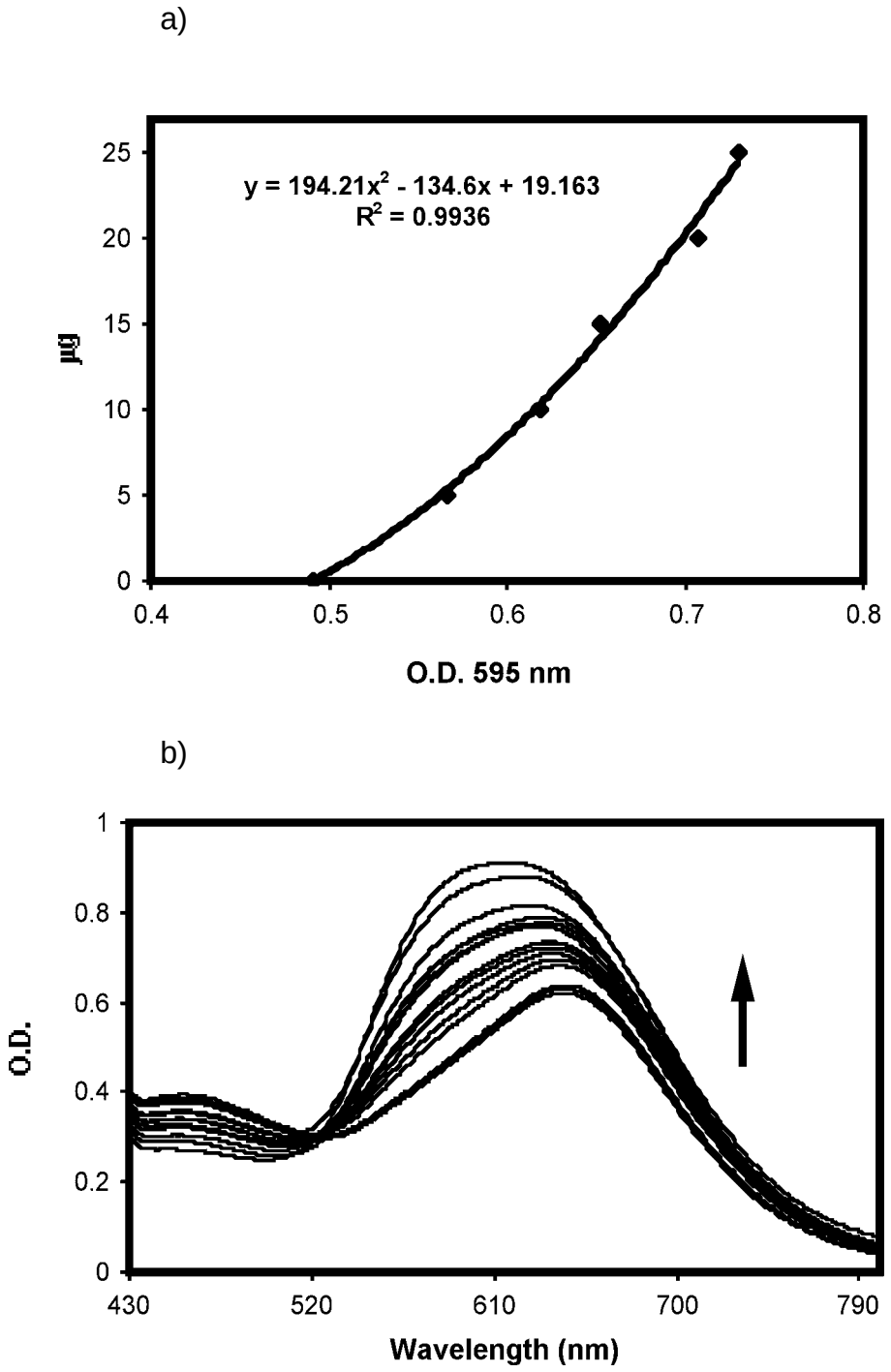


Figure 5.12. a) calibration curve used for the determination of labelled NPGK concentration. b) UV/Vis. spectra of the diluted solution of Coomassie brilliant blue G-250 upon addition of protein. The OD at 595 increases with an increase in the amount of protein added (0-30 μg in 100 μl of protein solution)).

Solutions of protein and reagents used in the labelling reactions were deoxygenated using the following method: the sample container, provided with a magnetic stirrer was transferred into a larger flask and submitted to a partial vacuum that was broken with nitrogen. This process was repeated al least 5 times.

Analytical gel filtration

Protein samples for analytical gel filtration were prepared as follow:

25 μL of 40 μM protein stock solutions (**1** and **6**) in PBS (pH 7) were diluted to a 20 μM concentration in the same buffers. Samples were transferred to the autosampler of the HPLC system and 20 μL were injected on the column, which was re-equilibrated overnight with PBS (pH 7). The flow was 1 mL/min.

Equilibrium unfolding measurements

Unfolding profiles were obtained recording the fluorescence or CD spectra of protein solutions with increasing amount of denaturant. A number of protein solutions (between 10 and 20) with the same concentration of protein and a concentration of GuHCl ranging from 0 to 3 M were prepared from stock solutions of protein 40 μM and GuHCl 6 M, both in PBS (pH 7) buffer. The desired concentration was achieved by diluting the stock solutions with PBS (pH 7) buffer. The samples were then allowed to re-equilibrate for a minimum of 20 minutes at 25 ^{O}C. For CD experiments, the final concentration of protein was 10 μM, and the spectra were collected over a 300 – 200 nm range using a 2 nm slit width (Figure 5.13). Unfolding profiles were obtained as a plot of absorbance at 222 nm *versus* denaturant concentration. For fluorescence, concentration of protein was between 4 and 5 μM, and the spectra were obtained by excitation of the tyrosines at 272 nm. Emission was recorded over a range of 280 – 400 nm using a slit width of 5 nm (Figure 5.14).

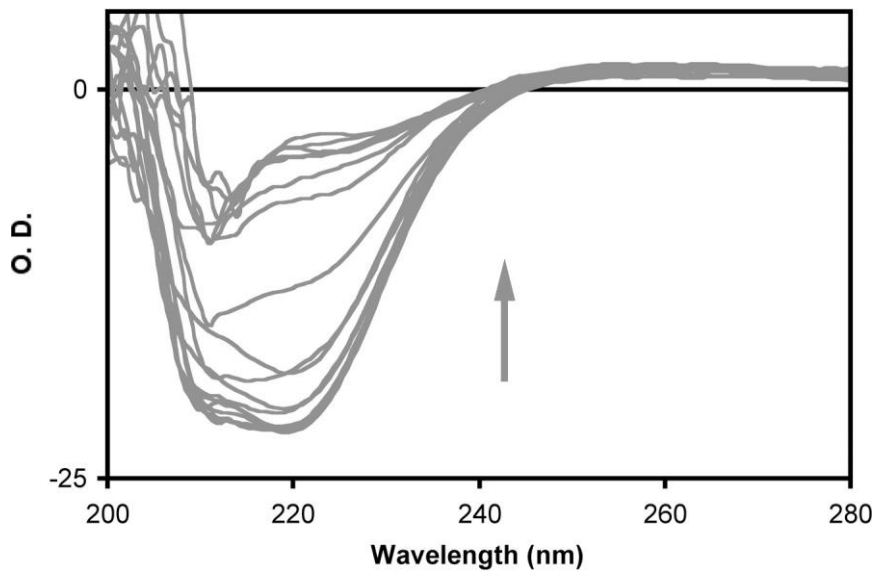


Figure 5.13. Circular dichroism spectra of **6** (a similar spectrum was obtained for **1**) at different denaturant concentrations (GuHCl) The arrow indicates increasing concentration of denaturant. The OD at 222 nm decreases upon unfolding.

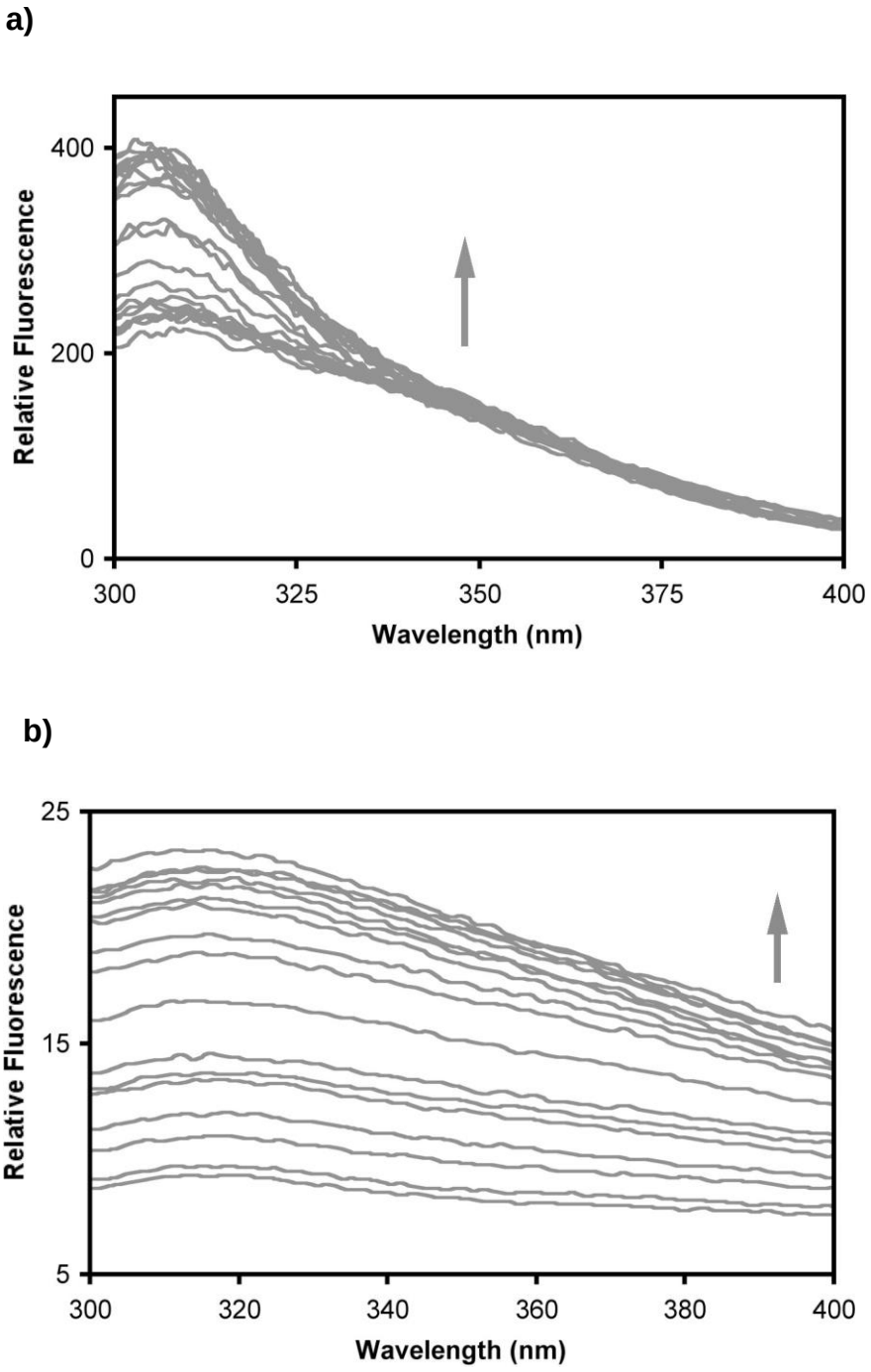


Figure 5.14. Fluorescence emission spectra of **1** (a), **6** (b) at different denaturant (GuHCl) concentration (the arrows indicate increasing denaturant concentration). The fluorescence intensity increases upon unfolding.

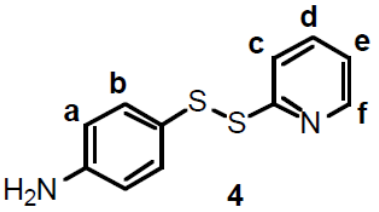
Unfolding profiles were obtained as a plot of the fluorescence intensity at 305 nm

versus denaturant concentration. Experimental data were fitted to equation 4.2 using Microcal Origin (version 5.0, Microcal Software).

^{1}H NMR spectroscopy experiments for proteins solutions

For ^{1}H NMR spectroscopy experiments, protein solutions were: 45 μM protein in PBS (pH 7), 10% (v/v) D_2O, 800 μM TSP. Spectra were recorded with 1024 scans. Water suppression was achieved using a presaturation pulse during the relaxation delay between scans. Spectra were processed using FELIX 2000 (MSI, San Diego) and referenced against TSP present in the sample.

5.4.4 Synthesis of 4



4

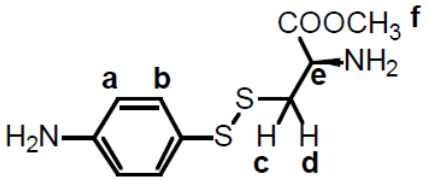
To a solution of 0.7 g (3.17 mmol) of 2-2' dipyridyldisulphide in 21 mL of a mixture of ethanol/acetic acid 20:1 v/v, were added 0.44 g (3.17 mmol) of 4-aminothiophenol and the solution was stirred for 1 hour. The solvent was then removed under reduced pressure and the remaining solid redisolved in CH_2Cl_2 (30 mL). The solution was washed with a saturated solution of $NaHCO_4$ in water (5 x 20 mL), dried with anhydrous Na_2SO_4 and the solvent removed under reduced pressure. The residue was purified by flash chromatography on silica, using a mixture chloroform /acetic acid/THF 95:3:2 v/v. The fraction containing **4** (Rf = 0.28) was washed with a saturated solution of $NaHCO_4$, dried with Na_2SO_4.and the solvents removed under reduced pressure (0.31 g, 41.7% yield).
^{1}H-NMR ($CDCl_3$): 8.47 (d, 1H, **f**, J = 3.97), 7.76 (d, 2H, **c**, J = 7.95), 7.63 (m, 1H, **d**), 7.37 (d, 2H, **b**, J= 8.85), 7.07 (m, 1H, **e**), 6.58 (d, 2H, **a**, J = 8.55).[13] C-NMR ($CDCl_3$): 160.5, 149.5, 147.2, 137.1, 132.5, 123.9, 120.6, 119.9, 115.5. MS (CI^+, gas NH_3) m/z: 112, 124, 235(MH^+). $C_{11}H_{10}N_2S_2$ requires 234.34.
*Crystal data for **4** ($C_{11}H_{10}N_2S_2$):* M = 234.33; crystallises from dichloromethane/hexane as colourless blocks; crystal dimensions 0.42 x 0.23 x 0.12 mm^3. Monoclinic, a = 8.9677(10), b = 13.0644(14), c = 9.4776(10) Å, β = 103.096(2)°, U = 1081.5(2) $Å^3$, Z = 4, D_C = 1.439 Mg/m^3, space group $P2_1/n$ (a

non-standard setting of $P2_1/c$ $C^{5,2}_{h}$, No.14), Mo-K_α radiation (λ, = 0.71073 Å), μ(Mo-K_α) = 0.457 mm^{-1}, F(000) = 488. Data collected were measured on a Bruker Smart CCD area detector with Oxford Cryosystems low temperature system. Cell parameters were refined from the setting angles of 66 reflections (θ range 2.70 to 28.29°). Reflections were measured from a hemisphere of data collected of frames each covering 0.3 degrees in omega. Of the 6677 reflections measured, all of which were corrected for Lorentz and polarisation effects and for absorption by semi empirical methods based on symmetry-equivalent and repeated reflections (minimum and maximum transmission coefficients 0.8312 and 0.9472) 2265. Independent reflections exceeded the significance level $|F|/\sigma(|F|) > 4.0$. The structure was solved by direct methods and refined by full matrix least squares methods on F^2. Hydrogen atoms were placed geometrically and refined with a riding model and with U_{iso} constrained to be 1.2 times U_{eq} of the carrier atom. Refinement converged at a final R = 0.0422 (wR_2 = 0.1162, for all 2603 data, 136 parameters, mean and maximum δ/σ 0.000, 0.000) with allowance for the thermal anisotropy of all non-hydrogen atoms. Minimum and maximum final electron density -0.654 and 0.412 $e.Å^{-3}$. A weighting scheme $w = 1/[\sigma^2(Fo^2) + (0.0735*P)^2+0.09*P]$ where $P=(Fo^2+ 2 * Fc^2)/3$ was used in the latter stages of refinement. Complex scattering factors were taken from the program package SHELXTL[Y] as implemented on the Viglen Pentium computer.

Supplementary material (see Appendix I)

anisotropic thermal vibrational parameters with e.s.d.s
hydrogen atom position parameters
observed structure amplitudes and calculated structure factors

5.4.5 Synthesis of 5



5

The same procedure described for syntheses of **4** was used, starting from 0.12 g (0.5 mmol) of **4** and 0.09 g (0.5 mmol) of L-cysteine methyl ester hydrochloride. Eluents for flash column chromatography on silica were: CH_2Cl_2/ethylacetate 70:30 v/v to

elute the impurity of bis (4-aminophenyl) disulphide (Rf = 0.9), followed by CH_2Cl_2/ethylacetae/methanol 80:15:5 v/v to elute **5** (Rf = 0.42) (0.04 g, 27 % yield).

^{1}H-NMR ($CDCl_3$): 7.85 (d, 2H, **b**, J = 7.95), 6.6 (d, 2H, **a**, J = 7.95), 3.82 (dd, 1H, **e**), 3.7 (s, 3H, **f**), 3.15 (dd, 1H, **c** or **d**), 2.85 (dd, 1H, **c** or **d**). MS (ES^+) m/z: = 94, 124, 259(MH^+), 281(MNa^+). ^{13} C-NMR ($CDCl_3$): 174.5, 147.5, 133.9, 124.9, 115.7, 53.4, 52.5, 43.5. HRMS (FAB^+) m/z: = 259.056736 (MH^+), $C_{10}H_{14}N_2O_2S_2$ requires 259.057496. UV/Vis. (H_2O, pH 5): ε_{260}= 19750, ε_{280}= 15250, ε_{310}= 8500.

5.4.6 Preparation of an analytical sample of 6 (~ 0.3 mg)

An analytical sample of **6** was prepared from protein **1** and disulfide **4** with the following procedure:
to a 10 mL vial equipped with a magnetic strirrer and inserted in a 50 mL flask, were added in the following order:

800 μL of 6 M GuHCl (TEA, pH 6)

650 μL of distilled water

75 μl of 0.1M sodium azide solution

760 μl of acetonitrile

40 μl of 10 mM solution of **4** in acetonitrile

The solution was deoxygenated under nitrogen atmosphere and 70 μl of 0.25 mM **1** in TEA buffer (pH 6) was added under gentle stirring. The solution was deoxygenated again and then left under gentle stirring in nitrogen atmosphere and in the dark. After 6 hours, the solution was transferred to a centrifugal filter device and dialysed in distilled water. The resulting sample (400 μL) was lyophilised. For mass spectroscopy analysis (ESI–FTICR), the dry sample was dissolved in 200 μL of a 1:1 mixture of acetonitrile: 4% formic acid v/v.
Expected for **6**, $C_{859}H_{1383}N_{247}O_{253}S_5$; 19381. Found (calculated as average of the 9^+ and 10^+ protein charge states): 19381.

5.4.7 Synthesis of a preparative sample of 6 (~ 3 mg)

For the preparative synthesis of **6**, the procedure described above was scaled up 10 fold to a final volume for the reaction mixture of 24 mL. After 6 hours, the reaction mixture was split into 4 portions of 6 mL, each transferred into a centrifugal filter device and diluted to 20 mL with a solution 2M GuHCl, 50 mM PBS (pH 7), 3 mM sodium azide. The solutions were centrifuged to a final volume each of 1 mL and

then they were reunited (4 mL) in a single centrifugal filter device, and the resulting solution was centrifuged to a final volume of 2 mL. The 2 mL were diluted to 2.5 mL with PBS buffer (pH 7) and then loaded on a PD10 column previously re-equilibrated with PBS buffer (pH 7). The protein was eluted in the first 3.2 ml (stock solution) ($OD_{280\ nm}$ = 0.42, $OD_{310\ nm}$ = 0.18). For the determination of protein concentration, 24 μl of the 3.2 ml stock solutions were added to a diluted solution of Comassie dye for the Bradfora assay (OD_{595} = 0.60). The calibration curve showed in Figure 5.11 was used to calculate the protein concentration (1mg/mL, 48 μM, yield 90%).

5.5 References

1. Y. Yang, H. M. Zhou, *Biochim. Biophys. Acta*, **1998**, *1388*, 190-198.
2. Y. Yang, H. M. Zhou, Y. D. Park, *Biochem. Biophys. Res. Commun*. **1999**, *259*, 450-454.
3. T. Okuno, S. Hirota, O. Yamauchi, *Biochemistry* **2000**, *39*, 7538-7545.
4. G. J. Davies, S. J. Gamblin, J. A. Littlechild, H. C. Watson, *Proteins: Struct. Funct. Genet*. **1993**, *15*, 283-289.
5. Y. Hirata, T. Niga, T. Okada, *Chem. Phys. Lett*. **1994**, *221*, 283–288.
6. G. S. Beddard, M. G. Hyde, *Chemical Physics* **1991**, *151*, 239–248.
7. J. A. Riddick, W. B. Bunger, T. K. Sakan, *Organic Solvents (Physical Properties and Methods of purification)* **4th Edition**, Techniques of Chemistry, Vol. II, Wiley Interscience.
8. M. J. Parker, J. Spencer, G. S. Jackson, S. G. Burston, L. L. P. Hosszu, C. J. Craven, J. P. Waltho, A. R. Clarke, *Biochemistry* **1996**, *35*, 15740-15752.
9. J. C. Janson, L. Ryden, *Protein purification* **1989**, VCH Publishers, 258.
10. J. Carlsson, H. Drevin, R. Axen, *Biochem. J.* **1978**, *173*, 723-737.
11. M. M. Bradford, *Anal. Biochem*. **1976**, *72*, 248-254.
12 . M. J. Parker, J. Spencer, A. R. Clarke, *J. Mol. Biol*. **1995**, *253*, 771-786.
13. T. E. Creighton, *Proteins: structure and molecular properties*, 2nd Edition **1993**, 177, W. H. Freeman and Company.
14. G. D. Reid, D. J. Whittaker, M.A. Day, D. A. Turton, V. Kayser, J.M. Kelly, G. B. Beddard, *J. Am. Chem. Soc.* **2001**, *124*, 5518-5527.
15. M. Palmblad, K. Hakansson, P. . Hakansson, X. Feng, H. J. Cooper, A. E. Giuannakopulus, P. S. Green, P. J. Derrick, *Eur. J. Mass. Spectr*. **2000**, *6*, 267-275.
16. http: // *ca. expasy. org./cgi.-bin/protparam* (entry code for Bacillus sterarothermophilus PGK: P18912).

Chapter 6

6. Implications for Future Work

6.1 Introduction

Although preparation of novel photocaged proteins has been achieved, the crosslinked NPGK has been prepared as a mixture of products and the single amino acid modification approach is not enough to perturb the native fold of NPGK. Further experiments to improve synthetic procedures, better understand the structural properties of the modified proteins and exploit these developments are:

6.2 Crosslinked NPGK

In the case of the crosslinked protein, the synthetic procedure may be improved by carrying out the ligand functionalisation in the folded state and the croosslinking reaction in the unfolded state. This approach may have two advantages: first it may result in formation of a pure product thus eliminating the needed of purification protocols; second, it may result in a better yield of modified protein if the precipitation observed with the protocol outlined in chapter 4 is due to the presence of side products.
To resolve the issue of whether or not the cyclic protein resembles a molten- globule state, an obvious experiment would be to carry out thermal unfolding measurements. Since the enthalpy of the molten-globule state is similar to that of an unfolded state, these data may be complementary to the CD, fluorescence and NMR data presented in chapter 4.[1-3]

6.3 Photocaged ferrochelatase

Although modification of the single cysteine in wild-type NPGK has not provided a suitable substrate for the observation of fast refolding reactions, we have decided to apply this procedure to *Bacillus subtilis* ferrochelatase, a single domain protein which is currently grown and purified in the group of Prof. Neil Hunter in the department of

Molecular Biology at Sheffield University. It has been shown that chemical modification of the single cysteine of this ferrochelatase results in a complete loss of enzymatic activity.[4] Thus we believe that functionalising the single thiol with our optical trigger, may provide insight in the conformational changes associated with the catalytic activity of this protein in a time domain not accessible to traditional stop flow measurements.

6.4 References

1. T. E. Creighton, *Proteins: structure and molecular properties*, Chapter 6, 2nd Edition **1993**, 290-292, W. H. Freeman and Company.
2. O. B. Ptitsyn, *J. of Protein Chemistry* **1987**, *6*, 273-293.
3. K. Kuwajima, *Proteins: Structure, Function, and Genetics* **1989**, *6*, 87-103.
4. M. Hansson, L. Hederstedt, *European Journal of Biochemistry* **1994**, *220*, 201-208.

Appendix I

Appendix I X-Ray Crystallography Data for Chapter 3 and 5

I.1 X-ray crystal structure of 2 (Chapter 3)

Table 1. Crystal data and structure refinement for **2**

Identification code	och103s	
Empirical formula	C16 H20 N2 S2	
Formula weight	304.46	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	$P2_1/c$	
Unit cell dimensions	a = 16.0946(12) Å	α= 90°.
	b = 14.9471(12) Å	β= 115.7380(10)°.
	c = 14.6262(12) Å	γ = 90°.
Volume	3169.5(4) $Å^3$	
Z	8	
Density (calculated)	1.276 Mg/m^3	
Absorption coefficient	0.328 mm^{-1}	
F(000)	1296	
Crystal size	0.16 x 0.16 x 0.05 mm^3	
Theta range for data collection	1.40 to 28.29°.	
Index ranges	-12<=h<=21, -19<=k<=19, -19<=l<=19	
Reflections collected	20877	
Independent reflections	7602 [R(int) = 0.1257]	
Completeness to theta = 28.29°	96.4 %	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.9838 and 0.9494	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	7602 / 0 / 361	
Goodness-of-fit on F^2	0.960	
Final R indices [I>2sigma(I)]	R1 = 0.0628, wR2 = 0.1243	
R indices (all data)	R1 = 0.1140, wR2 = 0.1429	
Largest diff. peak and hole	0.524 and -0.472 $e.Å^{-3}$	

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3) for och103s. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
S(1A)	4610(1)	4799(1)	6557(1)	31(1)
S(2A)	5406(1)	4786(1)	8122(1)	35(1)
N(1A)	2085(1)	1696(1)	5888(2)	27(1)
N(2A)	8388(2)	2148(1)	8777(2)	35(1)
C(1A)	3859(2)	3879(2)	6340(2)	25(1)
C(2A)	4115(2)	3018(2)	6196(2)	25(1)
C(3A)	3521(2)	2305(2)	6027(2)	25(1)
C(4A)	2638(2)	2430(2)	5994(2)	22(1)
C(5A)	2378(2)	3297(2)	6107(2)	28(1)
C(6A)	2982(2)	4006(2)	6284(2)	29(1)
C(7A)	6290(2)	3998(2)	8323(2)	29(1)
C(8A)	7156(2)	4281(2)	8423(2)	30(1)
C(9A)	7846(2)	3670(2)	8586(2)	31(1)
C(10A)	7698(2)	2743(2)	8636(2)	28(1)
C(11A)	6828(2)	2468(2)	8541(2)	31(1)
C(12A)	6145(2)	3089(2)	8398(2)	30(1)
C(13A)	1150(2)	1788(2)	5789(2)	29(1)
C(14A)	779(2)	892(2)	5926(2)	34(1)
C(15A)	8241(2)	1186(2)	8674(2)	33(1)
C(16A)	9120(2)	706(2)	8846(3)	47(1)
S(1B)	-219(1)	7428(1)	2094(1)	31(1)
S(2B)	578(1)	7279(1)	3649(1)	33(1)
N(1B)	-2857(1)	4364(1)	952(2)	29(1)
N(2B)	3468(2)	4643(1)	3968(2)	29(1)
C(1B)	-999(2)	6521(2)	1758(2)	25(1)
C(2B)	-776(2)	5687(2)	1486(2)	27(1)
C(3B)	-1391(2)	4985(2)	1206(2)	27(1)
C(4B)	-2273(2)	5088(2)	1187(2)	25(1)
C(5B)	-2492(2)	5926(2)	1449(2)	28(1)
C(6B)	-1873(2)	6622(2)	1727(2)	29(1)
C(7B)	1425(2)	6491(2)	3741(2)	28(1)
C(8B)	2307(2)	6773(2)	3872(2)	31(1)

C(9B)	2974(2)	6166(2)	3962(2)	31(1)
C(10B)	2796(2)	5238(2)	3908(2)	27(1)
C(11B)	1923(2)	4956(2)	3796(2)	28(1)
C(12B)	1264(2)	5577(2)	3724(2)	29(1)
C(13B)	-3783(2)	4440(2)	882(2)	33(1)
C(14B)	-4236(2)	3532(2)	724(2)	39(1)
C(15B)	3269(2)	3709(2)	3692(2)	30(1)
C(16B)	4103(2)	3238(2)	3712(2)	37(1)

Table 3. Bond lengths [Å] and angles [°] for och103s.

S(1A)-C(1A)	1.766(3)
S(1A)-S(2A)	2.0817(10)
S(2A)-C(7A)	1.771(3)
N(1A)-C(4A)	1.378(3)
N(1A)-C(13A)	1.454(3)
N(2A)-C(10A)	1.367(3)
N(2A)-C(15A)	1.454(3)
C(1A)-C(6A)	1.391(4)
C(1A)-C(2A)	1.395(3)
C(2A)-C(3A)	1.381(3)
C(3A)-C(4A)	1.413(3)
C(4A)-C(5A)	1.394(3)
C(5A)-C(6A)	1.385(3)
C(7A)-C(12A)	1.390(3)
C(7A)-C(8A)	1.402(4)
C(8A)-C(9A)	1.376(4)
C(9A)-C(10A)	1.413(3)
C(10A)-C(11A)	1.407(4)
C(11A)-C(12A)	1.384(4)
C(13A)-C(14A)	1.515(3)
C(15A)-C(16A)	1.506(4)
S(1B)-C(1B)	1.766(2)
S(1B)-S(2B)	2.0821(10)
S(2B)-C(7B)	1.762(3)
N(1B)-C(4B)	1.376(3)
N(1B)-C(13B)	1.454(3)
N(2B)-C(10B)	1.373(3)
N(2B)-C(15B)	1.449(3)
C(1B)-C(6B)	1.396(3)
C(1B)-C(2B)	1.402(3)
C(2B)-C(3B)	1.378(3)
C(3B)-C(4B)	1.415(3)
C(4B)-C(5B)	1.400(3)
C(5B)-C(6B)	1.374(3)
C(7B)-C(12B)	1.389(3)
C(7B)-C(8B)	1.411(4)

C(8B)-C(9B)	1.368(4)
C(9B)-C(10B)	1.412(3)
C(10B)-C(11B)	1.407(4)
C(11B)-C(12B)	1.378(4)
C(13B)-C(14B)	1.509(4)
C(15B)-C(16B)	1.504(4)

C(1A)-S(1A)-S(2A)	103.68(9)
C(7A)-S(2A)-S(1A)	104.79(9)
C(4A)-N(1A)-C(13A)	121.8(2)
C(10A)-N(2A)-C(15A)	122.9(2)
C(6A)-C(1A)-C(2A)	118.3(2)
C(6A)-C(1A)-S(1A)	120.14(19)
C(2A)-C(1A)-S(1A)	121.5(2)
C(3A)-C(2A)-C(1A)	120.9(2)
C(2A)-C(3A)-C(4A)	120.8(2)
N(1A)-C(4A)-C(5A)	122.7(2)
N(1A)-C(4A)-C(3A)	119.4(2)
C(5A)-C(4A)-C(3A)	117.9(2)
C(6A)-C(5A)-C(4A)	120.7(2)
C(5A)-C(6A)-C(1A)	121.3(2)
C(12A)-C(7A)-C(8A)	118.7(2)
C(12A)-C(7A)-S(2A)	120.9(2)
C(8A)-C(7A)-S(2A)	120.5(2)
C(9A)-C(8A)-C(7A)	120.6(2)
C(8A)-C(9A)-C(10A)	121.2(3)
N(2A)-C(10A)-C(11A)	122.2(2)
N(2A)-C(10A)-C(9A)	120.1(2)
C(11A)-C(10A)-C(9A)	117.7(2)
C(12A)-C(11A)-C(10A)	120.6(2)
C(11A)-C(12A)-C(7A)	121.3(3)
N(1A)-C(13A)-C(14A)	110.7(2)
N(2A)-C(15A)-C(16A)	110.8(2)
C(1B)-S(1B)-S(2B)	104.40(8)
C(7B)-S(2B)-S(1B)	104.12(9)
C(4B)-N(1B)-C(13B)	121.7(2)
C(10B)-N(2B)-C(15B)	122.2(2)
C(6B)-C(1B)-C(2B)	117.8(2)

C(6B)-C(1B)-S(1B)	120.66(19)
C(2B)-C(1B)-S(1B)	121.5(2)
C(3B)-C(2B)-C(1B)	121.4(2)
C(2B)-C(3B)-C(4B)	120.6(2)
N(1B)-C(4B)-C(5B)	123.1(2)
N(1B)-C(4B)-C(3B)	119.4(2)
C(5B)-C(4B)-C(3B)	117.5(2)
C(6B)-C(5B)-C(4B)	121.4(2)
C(5B)-C(6B)-C(1B)	121.2(2)
C(12B)-C(7B)-C(8B)	117.8(2)
C(12B)-C(7B)-S(2B)	121.6(2)
C(8B)-C(7B)-S(2B)	120.6(2)
C(9B)-C(8B)-C(7B)	121.1(2)
C(8B)-C(9B)-C(10B)	121.0(2)
N(2B)-C(10B)-C(11B)	122.2(2)
N(2B)-C(10B)-C(9B)	119.9(2)
C(11B)-C(10B)-C(9B)	117.9(2)
C(12B)-C(11B)-C(10B)	120.3(2)
C(11B)-C(12B)-C(7B)	121.9(2)
N(1B)-C(13B)-C(14B)	110.7(2)
N(2B)-C(15B)-C(16B)	111.0(2)

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for och103s. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
S(1A)	26(1)	26(1)	39(1)	2(1)	11(1)	-4(1)
S(2A)	30(1)	35(1)	38(1)	-11(1)	13(1)	-3(1)
N(1A)	22(1)	24(1)	34(1)	-1(1)	12(1)	-1(1)
N(2A)	25(1)	31(1)	40(1)	1(1)	6(1)	-3(1)
C(1A)	24(1)	23(1)	25(1)	2(1)	8(1)	-1(1)
C(2A)	20(1)	29(1)	27(1)	1(1)	10(1)	2(1)
C(3A)	24(1)	21(1)	29(1)	1(1)	10(1)	5(1)
C(4A)	21(1)	25(1)	20(1)	1(1)	7(1)	-1(1)
C(5A)	19(1)	28(1)	35(2)	1(1)	10(1)	2(1)
C(6A)	29(2)	21(1)	37(2)	-1(1)	13(1)	4(1)
C(7A)	25(2)	33(2)	24(1)	-5(1)	7(1)	-3(1)
C(8A)	33(2)	26(1)	27(1)	-2(1)	10(1)	-6(1)
C(9A)	26(2)	31(2)	33(2)	-1(1)	9(1)	-5(1)
C(10A)	28(2)	30(1)	21(1)	0(1)	6(1)	-3(1)
C(11A)	35(2)	29(1)	26(1)	0(1)	10(1)	-6(1)
C(12A)	30(2)	35(2)	26(1)	-1(1)	13(1)	-9(1)
C(13A)	23(1)	32(2)	31(1)	-4(1)	11(1)	-3(1)
C(14A)	28(2)	35(2)	40(2)	-6(1)	16(1)	-7(1)
C(15A)	40(2)	27(1)	29(1)	-2(1)	10(1)	-3(1)
C(16A)	46(2)	34(2)	57(2)	5(2)	18(2)	5(1)
S(1B)	28(1)	27(1)	34(1)	2(1)	10(1)	-1(1)
S(2B)	27(1)	33(1)	33(1)	-8(1)	8(1)	2(1)
N(1B)	26(1)	25(1)	37(1)	-2(1)	14(1)	0(1)
N(2B)	23(1)	26(1)	34(1)	-1(1)	9(1)	-2(1)
C(1B)	23(1)	24(1)	23(1)	1(1)	5(1)	-1(1)
C(2B)	24(1)	29(1)	30(1)	2(1)	13(1)	4(1)
C(3B)	26(1)	27(1)	29(1)	-2(1)	11(1)	2(1)
C(4B)	24(1)	28(1)	20(1)	4(1)	7(1)	3(1)
C(5B)	21(1)	29(1)	34(1)	0(1)	11(1)	4(1)
C(6B)	25(2)	29(1)	30(1)	-2(1)	9(1)	5(1)
C(7B)	26(1)	30(1)	24(1)	-3(1)	7(1)	1(1)
C(8B)	30(2)	25(1)	33(2)	-5(1)	10(1)	-3(1)
C(9B)	25(2)	30(2)	33(2)	-6(1)	9(1)	-6(1)

C(10B)	26(1)	29(1)	23(1)	-1(1)	8(1)	0(1)
C(11B)	27(2)	26(1)	27(1)	1(1)	10(1)	-3(1)
C(12B)	23(1)	34(2)	27(1)	-1(1)	9(1)	-6(1)
C(13B)	22(1)	37(2)	36(2)	3(1)	10(1)	1(1)
C(14B)	29(2)	41(2)	48(2)	-2(1)	19(1)	-5(1)
C(15B)	29(2)	29(1)	26(1)	1(1)	6(1)	0(1)
C(16B)	40(2)	28(2)	42(2)	-2(1)	17(2)	0(1)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for och103s.

	x	y	z	U(eq)
H(1AA)	2303	1158	5881	32
H(2AA)	8945	2356	8938	42
H(2AB)	4705	2921	6214	30
H(3AA)	3708	1723	5932	30
H(5AA)	1779	3403	6061	34
H(6AA)	2794	4590	6369	35
H(8AA)	7268	4900	8379	35
H(9AA)	8432	3875	8666	37
H(11A)	6708	1849	8577	37
H(12A)	5567	2892	8349	36
H(13A)	747	2032	5110	35
H(13B)	1146	2214	6306	35
H(14A)	150	968	5857	51
H(14B)	1173	655	6602	51
H(14C)	775	473	5408	51
H(15A)	7760	1047	7985	40
H(15B)	8022	974	9172	40
H(16A)	9006	60	8775	71
H(16B)	9594	838	9531	71
H(16C)	9331	908	8345	71
H(1BA)	-2665	3842	841	35
H(2BA)	4039	4835	4181	35
H(2BB)	-187	5604	1495	33
H(3BA)	-1223	4427	1024	33
H(5BA)	-3082	6016	1435	33
H(6BA)	-2043	7183	1901	35
H(8BA)	2439	7394	3899	37
H(9BA)	3565	6372	4062	37
H(11B)	1788	4336	3771	33
H(12B)	682	5373	3660	35
H(13C)	-4158	4837	309	39
H(13D)	-3754	4710	1514	39

H(14D)	-4860	3597	675	58
H(14E)	-3872	3143	1298	58
H(14F)	-4270	3267	96	58
H(15C)	2755	3666	3003	36
H(15D)	3073	3413	4171	36
H(16D)	3951	2610	3521	56
H(16E)	4608	3271	4398	56
H(16F)	4293	3527	3232	56

I.2 X-ray crystal structure of 3a (Chapter 3)

Table 1. Crystal data and structure refinement for ova55m.

Identification code	ova55m	
Empirical formula	C20 H28 N2 S3	
Formula weight	392.62	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	$P2_1/n$	
Unit cell dimensions	a = 10.542(2) Å	α= 90°.
	b = 10.2720(19) Å	β= 94.801(4)°.
	c = 19.133(3) Å	γ = 90°.
Volume	2064.5(7) $Å^3$	
Z	4	
Density (calculated)	1.263 Mg/m^3	
Absorption coefficient	0.365 mm^{-1}	
F(000)	840	
Crystal size	0.32 x 0.18 x 0.06 mm^3	
Theta range for data collection	2.13 to 28.28°.	
Index ranges	-11<=h<=13, -13<=k<=13, -25<=l<=21	
Reflections collected	12843	
Independent reflections	4987 [R(int) = 0.1022]	
Completeness to theta = 28.28°	97.4 %	
Absorption correction	Semi-empirical	
Max. and min. transmission	0.9784 and 0.8922	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	4987 / 0 / 226	
Goodness-of-fit on F^2	0.921	
Final R indices [I>2sigma(I)]	R1 = 0.0711, wR2 = 0.1576	
R indices (all data)	R1 = 0.1383, wR2 = 0.1958	
Largest diff. peak and hole	0.630 and -0.378 $e.Å^{-3}$	

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3) for ova55m. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
S(1)	1827(1)	6048(1)	1787(1)	34(1)
S(2)	239(1)	6843(1)	2179(1)	38(1)
S(3)	2958(1)	7559(1)	1508(1)	39(1)
N(1)	1368(3)	8380(3)	-1534(2)	31(1)
N(2)	1996(3)	7807(3)	5181(2)	32(1)
C(1)	2476(4)	7828(4)	612(2)	29(1)
C(2)	3327(4)	7612(4)	111(2)	30(1)
C(3)	2983(3)	7814(4)	-594(2)	30(1)
C(4)	1735(3)	8207(4)	-830(2)	29(1)
C(5)	886(4)	8443(4)	-311(2)	33(1)
C(6)	1263(4)	8272(4)	391(2)	32(1)
C(7)	735(3)	7144(4)	3072(2)	31(1)
C(8)	844(3)	8397(4)	3334(2)	32(1)
C(9)	1251(3)	8638(4)	4027(2)	30(1)
C(10)	1572(3)	7594(3)	4495(2)	26(1)
C(11)	1436(4)	6323(4)	4221(2)	32(1)
C(12)	1018(3)	6112(4)	3529(2)	33(1)
C(13)	2231(4)	9127(4)	5462(2)	32(1)
C(14)	3524(4)	9649(4)	5309(2)	38(1)
C(15)	2447(4)	6746(4)	5645(2)	31(1)
C(16)	1370(4)	6108(4)	6007(2)	36(1)
C(18)	2212(4)	7959(4)	-2063(2)	33(1)
C(19)	2352(4)	6507(4)	-2152(2)	39(1)
C(20)	22(4)	8539(4)	-1767(2)	38(1)
C(21)	-774(4)	7335(5)	-1707(2)	46(1)

Table 3. Bond lengths [Å] and angles [°] for ova55m.

S(1)-S(3)	2.0548(15)
S(1)-S(2)	2.0592(15)
S(2)-C(7)	1.772(4)
S(3)-C(1)	1.769(4)
N(1)-C(4)	1.383(5)
N(1)-C(20)	1.461(5)
N(1)-C(18)	1.467(5)
N(2)-C(10)	1.369(5)
N(2)-C(13)	1.472(5)
N(2)-C(15)	1.460(5)
C(1)-C(2)	1.385(5)
C(1)-C(6)	1.389(5)
C(2)-C(3)	1.381(5)
C(3)-C(4)	1.414(5)
C(4)-C(5)	1.411(5)
C(5)-C(6)	1.380(5)
C(7)-C(12)	1.391(5)
C(7)-C(8)	1.383(5)
C(8)-C(9)	1.381(5)
C(9)-C(10)	1.419(5)
C(10)-C(11)	1.410(5)
C(11)-C(12)	1.376(5)
C(13)-C(14)	1.516(5)
C(15)-C(16)	1.526(5)
C(18)-C(19)	1.510(5)
C(20)-C(21)	1.504(6)

S(3)-S(1)-S(2)	107.59(7)
C(7)-S(2)-S(1)	103.67(13)
C(1)-S(3)-S(1)	104.15(13)
C(4)-N(1)-C(20)	119.8(3)
C(4)-N(1)-C(18)	119.9(3)
C(20)-N(1)-C(18)	116.9(3)
C(10)-N(2)-C(13)	121.9(3)
C(10)-N(2)-C(15)	121.9(3)
C(13)-N(2)-C(15)	115.4(3)

C(2)-C(1)-C(6)	118.4(3)
C(2)-C(1)-S(3)	119.7(3)
C(6)-C(1)-S(3)	121.9(3)
C(3)-C(2)-C(1)	121.3(4)
C(2)-C(3)-C(4)	121.0(3)
N(1)-C(4)-C(5)	121.4(3)
N(1)-C(4)-C(3)	121.7(3)
C(5)-C(4)-C(3)	116.9(3)
C(6)-C(5)-C(4)	120.9(4)
C(5)-C(6)-C(1)	121.4(4)
C(12)-C(7)-C(8)	118.3(4)
C(12)-C(7)-S(2)	120.2(3)
C(8)-C(7)-S(2)	121.4(3)
C(9)-C(8)-C(7)	121.6(4)
C(8)-C(9)-C(10)	120.6(4)
N(2)-C(10)-C(11)	121.3(3)
N(2)-C(10)-C(9)	121.7(3)
C(11)-C(10)-C(9)	117.0(3)
C(12)-C(11)-C(10)	121.1(4)
C(11)-C(12)-C(7)	121.3(4)
N(2)-C(13)-C(14)	112.4(3)
N(2)-C(15)-C(16)	112.3(3)
N(1)-C(18)-C(19)	116.0(3)
N(1)-C(20)-C(21)	114.7(3)

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for ova55m. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
S(1)	37(1)	34(1)	33(1)	-2(1)	7(1)	-1(1)
S(2)	27(1)	55(1)	33(1)	-1(1)	5(1)	-3(1)
S(3)	35(1)	52(1)	31(1)	-2(1)	6(1)	-15(1)
N(1)	26(2)	34(2)	33(2)	2(1)	8(1)	0(1)
N(2)	36(2)	18(2)	40(2)	2(1)	4(2)	-7(1)
C(1)	29(2)	33(2)	26(2)	-1(1)	1(2)	-9(2)
C(2)	21(2)	36(2)	34(2)	-4(2)	2(2)	-7(2)
C(3)	24(2)	31(2)	35(2)	-4(2)	10(2)	-4(2)
C(4)	27(2)	24(2)	36(2)	0(2)	7(2)	-6(2)
C(5)	26(2)	30(2)	44(2)	2(2)	5(2)	3(2)
C(6)	32(2)	32(2)	34(2)	-4(2)	16(2)	-2(2)
C(7)	23(2)	32(2)	37(2)	1(2)	6(2)	-3(2)
C(8)	23(2)	34(2)	39(2)	9(2)	7(2)	2(2)
C(9)	27(2)	23(2)	42(2)	5(2)	9(2)	-4(2)
C(10)	24(2)	23(2)	32(2)	4(2)	7(2)	-3(2)
C(11)	33(2)	27(2)	36(2)	2(2)	5(2)	-3(2)
C(12)	28(2)	36(2)	35(2)	-4(2)	8(2)	-7(2)
C(13)	33(2)	26(2)	38(2)	-5(2)	4(2)	-4(2)
C(14)	36(2)	22(2)	54(2)	-1(2)	1(2)	-6(2)
C(15)	32(2)	25(2)	35(2)	2(2)	3(2)	1(2)
C(16)	42(2)	30(2)	36(2)	7(2)	8(2)	3(2)
C(18)	35(2)	34(2)	31(2)	2(2)	9(2)	-4(2)
C(19)	39(2)	40(2)	39(2)	-8(2)	11(2)	-4(2)
C(20)	30(2)	47(3)	36(2)	7(2)	1(2)	8(2)
C(21)	27(2)	57(3)	52(3)	4(2)	1(2)	-5(2)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters (Å^{2}x 10^3) for ova55m.

	x	y	z	U(eq)
H(2A)	4164	7320	254	36
H(3A)	3595	7686	-925	35
H(5A)	42	8723	-448	40
H(6A)	682	8463	731	38
H(8A)	633	9110	3030	38
H(9A)	1316	9509	4192	36
H(11A)	1637	5599	4519	38
H(12A)	921	5245	3361	39
H(13A)	2173	9114	5976	39
H(13B)	1561	9720	5254	39
H(14A)	3633	10530	5501	56
H(14B)	3584	9673	4801	56
H(14C)	4192	9081	5527	56
H(15A)	3090	7088	6006	37
H(15B)	2867	6081	5369	37
H(16A)	1714	5401	6310	54
H(16B)	737	5757	5652	54
H(16C)	966	6758	6291	54
H(18A)	3068	8332	-1939	40
H(18B)	1891	8334	-2521	40
H(19A)	2932	6332	-2515	58
H(19B)	1517	6124	-2290	58
H(19C)	2698	6123	-1707	58
H(20A)	-50	8825	-2263	46
H(20B)	-334	9241	-1486	46
H(21A)	-1660	7526	-1870	68
H(21B)	-729	7053	-1216	68
H(21C)	-450	6641	-1995	68

I.2 X-ray crystal structure of 5 (Chapter 3)

Table 1. Crystal data and structure refinement for **5**.

Identification code	och107m	
Empirical formula	C28 H30 N4 O4 S2	
Formula weight	550.68	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	C2/c	
Unit cell dimensions	a = 25.908(3) Å	α= 90°.
	b = 8.2753(12) Å	β= 117.428(2)°.
	c = 14.3757(19) Å	γ = 90°.
Volume	2735.7(6) $Å^3$	
Z	4	
Density (calculated)	1.337 Mg/m^3	
Absorption coefficient	0.236 mm^{-1}	
F(000)	1160	
Crystal size	0.26 x 0.15 x 0.05 mm^3	
Theta range for data collection	1.77 to 28.32°.	
Index ranges	-34<=h<=17, -10<=k<=9, -18<=l<=19	
Reflections collected	8455	
Independent reflections	3247 [R(int) = 0.1030]	
Completeness to theta = 28.32°	95.3 %	
Absorption correction	Semi-empirical	
Max. and min. transmission	0.9883 and 0.9412	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	3247 / 0 / 172	
Goodness-of-fit on F^2	0.909	
Final R indices [I>2sigma(I)]	R1 = 0.0586, wR2 = 0.1176	
R indices (all data)	R1 = 0.0977, wR2 = 0.1308	
Largest diff. peak and hole	0.338 and -0.616 e.$Å^{-3}$	

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3) for Och107m. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
S(1)	451(1)	-6067(1)	2854(1)	24(1)
N(1)	1112(1)	-501(2)	5835(2)	23(1)
N(2)	1935(1)	3079(2)	5649(2)	26(1)
O(1)	1834(1)	2931(2)	3974(1)	37(1)
O(2)	2170(1)	4019(2)	7306(2)	46(1)
C(1)	644(1)	-4424(3)	3738(2)	22(1)
C(2)	827(1)	-2965(3)	3508(2)	24(1)
C(3)	989(1)	-1678(3)	4194(2)	23(1)
C(4)	963(1)	-1781(3)	5155(2)	21(1)
C(5)	778(1)	-3276(3)	5377(2)	23(1)
C(6)	625(1)	-4550(3)	4692(2)	22(1)
C(7)	1199(1)	1105(3)	5527(2)	23(1)
C(8)	1831(1)	1397(3)	5795(2)	29(1)
C(9)	1940(1)	3693(3)	4754(2)	29(1)
C(10)	2111(1)	5428(3)	4990(2)	33(1)
C(11)	2210(1)	5753(3)	5958(2)	36(1)
C(12)	2111(1)	4243(3)	6429(2)	32(1)
C(13)	1048(1)	-587(3)	6792(2)	25(1)
C(14)	422(1)	-270(3)	6576(2)	30(1)

Table 3. Bond lengths [Å] and angles [°] for Och107m.

S(1)-C(1)	1.769(2)
S(1)-S(1)#1	2.0734(12)
N(1)-C(4)	1.371(3)
N(1)-C(7)	1.452(3)
N(1)-C(13)	1.460(3)
N(2)-C(12)	1.386(3)
N(2)-C(9)	1.390(3)
N(2)-C(8)	1.450(3)
O(1)-C(9)	1.203(3)
O(2)-C(12)	1.212(3)
C(1)-C(2)	1.393(3)
C(1)-C(6)	1.399(3)
C(2)-C(3)	1.379(3)
C(3)-C(4)	1.417(3)
C(4)-C(5)	1.415(3)
C(5)-C(6)	1.371(3)
C(7)-C(8)	1.519(3)
C(9)-C(10)	1.495(4)
C(10)-C(11)	1.322(4)
C(11)-C(12)	1.498(4)
C(13)-C(14)	1.527(3)
C(1)-S(1)-S(1)#1	103.54(8)
C(4)-N(1)-C(7)	121.43(19)
C(4)-N(1)-C(13)	121.16(19)
C(7)-N(1)-C(13)	115.71(18)
C(12)-N(2)-C(9)	110.7(2)
C(12)-N(2)-C(8)	124.3(2)
C(9)-N(2)-C(8)	124.5(2)
C(2)-C(1)-C(6)	117.8(2)
C(2)-C(1)-S(1)	120.72(17)
C(6)-C(1)-S(1)	121.43(18)
C(3)-C(2)-C(1)	121.7(2)
C(2)-C(3)-C(4)	121.0(2)
N(1)-C(4)-C(3)	121.7(2)
N(1)-C(4)-C(5)	121.8(2)

C(3)-C(4)-C(5)	116.5(2)
C(6)-C(5)-C(4)	121.8(2)
C(5)-C(6)-C(1)	121.2(2)
N(1)-C(7)-C(8)	111.72(19)
N(2)-C(8)-C(7)	111.44(19)
O(1)-C(9)-N(2)	125.2(2)
O(1)-C(9)-C(10)	129.2(2)
N(2)-C(9)-C(10)	105.6(2)
C(11)-C(10)-C(9)	109.1(2)
C(10)-C(11)-C(12)	108.4(2)
O(2)-C(12)-N(2)	124.9(2)
O(2)-C(12)-C(11)	129.1(2)
N(2)-C(12)-C(11)	106.1(2)
N(1)-C(13)-C(14)	111.73(19)

Symmetry transformations used to generate equivalent atoms: #1 -x,y,-z+1/2

Table 4. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for Och107m. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
S(1)	22(1)	24(1)	24(1)	-3(1)	8(1)	2(1)
N(1)	27(1)	20(1)	23(1)	-2(1)	14(1)	-5(1)
N(2)	22(1)	20(1)	33(1)	0(1)	10(1)	-3(1)
O(1)	28(1)	47(1)	36(1)	-4(1)	13(1)	1(1)
O(2)	55(1)	41(1)	34(1)	-5(1)	13(1)	-7(1)
C(1)	17(1)	23(1)	23(1)	-2(1)	8(1)	0(1)
C(2)	26(1)	27(1)	20(1)	4(1)	10(1)	1(1)
C(3)	23(1)	22(1)	27(1)	3(1)	13(1)	-2(1)
C(4)	15(1)	24(1)	23(1)	1(1)	7(1)	2(1)
C(5)	24(1)	24(1)	23(1)	1(1)	14(1)	-3(1)
C(6)	21(1)	20(1)	24(1)	3(1)	10(1)	0(1)
C(7)	24(1)	17(1)	30(1)	0(1)	13(1)	1(1)
C(8)	25(1)	20(1)	37(2)	1(1)	11(1)	-1(1)
C(9)	14(1)	32(2)	39(2)	3(1)	9(1)	0(1)
C(10)	21(1)	28(2)	47(2)	10(1)	14(1)	-1(1)
C(11)	29(2)	22(2)	53(2)	-2(1)	15(1)	-5(1)
C(12)	26(1)	26(2)	34(2)	-3(1)	5(1)	-3(1)
C(13)	28(1)	27(1)	19(1)	-3(1)	10(1)	-4(1)
C(14)	31(1)	32(2)	31(1)	-4(1)	18(1)	-3(1)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for Och107m.

	x	y	z	U(eq)
H(2A)	841	-2853	2862	29
H(3A)	1120	-707	4019	28
H(5A)	761	-3400	6019	27
H(6A)	504	-5536	4869	26
H(7A)	1075	1921	5888	28
H(7B)	954	1239	4763	28
H(8A)	2083	1085	6533	34
H(8B)	1936	711	5343	34
H(10A)	2144	6177	4520	39
H(11A)	2323	6772	6298	43
H(13A)	1307	220	7301	30
H(13B)	1168	-1672	7109	30
H(14A)	396	-331	7234	45
H(14B)	165	-1083	6085	45
H(14C)	303	809	6270	45

I.2 X-ray crystal structure of 7 (Chapter 3)

Table 1. Crystal data and structure refinement for **7**.

Identification code	och123	
Empirical formula	C22 H18 N4 O4 S2	
Formula weight	466.52	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	$P2_1/c$	
Unit cell dimensions	a = 6.210(4) Å	α= 90°.
	b = 23.380(15) Å	β= 90.000(14)°.
	c = 14.011(9) Å	γ = 90°.
Volume	2034(2) $Å^3$	
Z	4	
Density (calculated)	1.523 Mg/m^3	
Absorption coefficient	0.302 mm^{-1}	
F(000)	968	
Crystal size	0.12 x 0.06 x 0.03 mm^3	
Theta range for data collection	1.69 to 28.38°.	
Index ranges	-7<=h<=8, -21<=k<=31, -18<=l<=18	
Reflections collected	10256	
Independent reflections	4685 [R(int) = 0.3993]	
Completeness to theta = 28.38°	91.8 %	
Absorption correction	Semi-empirical	
Max. and min. transmission	0.9910 and 0.9646	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	4685 / 0 / 289	
Goodness-of-fit on F^2	0.869	
Final R indices [I>2sigma(I)]	R1 = 0.0956, wR2 = 0.1601	
R indices (all data)	R1 = 0.3005, wR2 = 0.2329	
Largest diff. peak and hole	0.475 and -0.481 $e.Å^{-3}$	

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3) for och123. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
S(1)	1807(3)	2826(1)	-156(1)	38(1)
S(2)	1285(3)	2919(1)	1288(2)	39(1)
O(1)	1548(8)	192(2)	-2311(4)	39(1)
O(2)	-4168(10)	-459(2)	-600(4)	47(2)
O(3)	7416(9)	1159(2)	5701(4)	46(2)
O(4)	2315(9)	343(2)	3827(4)	49(2)
N(1)	-3804(10)	880(2)	-1365(5)	43(2)
N(2)	-1650(10)	14(2)	-1517(4)	31(2)
N(3)	7249(10)	1386(2)	3406(4)	35(2)
N(4)	5227(10)	797(2)	4528(4)	30(2)
C(1)	136(13)	2238(3)	-479(5)	30(2)
C(2)	889(12)	1870(3)	-1180(5)	30(2)
C(3)	-394(13)	1422(3)	-1498(5)	34(2)
C(4)	-2455(13)	1331(3)	-1098(5)	33(2)
C(5)	-3178(12)	1704(3)	-391(5)	33(2)
C(6)	-1918(13)	2152(3)	-97(6)	39(2)
C(7)	-3085(13)	429(3)	-1995(5)	35(2)
C(8)	459(13)	-84(3)	-1774(5)	29(2)
C(9)	1143(15)	-609(3)	-1233(6)	46(2)
C(10)	-474(13)	-787(3)	-730(6)	37(2)
C(11)	-2344(15)	-413(3)	-907(5)	33(2)
C(12)	3135(12)	2446(3)	1823(5)	31(2)
C(13)	2468(13)	1902(3)	2096(5)	37(2)
C(14)	3784(13)	1540(3)	2606(6)	39(2)
C(15)	5825(12)	1725(3)	2886(5)	28(2)
C(16)	6477(13)	2279(3)	2634(5)	34(2)
C(17)	5178(13)	2628(3)	2121(5)	34(2)
C(18)	6696(13)	823(3)	3748(6)	36(2)
C(19)	3148(14)	548(3)	4520(7)	39(2)
C(20)	2363(13)	573(3)	5509(6)	40(2)
C(21)	3841(13)	810(3)	6049(6)	33(2)
C(22)	5686(15)	957(3)	5466(6)	37(2)

Table 3. Bond lengths [Å] and angles [°] for och123.

S(1)-C(1)	1.781(7)
S(1)-S(2)	2.060(3)
S(2)-C(12)	1.763(7)
O(1)-C(8)	1.200(8)
O(2)-C(11)	1.216(9)
O(3)-C(22)	1.219(9)
O(4)-C(19)	1.201(9)
N(1)-C(4)	1.397(9)
N(1)-C(7)	1.446(9)
N(2)-C(8)	1.378(9)
N(2)-C(11)	1.382(9)
N(2)-C(7)	1.477(8)
N(3)-C(15)	1.393(8)
N(3)-C(18)	1.443(9)
N(4)-C(22)	1.397(9)
N(4)-C(19)	1.416(10)
N(4)-C(18)	1.424(10)
C(1)-C(2)	1.386(10)
C(1)-C(6)	1.398(11)
C(2)-C(3)	1.389(10)
C(3)-C(4)	1.414(11)
C(4)-C(5)	1.395(10)
C(5)-C(6)	1.371(10)
C(8)-C(9)	1.504(10)
C(9)-C(10)	1.296(11)
C(10)-C(11)	1.475(11)
C(12)-C(13)	1.393(10)
C(12)-C(17)	1.401(10)
C(13)-C(14)	1.375(10)
C(14)-C(15)	1.395(10)
C(15)-C(16)	1.402(10)
C(16)-C(17)	1.353(9)
C(19)-C(20)	1.469(12)
C(20)-C(21)	1.312(10)
C(21)-C(22)	1.448(12)

C(1)-S(1)-S(2)	103.9(3)
C(12)-S(2)-S(1)	104.4(3)
C(4)-N(1)-C(7)	121.9(7)
C(8)-N(2)-C(11)	109.7(7)
C(8)-N(2)-C(7)	124.4(7)
C(11)-N(2)-C(7)	124.5(7)
C(15)-N(3)-C(18)	122.8(6)
C(22)-N(4)-C(19)	107.7(7)
C(22)-N(4)-C(18)	125.4(7)
C(19)-N(4)-C(18)	126.6(6)
C(2)-C(1)-C(6)	119.4(7)
C(2)-C(1)-S(1)	117.5(6)
C(6)-C(1)-S(1)	123.0(6)
C(1)-C(2)-C(3)	120.1(8)
C(2)-C(3)-C(4)	120.4(8)
C(5)-C(4)-N(1)	118.0(8)
C(5)-C(4)-C(3)	118.6(7)
N(1)-C(4)-C(3)	123.5(7)
C(6)-C(5)-C(4)	120.5(8)
C(5)-C(6)-C(1)	121.0(8)
N(1)-C(7)-N(2)	112.9(6)
O(1)-C(8)-N(2)	127.6(7)
O(1)-C(8)-C(9)	126.6(7)
N(2)-C(8)-C(9)	105.9(7)
C(10)-C(9)-C(8)	108.5(8)
C(9)-C(10)-C(11)	109.1(7)
O(2)-C(11)-N(2)	125.0(8)
O(2)-C(11)-C(10)	128.3(7)
N(2)-C(11)-C(10)	106.7(8)
C(13)-C(12)-C(17)	117.7(7)
C(13)-C(12)-S(2)	119.8(6)
C(17)-C(12)-S(2)	121.8(5)
C(14)-C(13)-C(12)	121.8(7)
C(13)-C(14)-C(15)	119.7(7)
N(3)-C(15)-C(14)	123.2(7)
N(3)-C(15)-C(16)	118.3(6)
C(14)-C(15)-C(16)	118.6(7)
C(17)-C(16)-C(15)	121.1(7)

C(16)-C(17)-C(12)	121.1(6)
N(4)-C(18)-N(3)	116.6(6)
O(4)-C(19)-N(4)	124.3(9)
O(4)-C(19)-C(20)	129.5(8)
N(4)-C(19)-C(20)	106.2(7)
C(21)-C(20)-C(19)	109.2(8)
C(20)-C(21)-C(22)	109.2(8)
O(3)-C(22)-N(4)	122.5(8)
O(3)-C(22)-C(21)	129.6(7)
N(4)-C(22)-C(21)	107.8(7)

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for och123. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
S(1)	45(1)	26(1)	42(1)	4(1)	-9(1)	-4(1)
S(2)	49(1)	25(1)	43(1)	-1(1)	-12(1)	2(1)
O(1)	37(3)	31(3)	48(4)	2(3)	9(3)	1(3)
O(2)	43(4)	39(3)	59(4)	0(3)	0(3)	-9(3)
O(3)	44(4)	36(3)	57(4)	2(3)	-18(3)	-7(3)
O(4)	51(4)	30(3)	66(4)	-1(3)	-15(3)	-11(3)
N(1)	30(4)	33(4)	66(5)	-18(3)	-10(4)	6(3)
N(2)	33(4)	28(3)	32(4)	-3(3)	0(3)	6(3)
N(3)	29(4)	29(3)	48(4)	7(3)	-8(3)	-4(3)
N(4)	32(4)	20(3)	38(4)	-2(3)	-5(3)	0(3)
C(1)	35(5)	30(4)	26(4)	7(3)	-9(4)	2(4)
C(2)	28(4)	25(4)	37(5)	5(4)	-1(4)	3(4)
C(3)	37(5)	31(4)	34(5)	2(4)	-5(4)	0(4)
C(4)	33(5)	29(4)	39(5)	-1(4)	-13(4)	5(4)
C(5)	26(4)	35(4)	38(5)	7(4)	2(4)	6(4)
C(6)	43(6)	29(4)	44(5)	5(4)	-15(4)	-4(4)
C(7)	47(5)	23(4)	34(5)	1(3)	-11(4)	-1(4)
C(8)	32(5)	27(4)	28(5)	-4(4)	0(4)	2(4)
C(9)	41(6)	33(4)	64(6)	12(4)	-15(5)	0(4)
C(10)	37(5)	32(4)	42(5)	4(4)	-5(4)	-8(4)
C(11)	39(5)	34(4)	28(5)	5(4)	2(4)	-14(4)
C(12)	41(5)	22(4)	29(4)	2(3)	-2(4)	-2(4)
C(13)	42(5)	32(4)	38(5)	2(4)	-7(4)	-19(4)
C(14)	48(6)	25(4)	43(5)	-4(4)	-8(4)	-4(4)
C(15)	28(4)	34(4)	23(4)	1(3)	-1(3)	5(4)
C(16)	37(5)	26(4)	38(5)	1(4)	-8(4)	-10(4)
C(17)	44(5)	20(3)	37(5)	-1(3)	-7(4)	-14(4)
C(18)	42(5)	27(4)	38(5)	-10(4)	-3(4)	6(4)
C(19)	45(5)	16(4)	58(6)	3(4)	-15(5)	0(4)
C(20)	41(5)	21(4)	58(6)	6(4)	-3(5)	2(4)
C(21)	32(5)	21(4)	46(5)	-2(4)	1(4)	0(4)
C(22)	49(6)	18(4)	46(5)	-5(4)	-12(5)	15(4)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for och123.

	x	y	z	U(eq)
H(1A)	-5130	870	-1143	51
H(3A)	8541	1522	3532	42
H(2A)	2284	1925	-1443	36
H(3B)	117	1176	-1987	41
H(5A)	-4555	1647	-111	39
H(6A)	-2449	2408	374	47
H(7A)	-2313	601	-2542	42
H(7B)	-4356	224	-2248	42
H(9A)	2528	-781	-1256	55
H(10A)	-458	-1107	-312	44
H(13A)	1065	1776	1925	45
H(14A)	3305	1167	2768	47
H(16A)	7855	2413	2828	40
H(17A)	5659	3001	1959	40
H(18A)	6073	604	3210	43
H(18B)	8041	627	3941	43
H(20A)	1002	440	5723	48
H(21A)	3720	875	6716	40

I.2 X-ray crystal structure of 4 (Chapter 5)

Table 1. Crystal data and structure refinement for **4**.

Identification code	och149m	
Empirical formula	C11 H10 N2 S2	
Formula weight	234.33	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	$P2_1/n$	
Unit cell dimensions	a = 8.9677(10) Å	α= 90°.
	b = 13.0644(14) Å	β= 103.096(2)°.
	c = 9.4776(10) Å	γ = 90°.
Volume	1081.5(2) $Å^3$	
Z	4	
Density (calculated)	1.439 Mg/m^3	
Absorption coefficient	0.457 mm^{-1}	
F(000)	488	
Crystal size	0.42 x 0.23 x 0.12 mm^3	
Theta range for data collection	2.70 to 28.29°.	
Index ranges	-11<=h<=11, -17<=k<=16, -12<=l<=11	
Reflections collected	6677	
Independent reflections	2603 [R(int) = 0.0547]	
Completeness to theta = 28.29°	97.1 %	
Absorption correction	Semi-empirical	
Max. and min. transmission	0.9472 and 0.8312	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	2603 / 0 / 136	
Goodness-of-fit on F^2	1.058	
Final R indices [I>2sigma(I)]	R1 = 0.0422, wR2 = 0.1125	
R indices (all data)	R1 = 0.0483, wR2 = 0.1162	
Largest diff. peak and hole	0.412 and -0.654 $e.Å^{-3}$	

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3) for och149m. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
S(1)	1991(1)	3545(1)	4243(1)	27(1)
S(2)	1567(1)	2739(1)	5971(1)	27(1)
N(1)	4450(2)	4157(1)	3499(2)	27(1)
N(2)	2630(2)	5602(1)	11009(2)	29(1)
C(1)	4018(2)	3554(1)	4476(2)	23(1)
C(2)	5039(2)	3011(1)	5540(2)	27(1)
C(3)	6587(2)	3098(1)	5576(2)	30(1)
C(4)	7059(2)	3717(1)	4570(2)	31(1)
C(5)	5956(2)	4226(1)	3558(2)	29(1)
C(6)	1886(2)	3640(1)	7413(2)	23(1)
C(7)	846(2)	4417(1)	7491(2)	26(1)
C(8)	1085(2)	5072(1)	8666(2)	26(1)
C(9)	2373(2)	4957(1)	9822(2)	23(1)
C(10)	3422(2)	4185(1)	9724(2)	24(1)
C(11)	3188(2)	3537(1)	8536(2)	23(1)

Table 3. Bond lengths [Å] and angles [°] for och149m.

S(1)-C(1)	1.7813(17)
S(1)-S(2)	2.0531(6)
S(2)-C(6)	1.7770(16)
N(1)-C(1)	1.337(2)
N(1)-C(5)	1.342(2)
N(2)-C(9)	1.383(2)
C(1)-C(2)	1.393(2)
C(2)-C(3)	1.386(2)
C(3)-C(4)	1.387(3)
C(4)-C(5)	1.382(2)
C(6)-C(7)	1.392(2)
C(6)-C(11)	1.397(2)
C(7)-C(8)	1.382(2)
C(8)-C(9)	1.408(2)
C(9)-C(10)	1.397(2)
C(10)-C(11)	1.385(2)
C(1)-S(1)-S(2)	105.83(5)
C(6)-S(2)-S(1)	104.33(5)
C(1)-N(1)-C(5)	117.22(14)
N(1)-C(1)-C(2)	123.72(16)
N(1)-C(1)-S(1)	111.44(12)
C(2)-C(1)-S(1)	124.84(13)
C(3)-C(2)-C(1)	117.79(16)
C(4)-C(3)-C(2)	119.41(16)
C(5)-C(4)-C(3)	118.40(17)
N(1)-C(5)-C(4)	123.46(16)
C(7)-C(6)-C(11)	118.94(14)
C(7)-C(6)-S(2)	122.07(12)
C(11)-C(6)-S(2)	118.94(12)
C(8)-C(7)-C(6)	120.74(15)
C(7)-C(8)-C(9)	120.72(15)
N(2)-C(9)-C(10)	120.28(14)
N(2)-C(9)-C(8)	121.59(15)
C(10)-C(9)-C(8)	118.11(14)
C(11)-C(10)-C(9)	120.99(14)
C(10)-C(11)-C(6)	120.48(15)

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for och149m. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
S(1)	27(1)	32(1)	20(1)	0(1)	0(1)	0(1)
S(2)	30(1)	27(1)	24(1)	-4(1)	5(1)	-7(1)
N(1)	32(1)	27(1)	19(1)	0(1)	3(1)	-1(1)
N(2)	30(1)	31(1)	25(1)	-5(1)	3(1)	3(1)
C(1)	28(1)	22(1)	18(1)	-4(1)	3(1)	0(1)
C(2)	33(1)	26(1)	21(1)	2(1)	5(1)	5(1)
C(3)	31(1)	34(1)	23(1)	-2(1)	0(1)	9(1)
C(4)	27(1)	35(1)	29(1)	-7(1)	6(1)	-1(1)
C(5)	35(1)	29(1)	24(1)	-3(1)	8(1)	-5(1)
C(6)	24(1)	24(1)	20(1)	-1(1)	3(1)	-5(1)
C(7)	22(1)	31(1)	24(1)	2(1)	1(1)	-1(1)
C(8)	24(1)	28(1)	24(1)	1(1)	3(1)	4(1)
C(9)	25(1)	24(1)	20(1)	1(1)	5(1)	-3(1)
C(10)	21(1)	28(1)	21(1)	3(1)	2(1)	0(1)
C(11)	24(1)	24(1)	23(1)	2(1)	4(1)	1(1)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for och149m.

	x	y	z	U(eq)
H(2A)	3457	5527	11706	35
H(2B)	1966	6088	11067	35
H(2C)	4686	2594	6219	32
H(3A)	7319	2737	6283	36
H(4A)	8115	3790	4577	37
H(5A)	6281	4646	2866	35
H(7A)	-38	4498	6728	32
H(8A)	373	5606	8694	31
H(10A)	4310	4103	10482	28
H(11A)	3918	3020	8485	28

Appendix II

Appendix II ^{1}H NMR and MS Spectra of the Dissucinimide Derivatives Used in Chapter 3

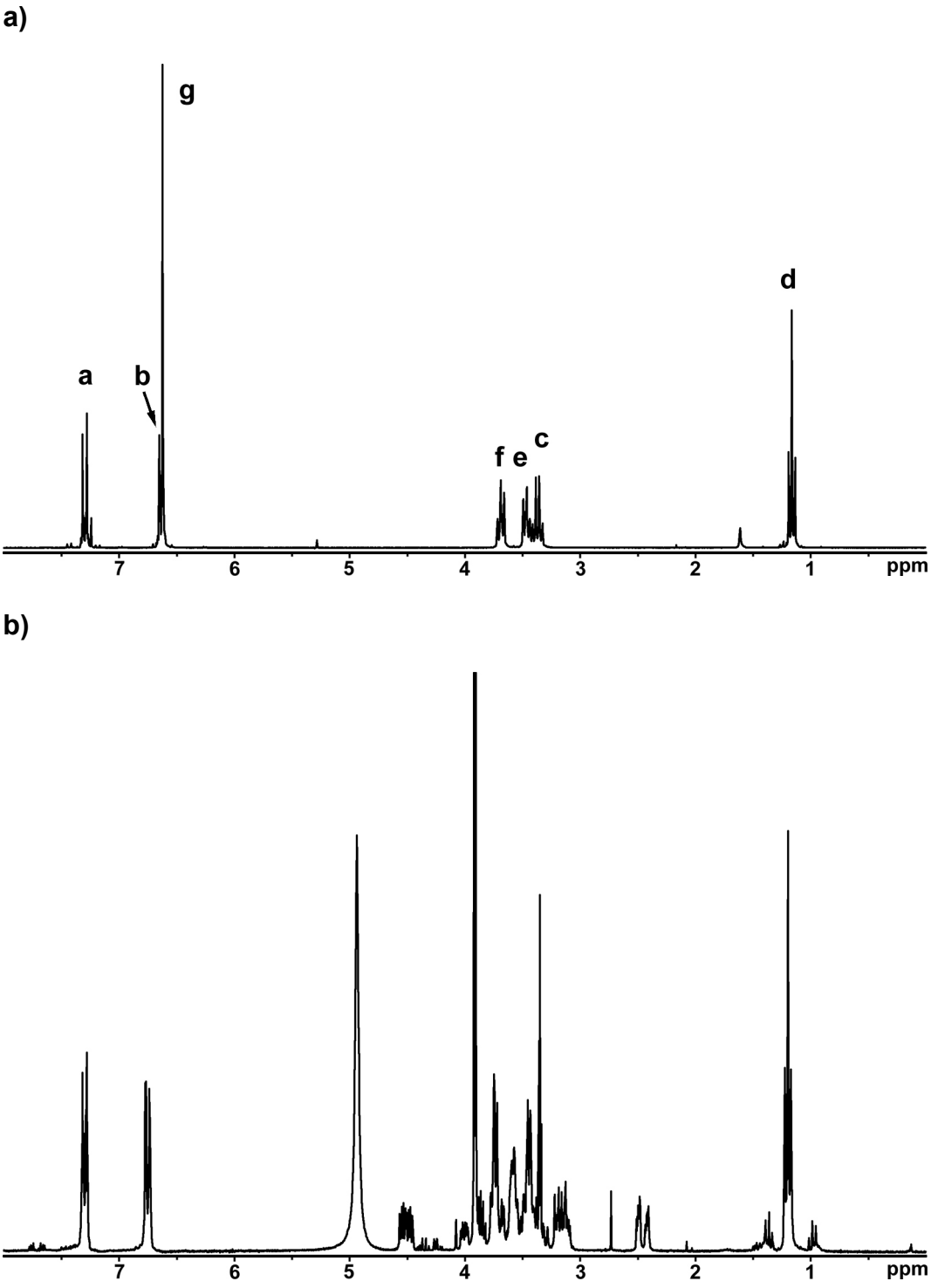
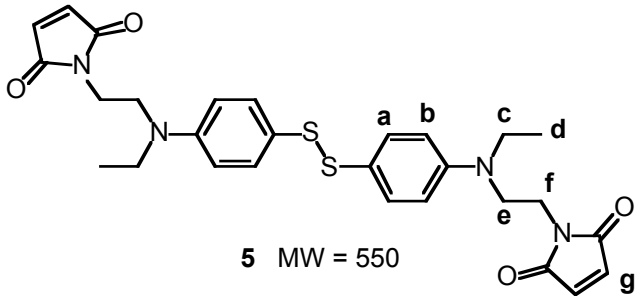
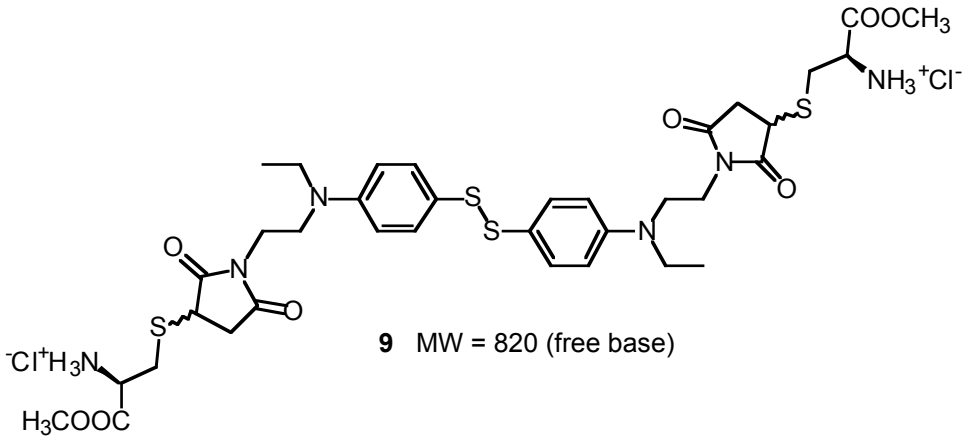


Figure II.1. ^{1}H NMR spectra of **5** in $CDCl_3$ (a) and **9** in CD_3OD (b).



5 MW = 550



9 MW = 820 (free base)

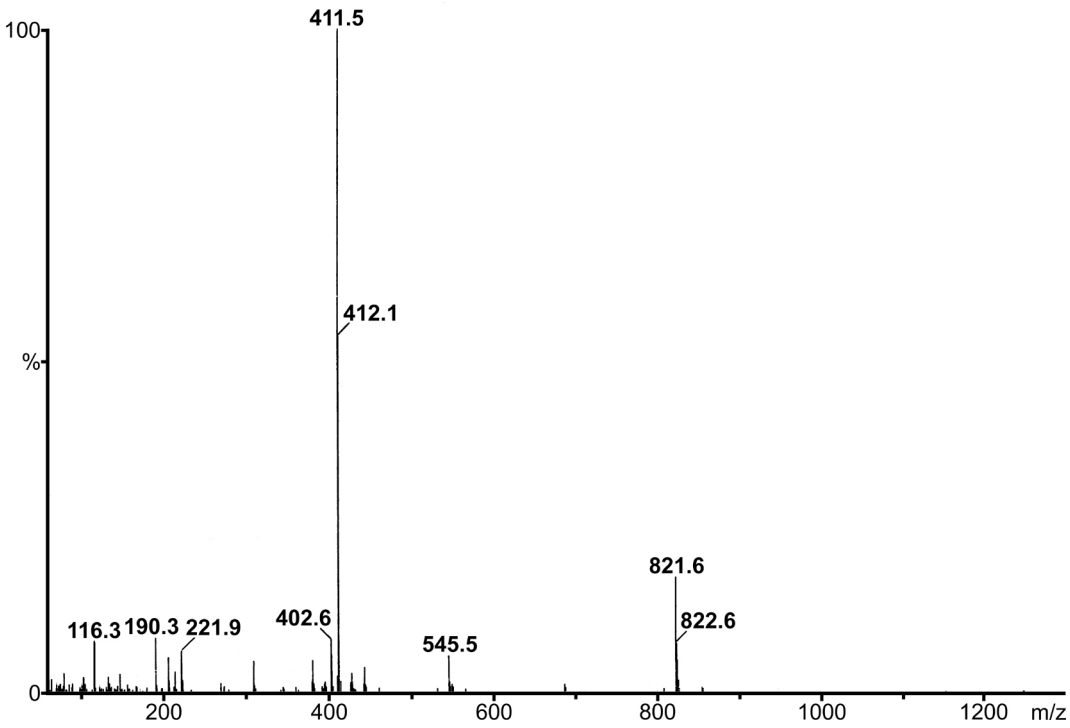


Figure II.2. MS-ES of **9**.

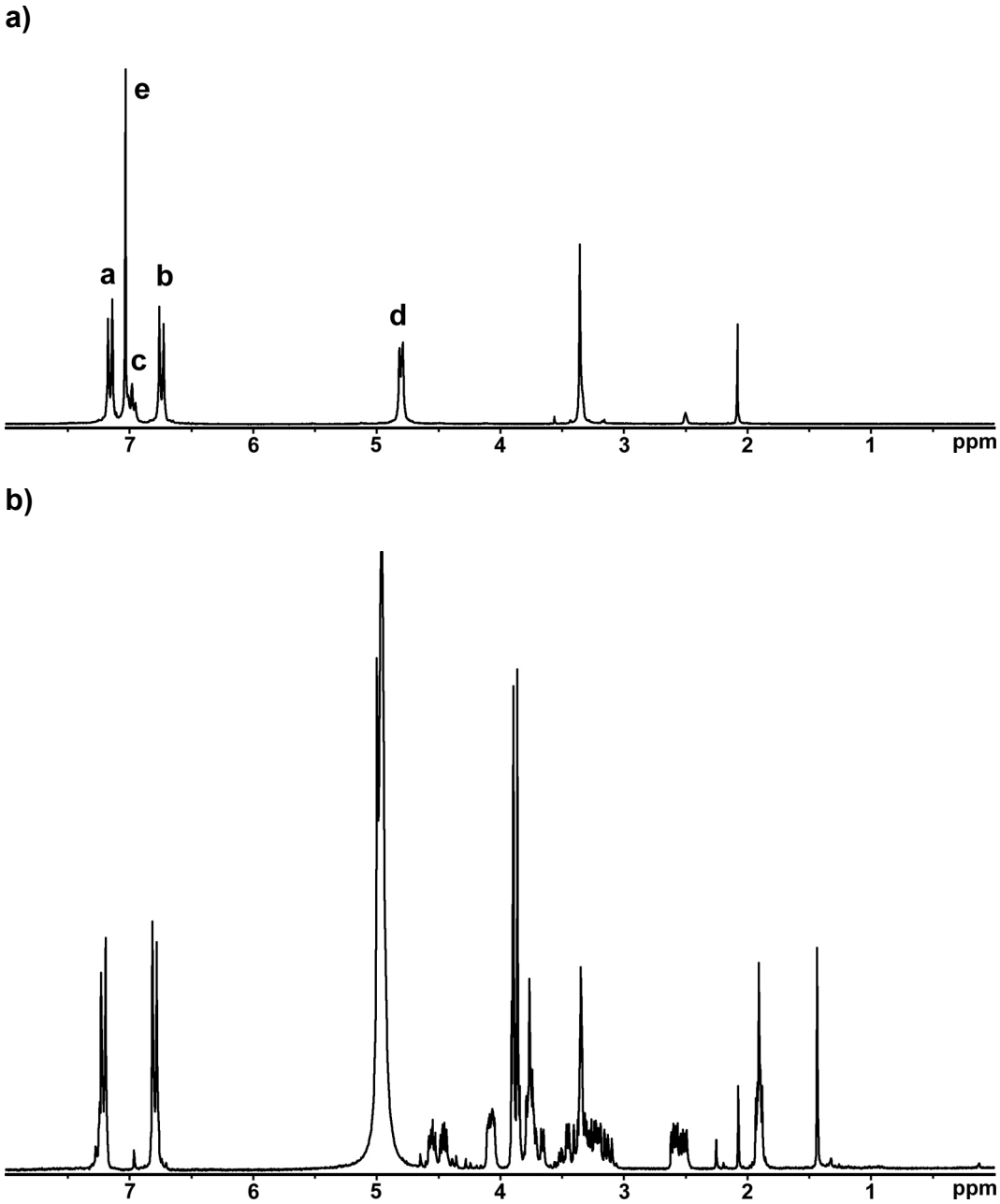
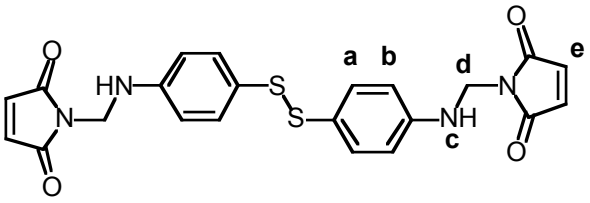
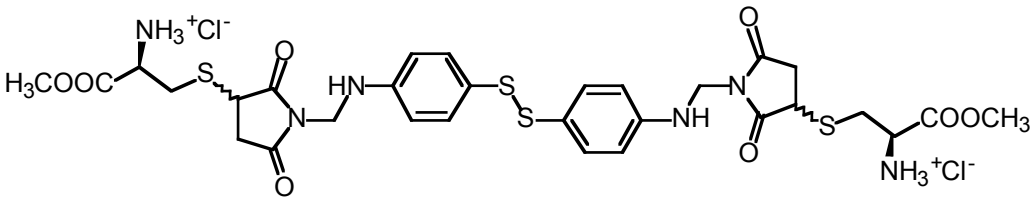


Figure II.3. ^{1}H NMR spectra of **7** in DMSO-d_6 (a) and **10** in CD_3OD (b).



7 MW = 466



10 MW = 736 (free base)

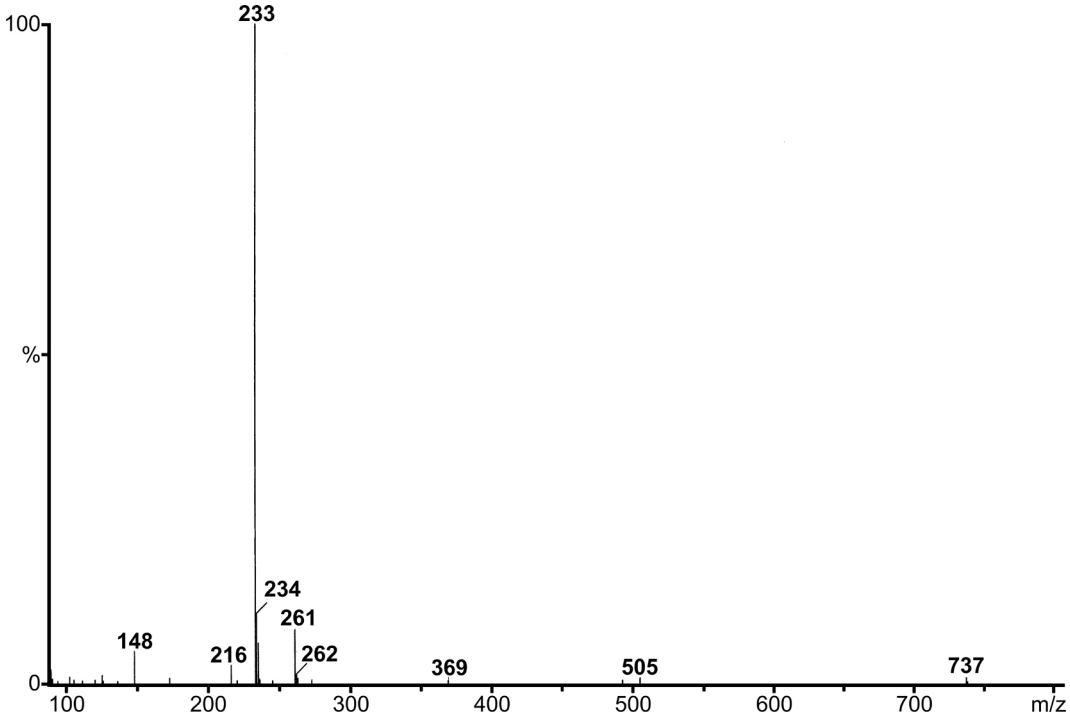


Figure II.4. MS-ES of **10**.

Appendix III

Appendix III UV/Vis. Absorption Spectra of Aromatic Disulfides Used in Chapter 3

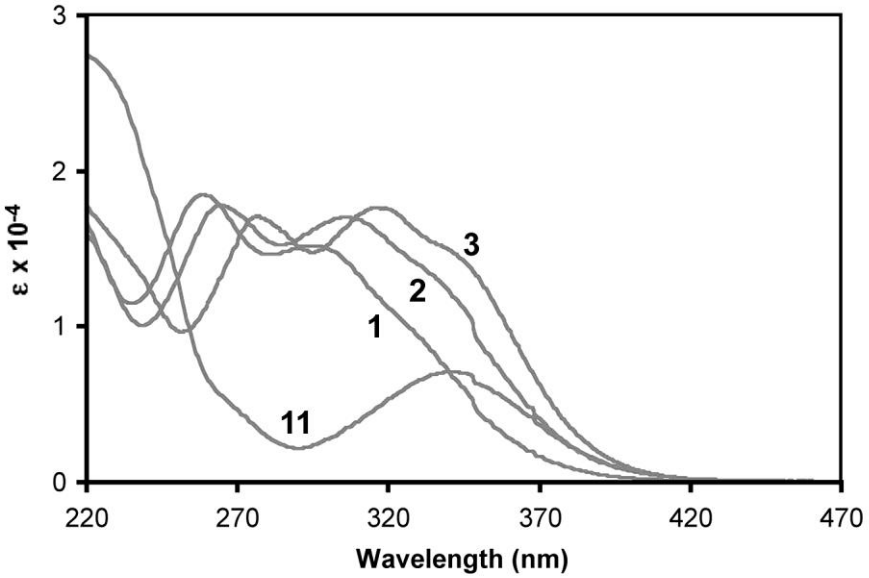


Figure III.1. UV/Vis. absortion spectra of **11**, **1**, **2**, and **3** in acetonitrile.

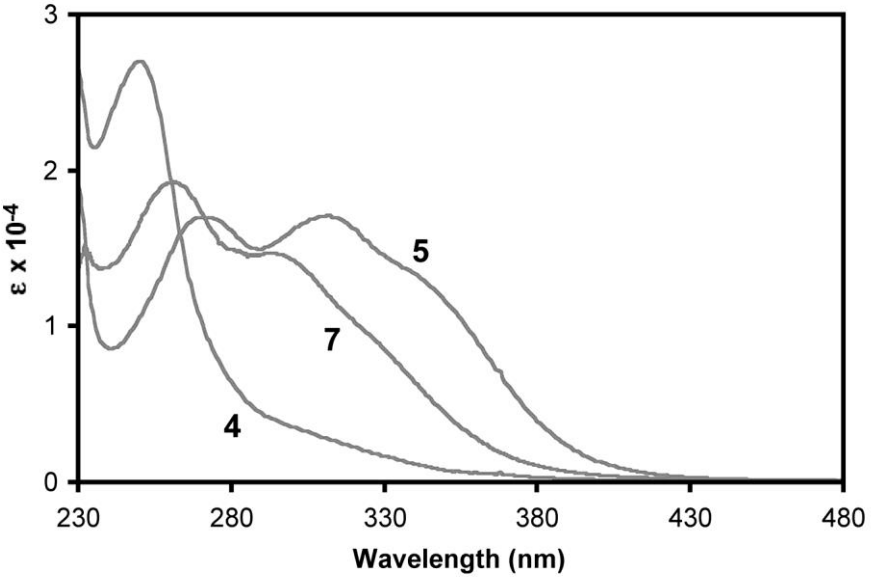


Figure III.2. UV/Vis. absortion spectra of **4** and **5** in acetonitrile and **7** in THF.

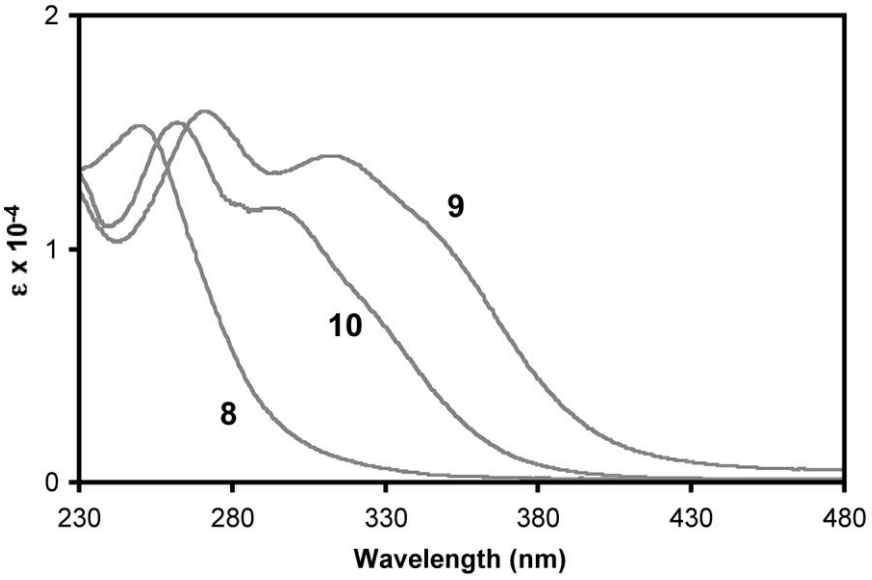


Figure III.3. UV/Vis. absortion spectra of **8**, **9** and **10** in water.